University of Zurich
Institute of Mathematics

Master Thesis

Kingman's coalescent

Author:
Chiara Romano
20-714-069

Supervisor:
Prof. Dr. Jean Bertoin

May 2026

Abstract

Coalescent processes are stochastic models that describe the merging of ancestral lineages, providing a robust mathematical framework for the study of genealogical relationships. The aim of this master's thesis is to provide an introduction to a specific coalescent process known as Kingman's coalescent.

Kingman's coalescent is defined as a continuous-time Markov chain on the space of partitions of $\mathbb{N}$. To rigorously analyze this model, we first establish the necessary probabilistic foundations. This includes the theory of continuous-time Markov chains, their infinitesimal generators, semigroups, and the characterization of their holding and jump times. We then develop the theory of exchangeable random partitions. In this context, the paintbox construction is introduced as a fundamental tool providing the most general representation of exchangeable partitions. Furthermore, as an important example, we study the Poisson-Dirichlet partition, which can be viewed from several different perspectives, including the stick-breaking construction, the Chinese restaurant process, and a construction based on a Poisson random measure.

Finally, we demonstrate that Kingman's coalescent arises as the universal limiting ancestral process for a broad class of classical population models, such as the Wright-Fisher and the Moran model. Additionally, we explore the duality between this model and a forward-in-time diffusion process, and discuss its connection to mutation models, including the infinite alleles and infinite sites model.

Acknowledgements

First of all, I would like to thank my supervisor, Prof. Dr. Jean Bertoin, for his guidance, patience, and support throughout this thesis, and also for guiding my interest in probability theory through his Probability I and II courses.

My sincere gratitude goes to my partner, friends and family, especially to my parents, for their continuous support and encouraging words whenever I needed them most.

Lastly, a big thanks goes to my university classmates, Philine, Clara and Hannah, with whom I have shared countless wonderful memories, and without whom I would not be where I am today.

Grazie a tutti!

Contents

List of Figures **IX**

Introduction **1**

1 Markov processes **3**
1.1 Discrete-time Markov chains 3
1.2 Continuous-time Markov chains 7
1.2.1 Poisson counting processes, Birth-death processes 14
1.3 General Markov processes 15

2 Exchangeable sequences **19**
2.1 Exchangeability and Independence 19
2.2 Concept of mixing 21
2.3 De Finetti's Theorem 23

3 Partitions **25**
3.1 Definitions and basic concepts 25
3.1.1 Size-biased sampling 36
3.2 Poisson-Dirichlet random partitions 40
3.2.1 $\alpha = 0,\ \theta > 0$ 40
3.2.2 $0 < \alpha < 1, \theta = 0$ 46
3.2.3 Poisson random measures and Poisson-Dirichlet laws 47
3.2.4 Example 59

4 Kingman's coalescent **63**
4.1 Introduction to the model 63
4.2 Aldous' construction 70
4.3 Genealogy of populations 77
4.4.1 Moran model 77
4.4.2 Wright-Fisher model 80
4.4.3 Möhle's Lemma 82
4.4.4 Wright-Fisher diffusion and duality 84
4.5 Infinite alleles model 90
4.6 Infinite sites model 97

Outlook **103**

Appendices **105**

A Theoretical background **107**
A.1 Distributions 107
A.2 Results from probability theory 110
A.3 Change of meassure 112
A.4 Analytical results 112
A.5 Kolmogorov's Extension Theorem 113

B Basic biological terminology **115**

Bibliography **117**

List of symbols **121**

List of Figures

1.1 Sample path of a ctMC. 12

3.1 A realization of a bond percolation on $\mathbb{Z}^2$. 28
3.2 Paintbox partition $\Pi_{|[8]}$. 31
3.3 Point process on $\mathbb{R}^2$ with $A \subset \mathbb{R}^2$. 48

4.1 Realization of Kingman's n-coalescent ($n = 8$). 65
4.2 Aldous' construction. Source: [5]. 70
4.3 Cutting a rooted segment. Source: [5]. 75
4.4 A simple visualization of the Moran model with two alleles. 78
4.5 Ancestral partition process associated with the Moran model. Source: [5]. 79
4.6 Ancestral partition process associated with the Wright-Fisher model. 81
4.7 Allelic partition generated by mutations for feather color. 91
4.8 Forest of subtrees. 93

B.1 Schematic representation of DNA base pairing. 115

Introduction

Population models such as the Wright-Fisher, Moran, and more general Cannings models describe the evolution of populations forward in time. However, tracing the ancestral history of a sample of individuals within these models becomes increasingly difficult as the population size grows. To address this issue, the British mathematician J.F.C. Kingman introduced a powerful stochastic model in 1982, known as Kingman's coalescent ([21],[20]). This model describes the genealogy of a sample as a continuous-time Markov chain on the space of partitions of $\mathbb{N}$.

The aim of this master's thesis is to provide a comprehensive and self-contained introduction to Kingman's coalescent. We develop the probabilistic framework underlying its construction and examine its main properties, as well as its connections to classical population models and more general Markov processes. The content is mainly based on Chapter 1 and Chapter 2 of the published lecture notes *Recent Progress in Coalescent Theory* by Nathanaël Berestycki [5]. Unless explicitly stated otherwise, the material in this thesis is primarily based on [5].

The structure of the thesis is as follows. Chapter 1 reviews the fundamental theory of discrete-time Markov chains on discrete state spaces. We then turn to continuous-time Markov chains and introduce key concepts such as infinitesimal generators, transition semigroups, holding times, and jump chains. After extending these ideas to continuous state spaces, the chapter concludes with a brief introduction to diffusion processes and stochastic differential equations, which are important for the subsequent analysis.

Chapter 2 explores the notion of exchangeability and its relation to independence. In this context, we introduce random measures and present de Finetti's Theorem, which states that, conditional on a directing random measure, the random variables of an infinite exchangeable sequence are independent and identically distributed.

The first part of Chapter 3 provides an introduction to the theory of random partitions, with particular emphasis on exchangeability. We present the paintbox construction as a canonical method for generating exchangeable partitions and as a key ingredient in Kingman's Representation Theorem. The second part studies several constructions of the Poisson-Dirichlet partition, a fundamental example of an exchangeable random partition. One of these approaches is based on Poisson random measures, whose basic theory is also introduced. In connection with Poisson-Dirichlet partitions, we naturally arrive at Ewens' sampling formula, which describes their probability distribution.

Finally, Chapter 4 is devoted to Kingman's coalescent. We present an explicit construction of the model, which is a multidimensional version of the paintbox construction. We then demonstrate how the coalescent arises as a universal limiting object for a broad class of population models under an appropriate time-scaling. Furthermore, we discuss its duality with the Wright-Fisher diffusion and extend the framework to include mutation models, specifically under the infinite alleles and infinite sites assumptions. Within the context of the infinite alleles model, the allelic partition follows a Poisson-Dirichlet distribution, leading us to a reformulation of Ewens' sampling formula.

Chapter 1

Markov processes

This chapter establishes the theoretical foundations of Markov processes. We begin by formalizing discrete-time Markov chains on discrete state spaces before extending these concepts to the continuous-time setting. Furthermore, we introduce the broader class of Markov processes that operate on continuous or more general state spaces, providing the necessary tools for the more complex models discussed in later chapters.

General setting. Throughout this thesis, we work on a filtered probability space

$$(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P}),$$

where Ω is the sample space, $\mathcal{F}$ is a σ-algebra on Ω, $\mathbb{P}$ is a probability measure, and $(\mathcal{F}_n)_{n\geq 0}$ is the canonical filtration generated by the sequence $(X_n)_{n\geq 0}$, defined by

$$\mathcal{F}_n \coloneqq \sigma(X_s : s \leq n).$$

This is the smallest filtration that makes the stochastic process $(X_n)_{n\geq 0}$ adapted.

1.1 Discrete-time Markov chains

Discrete-time Markov chains on a discrete (finite or countable) state space S form the simplest class of random processes satisfying the Markov property. This property can be stated as follows: Given a stochastic process $(X_n)_{n\in\mathbb{N}_0}$, the past $(X_i)_{i\leq n}$ and the future $(X_i)_{i\geq n}$ are conditionally independent given the present state X_n. [1]

Definition 1.1 (Stopping time)**.** A random variable τ with values in $\mathbb{N}_0 \cup \{\infty\}$ is called a stopping time if, for every $n \geq 0$, the event $\{\tau = n\}$ (equivalently $\{\tau \leq n\}$, $\{\tau \geq n+1\}$) is $\mathcal{F}_n$-measurable. For τ a stopping time, we can define a σ-algebra as follows:

$$\mathcal{F}_\tau \coloneqq \{\Lambda \in \mathcal{F} \mid \{\tau = n\} \cap \Lambda \in \mathcal{F}_n \text{ for all } n \geq 0\}.$$

[1] Unless explicitly cited otherwise, the material in this chapter is based on the Probability I and II lecture courses by Prof. Dr. Jean Bertoin.

Definition 1.2 (Discrete-time Markov chain)**.** A random process $X = (X_n)_{n\geq 0}$ with values in a discrete state space S is called a discrete-time Markov chain if

$$\mathbb{P}(X_{n+1} = s_{n+1} \mid X_0 = s_0, \ldots, X_n = s_n) = \mathbb{P}(X_{n+1} = s_{n+1} \mid X_n = s_n), \qquad (1.1)$$

for all $n \geq 0$ and for all states $s_0, s_1, \ldots, s_n, s_{n+1} \in S$ such that $\mathbb{P}(X_0 = s_0, \ldots, X_n = s_n) > 0$. Equation (1.1) is known as the Markov property.

Remark 1.3. In other words, this means that to predict the state X_{n+1}, the current state X_n provides all the necessary information, and knowing the previous states $X_0, \ldots, X_{n-1}$ does not provide any additional information.

Definition 1.4 (Transition probability matrix)**.** A matrix $P = (p(x, y))_{x,y\in S}$ is a transition probability matrix if $p(x, y) \geq 0\ \forall x, y \in S$ and $\sum_{y\in S} p(x, y) = 1\ \forall x \in S$.

The quantity $p(x, y)$ can be interpreted as the probability that, given the chain is in state x, its next state is y.

Definition 1.5 (Time-homogeneous Markov chain)**.** A Markov chain $X = (X_n)_{n\geq 0}$ with transition matrix $P = (p(x, y))_{x,y\in S}$ is time-homogeneous if

$$\mathbb{P}(X_0 = x_0, \ldots, X_n = s_n \mid X_0 = x_0) = p(x_0, s_1)p(s_1, s_2) \cdots p(s_{n-1}, s_n),$$

for all $n \geq 0$ and for all states $x_0, s_1, \ldots, s_n \in S$. The left-hand side can also be written as $\mathbb{P}_{x_0}(X_1 = s_1, \ldots, X_n = s_n)$, where $\mathbb{P}_{x_0}(\cdot)$ is called the law of X started from x_0.

More generally, let μ be a probability measure on the state space S. We can define $\mathbb{P}_\mu$ as the law of the chain X started from $X_0 \sim \mu$, which is called the initial distribution. Hence,

$$\mathbb{P}_\mu(X_0 = y) = \mu(y)$$

for $y \in S$. In this case, the equation (1.5) can be written as

$$\mathbb{P}_\mu(X_0 = x_0, \ldots, X_n = s_n) = \mu(x_0)p(x_0, s_1)p(s_1, s_2) \cdots p(s_{n-1}, s_n).$$

Proposition 1.6. *Let $f : S \to \mathbb{R}$ be a non-negative or bounded function. Let $(X_n)_{n\geq 0}$ be a discrete-time Markov chain on S. Then, for all $n \geq 1$ and for every $x \in S$, we have*

$$\mathbb{E}_x[f(X_n)] = \sum_{y\in S} f(y)p^n(x, y) \equiv P^n f(x),$$

where $p^n(x, y) = \mathbb{P}_x(X_n = y)$.

Proposition 1.7 (Simple Markov property)**.** *Let $f : S \to \mathbb{R}$ be a non-negative or bounded function. Let $(X_n)_{n\geq 0}$ be a discrete-time Markov chain on S. Then, for all $n, k \geq 1$ and for every $x \in S$, we have*

$$\mathbb{E}_x[f(X_{n+k}) \mid \mathcal{F}_k] = P^n f(X_k).$$

The strong Markov property extends the standard Markov property, asserting that the memoryless nature of the process holds not just at fixed, deterministic times, but also at random stopping times. To state this formally, we must first introduce the shift operator.

Definition 1.8 (k-shift operator)**.** The k-shift operator for $k \in \mathbb{N}_0$ is defined as

$$\theta_k : \Omega \to \Omega, \qquad (\omega_0, \omega_1, \ldots) \mapsto (\omega_k, \omega_{k+1}, \ldots).$$

Theorem 1.9 (Strong Markov property)**.** *Assume τ is a stopping time with respect to the filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ such that $\mathbb{P}_\mu(\tau < \infty) > 0$. Then, for every measurable $G : S^{\mathbb{N}} \to \mathbb{R}_+$*

$$\mathbb{E}_\mu[G(X \circ \theta_\tau)\mathbb{1}\{\tau < \infty\} \mid \mathcal{F}_\tau] = \phi(X_\tau),$$

where $\phi(x) = \mathbb{E}_x[G(X)]$ and $X \circ \theta_\tau = (X_\tau, X_{\tau+1}, \ldots)$.

Definition 1.10 (Recurrent and transient states)**.** Let $(X_n)_{n\geq 0}$ be a Markov chain on a discrete state space S. For $x \in S$, define the first return time to x by

$$\tau_x \coloneqq \inf\{n \geq 1 : X_n = x\},$$

with the convention $\inf(\emptyset) = \infty$. The state x is called

(i) recurrent if $\mathbb{P}_x(\tau_x < \infty) = 1$,

(ii) transient if $\mathbb{P}_x(\tau_x < \infty) > 0$.

Definition 1.11. Let $(X_n)_{n\geq 0}$ be a Markov chain on a discrete state space S and let x, y be in S with $x \neq y$. We say that

(i) x leads to y (denoted $x \to y$) if $\mathbb{P}_x(\tau_y < \infty) > 0$,

(ii) x communicates with y (denoted by $x \leftrightarrow y$) if $x \to y$ and $y \to x$.

Definition 1.12 (Communication class, [32])**.** The relation $\leftrightarrow$ satisfies the condition for an equivalence relation on a discrete state space S. Consequently, we can partition S into communication classes. A communication class $C \subseteq S$ is closed if

$$i \in C, i \to j \quad \Rightarrow \quad j \in C.$$

Remark 1.13 ([32])**.** Let C be a communication class. Then all states in C are either recurrent or transient.

Definition 1.14 (Absorbing state, [32])**.** A state $i \in S$ is called absorbing if $\{i\}$ is a closed communication class.

Definition 1.15 (Irreducible)**.** We say that a Markov chain $(X_n)_{n\geq 0}$ is irreducible if $x \leftrightarrow y$ for all x, $y \in S$.

Corollary 1.16. *Let $(X_n)_{n\geq 0}$ be an irreducible Markov chain with state space S. Then either all states in S are recurrent or all states are transient. Accordingly, the chain $(X_n)_{n\geq 0}$ is called recurrent or transient. Moreover, if S is finite and $(X_n)_{n\geq 0}$ is irreducible, then $(X_n)_{n\geq 0}$ is recurrent.*

We say that $(\mu(x))_{x\in S}$ with $\mu(x) \geq 0$ is a *measure* on the state space S. If additionally, we have the condition $\sum_{x\in S} \mu(x) = 1$, then μ is called a *probability measure.*

For any measure μ on S and transition matrix P, we write μP for the measure

$$\mu P(y) \coloneqq \sum_{x\in S} \mu(x)p(x,y), \quad \text{for all } y \in S.$$

Definition 1.17 (Invariant measure, stationary distribution)**.** A measure μ on S is called invariant (w.r.t. P) if and only if $\mu P(y) = \mu(y)$ for all $y \in S$. An invariant probability measure is called a stationary distribution.

Remark 1.18. If μ is a stationary distribution, then

$$\mu P^n = (\mu P)P^{n-1} = \cdots = \mu \quad \text{for all } n \geq 1.$$

That is, if the Markov chain is initialized with a stationary distribution μ, then the distribution of X_n under $\mathbb{P}_\mu$ is μ for every $n \geq 1$.

Definition 1.19 (Reversible measure)**.** A measure μ is a reversible measure w.r.t. P if and only if for all $x, y \in S$

$$\mu(x)p(x,y) = p(y,x)\mu(y).$$

This means that, if we have a reversible probability measure μ, then, under $\mathbb{P}_\mu$ and for any $n \geq 1$, we have

$$(X_0, \ldots, X_n) \overset{d}{=} (X_n, \ldots, X_0).$$

Hence, the initial process is indistinguishable from its time-reversal.

Lemma 1.20. *If μ is reversible w.r.t. P, then μ is invariant.*

Theorem 1.21. *Let $(X_n)_{n\geq 0}$ be an irreducible recurrent Markov chain on a state space S. Then exactly one of the following holds:*

(i) There exists a stationary distribution υ. Moreover, υ is unique and satisfies $\upsilon(x) > 0$ for all $x \in S$. In this case, the chain is called positive recurrent.

(ii) There exists no stationary distribution. Equivalently, $\mathbb{E}_x[\tau_x] = \infty$ for all $x \in S$. In this case, the chain is called null recurrent.

As already mentioned, for Markov chains the future and the past are independent given the present. The symmetry of this property implies that every Markov chain can be studied with time running backward.

Theorem 1.22 ([35]). *Let $(X_n)_{0\le n\le m}$ for a positive integer m be a time-homogeneous Markov chain with transition matrix P on a countable state space S. Define the time-reversed process by*

$$\widehat{X}_n := X_{m-n}, \qquad n = 0, 1, \dots, m.$$

Then

$$\widehat{X} = (\widehat{X}_0, \widehat{X}_1, \dots, \widehat{X}_m) = (X_m, X_{m-1}, \dots, X_0). \tag{1.2}$$

The reversed process is again a Markov chain with one-step transition probabilities at time $n \in \{0, 1, \dots, m-1\}$ given by

$$\mathbb{P}\Big(\widehat{X}_{n+1} = y \mid \widehat{X}_n = x\Big) = \frac{\mathbb{P}(X_{m-n-1} = y)}{\mathbb{P}(X_{m-n} = x)}\, p(y, x), \qquad (x, y) \in S^2.$$

Under the above assumptions, the time-reversed process is again a Markov chain, but in general it is not time-homogeneous. This changes when the original process is started in a stationary distribution.

Theorem 1.23 ([35]). *Let $(X_n)_{0\le n\le m}$ be irreducible and positive recurrent time-homogeneous Markov chain on a countable state space S with invariant distribution μ. Then, the reversed process defined as in (1.2) is again a time-homogeneous Markov chain with invariant distribution μ and with transition probabilities given by*

$$\widehat{p}(x, y) = \frac{\mu(y)}{\mu(x)}\, p(y, x), \qquad (x, y) \in S^2.$$

1.2 Continuous-time Markov chains

As discussed in the previous section, discrete-time Markov chains evolve at fixed time steps. In many real-world systems, however, state changes occur whenever an event happens rather than at predetermined times. Such dynamics are naturally modeled by continuous-time Markov chains.
In this section, we again consider a discrete state space S.

Definition 1.24 (Continuous-time stochastic process, [11]). A family $X := (X_t)_{t\ge 0}$ of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in a state space S is called a stochastic process. The functions $t \mapsto X_t(\omega)$ for $\omega \in \Omega$ are called sample paths of the stochastic process.

Definition 1.25 ([11])**.** Let X be a stochastic process.

(i) X is called right-continuous (resp. left-continuous) if a.s. $\lim_{u \downarrow t} X_u(\omega) = X_t(\omega)$ (resp. $\lim_{u \uparrow t} X_u(\omega) = X_t(\omega)$).

(ii) X is said to have left limits (resp. right limits) if a.s. $\lim_{s \uparrow t} X_s(\omega)$ (resp. $\lim_{s \downarrow t} X_s(\omega)$) exists.

(iii) X is càdlàg (French: continue à droite, limites à gauche) if it is right-continuous with left limits.

(iv) X is continuous if it is left- and right-continuous.

Definition 1.26 (Skorokhod space, [8])**.** The Skorokhod space, denoted by $D(\mathbb{R}_+, S)$, is the set of all càdlàg functions mapping from $\mathbb{R}_+$ to the state space S.

Definition 1.27 (Continuous-time Markov chain, [10])**.** A continuous-time stochastic process $(X_t)_{t \geq 0}$ with values in S is called a continuous-time Markov chain (ctMC) if for all states $x, y, i_1, \ldots, i_k \in S$ and any sequence $0 \leq t_0 < t_1 < \cdots < t_n < t$ of times, one has

$$\mathbb{P}(X_t = y \mid X_{t_n} = x, X_{t_{n-1}} = i_1, \ldots, X_{t_0} = i_k) = \mathbb{P}(X_t = y \mid X_{t_n} = x). \tag{1.3}$$

This equation is the Markov property in continuous time.

Definition 1.28 (Time-homogeneous ctMC, [10])**.** If the right-hand side of (1.3) does not depend on the starting time, i.e.,

$$\mathbb{P}(X_{t+s} = y \mid X_s = x) = \mathbb{P}(X_t = y \mid X_0 = x),$$

the ctMC is time-homogeneous.

Assumption. From this point on, we will only consider time-homogeneous ctMCs.

In the continuous-time setting, we introduce an object that plays the same role as the transition matrix in the discrete-time case. This object is the so-called *transition semigroup*, which is defined as follows:

Definition 1.29 (Transition semigroup, [10])**.** Let $(X_t)_{t \geq 0}$ be a ctMC with state space S. For each $t \geq 0$, define the transition matrix

$$P_t := \big(p_t(x, y)\big)_{x, y \in S},$$

where

$$p_t(x, y) := \mathbb{P}(X_t = y \mid X_0 = x).$$

Here, $p_t(x, y)$ is called the transition probability from state x to state y in time t. The family $(P_t)_{t \geq 0}$ is called the transition semigroup of the ctMC.

Remark 1.30 ([32])**.** For every $t \geq 0$ and every fixed $x \in S$, the function $p_t(x, \cdot)$ defines a probability measure on S, since $p_t(x, y) = \mathbb{P}(X_t = y \mid X_0 = x) \geq 0$ and $\sum_{y \in S} p_t(x, y) = 1$.

Proposition 1.31 (Chapman-Kolmogorov equation, [10])**.** *Let $(X_t)_{t \geq 0}$ be a ctMC with transition semigroup $(P_t)_{t \geq 0}$. Then, for all $s, t \geq 0$ and for all $x, y \in S$, we have*

$$p_{t+s}(x, y) = \sum_{k \in S} p_t(x, k) p_s(k, y).$$

Equivalently, in matrix form,

$$P_{t+s} = P_t P_s.$$

Moreover,

$$P_0 = I,$$

where I denotes the identity matrix on S, given by

$$p_0(x, y) = \begin{cases} 1, & x = y, \\ 0, & x \neq y. \end{cases}$$

Proposition 1.32 ([35])**.** *Let $(X_t)_{t \geq 0}$ be a ctMC with transition semigroup $(P_t)_{t \geq 0}$.*

(i) Action on functions: Let $f : S \to \mathbb{R}$ be a bounded or non-negative function. Then, for all $t \geq 0$ and all $x \in S$,

$$P_t f(x) \coloneqq \sum_{y \in S} p_t(x, y) f(y) = \mathbb{E}_x[f(X_t)].$$

(ii) Action on measures: Let μ be a probability measure on S. For any $A \subseteq S$, the measure μP_t is defined as

$$(\mu P_t)(A) \coloneqq \sum_{x \in S} \mu(x) p_t(x, A).$$

The measure μP_t represents the distribution of X_t given that the initial distribution of X_0 is μ.

So far, we have seen that a ctMC can be characterized by its transition probabilities. However, there are two other ways of defining such a process: First, through the infinitesimal generator Q, which describes the local rate of change; and second, via the jump chain and the associated holding times, which provide a pathwise construction of the process by separating the sequence of states from the duration spent in each state.

We suppose, following [10], that the semigroup of our ctMC $(X_t)_{t \geq 0}$ is continuous at the origin, that is,

$$\lim_{t \downarrow 0} P_t = P_0 = I.$$

This means that, at any time $h \geq 0$ and for all states $x, y \in S$,

$$\lim_{t \to 0} p_{t+h}(x, y) = p_h(x, y).$$

Definition 1.33 (Infinitesimal generator, [10],[15])**.** Let $(X_t)_{t\geq 0}$ be a ctMC with transition semigroup $(P_t)_{t\geq 0}$. The matrix

$$Q = \lim_{t\downarrow 0} \frac{P_t - I}{t} \tag{1.4}$$

is called the infinitesimal generator (or generator matrix) of the semigroup or of the ctMC. The entries of the matrix are denoted by $q(x, y)$ for $x, y \in S$. By acting on a Borel-measurable function f on the state space S, the generator Q is defined as the operator

$$(Qf)(x) = \lim_{t\downarrow 0} \frac{P_t f(x) - f(x)}{t} = \lim_{t\downarrow 0} \frac{\mathbb{E}_x[f(X_t)] - f(x)}{t}, \tag{1.5}$$

for x in S. We define $\mathcal{D}(Q) \subseteq \mathcal{B}(S)$ as the space of functions for which the limit exists.

Proposition 1.34 ([32])**.** *The infinitesimal generator Q, as defined in Definition 1.33, satisfies the following properties:*

(i) $0 \leq -q(x, x) < \infty$ *for all* $x \in S$*;*

(ii) $q(x, y) \geq 0$ *for all* $x \neq y$*;*

(iii) $\sum_{y\in S} q(x, y) = 0$ *for all* $x \in S$*.*

For each $x \in S$, assume that $q(x) := \sum_{y\neq x} q(x, y) < \infty$. Since each row of Q sums to zero, the diagonal entries satisfy $q(x, x) = -q(x)$.

Interpretation of the entries [32]: Each entry $q(x, y)$ with $x \neq y$ is the rate of going from x to y. The value $q(x) = -q(x, x)$ is the overall rate of leaving the state x.

Remark 1.35. For a process on a finite state space S, the action of the generator Q on a function f at state $x \in S$ can be naturally expressed in terms of the transition rates. We have

$$(Qf)(x) = \sum_{\substack{y\in S,\\ y\neq x}} q(x, y)\big(f(y) - f(x)\big). \tag{1.6}$$

Proof.

$$\begin{aligned}
(\mathbf{Q}f)(x) &= \lim_{t\downarrow 0} \frac{\mathbb{E}_x[f(X_t)] - f(x)}{t} \\
&= \lim_{t\downarrow 0} \frac{\sum_y f(y)p_t(x, y) - f(x)}{t} \\
&= \lim_{t\downarrow 0} \frac{f(x)p_t(x, x) + \sum_{y\neq x} f(y)p_t(x, y) - f(x)}{t} \\
&= \lim_{t\downarrow 0} \frac{f(x)(p_t(x, x) - 1)}{t} + \lim_{t\downarrow 0} \sum_{y\neq x} f(y)\frac{p_t(x, y)}{t}
\end{aligned}$$

$$
\begin{aligned}
&= f(x)q(x,x) + \sum_{y \neq x} f(y)q(x,y) \\
&= f(x)\Big(- \sum_{y \neq x} q(x,y)\Big) + \sum_{y \neq x} f(y)q(x,y) \\
&= \sum_{y \neq x} q(x,y)(f(y) - f(x)).
\end{aligned}
$$

■

Assumption. To prevent irregular behavior and ensure a mathematically rigorous decomposition into a jump chain and holding times, we will restrict our focus to right-continuous ctMCs. Since we are working on a discrete state space S, right-continuity can be rephrased as follows [32]: For all $\omega \in \Omega$ and $t \geq 0$ there exists $\epsilon > 0$ such that $X_s(\omega) = X_t(\omega)$ for $s \in [t, t+\epsilon]$.

Definition 1.36 (Jump and Holding times, Jump chain, [32])**.** Let $(X_t)_{t\geq 0}$ be a right-continuous ctMC.

(i) **Jump times:** For $n \in \mathbb{N}_0$, J_n is the time of the n-th jump. The sequence of jump times is defined recursively by

$$J_0 := 0, \quad J_{n+1} := \inf\{t > J_n : X_t \neq X_{J_n}\}$$

with the convention $\inf(\emptyset) = \infty$.

(ii) **Holding times:** For $n \in \mathbb{N}$, the holding time T_n is the duration the ctMC stays in a state before jumping for the n-th time. It can be calculated from the jump times by

$$T_n := \begin{cases} J_n - J_{n-1}, & \text{if } J_{n-1} < \infty, \\ \infty, & \text{otherwise.} \end{cases}$$

Note that right-continuity forces $T_n > 0$ for all n. For $J_n = \infty$, we define $X_\infty := X_{J_{n-1}}$.

(iii) **Jump chain:** The discrete-time process $(Y_n)_{n\in\mathbb{N}_0}$ given by $Y_n := X_{J_n}$ is called the jump chain or jump process of $(X_t)_{t\geq 0}$. This is simply the sequence of values taken by $(X_t)_{t\geq 0}$.

Remark 1.37. For a ctMC on a discrete state space, right-continuity is sufficient to guarantee the existence of left limits at all $t > 0$. At any jump time J_n, the left limit corresponds to the state held by the process up to the moment of the jump, i.e.,

$$\lim_{t \uparrow J_n} X_t = X_{J_{n-1}}.$$

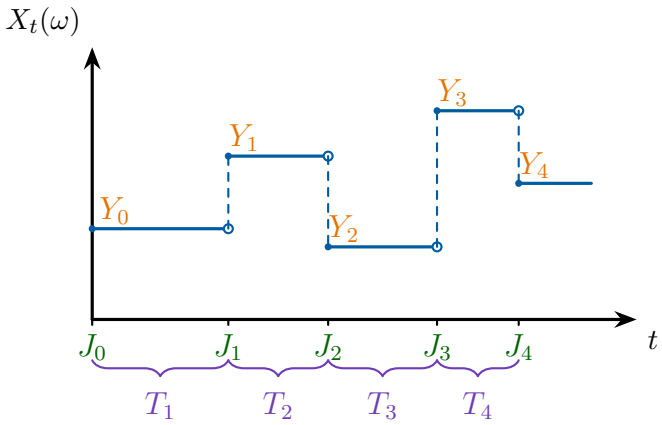


Figure 1.1: Sample path of a ctMC with jump times J_n, holding times T_n, and the jump chain $(Y_n)_{n\in\mathbb{N}_0}$.

As for the discrete case, we also have the strong Markov property in the continuous case. For this, we need to introduce stopping times in a continuous time setting:

Definition 1.38 (Stopping time, [32])**.** A random variable τ with values in $[0,\infty]$ is called a stopping time for a right-continuous ctMC $(X_t)_{t\geq 0}$ if, for each $t \in [0,\infty)$, the event $\{\tau \leq t\}$ depends only on $(X_s : s \leq t)$.

Remark 1.39 ([32])**.** The jump times are stopping times of the ctMC $(X_t)_{t\geq 0}$.

Proof. For a proof see [32], Lemma 6.5.2 on page 226. ■

Theorem 1.40 (Strong Markov property, [32])**.** *Let $X = (X_t)_{t\geq 0}$ be a right-continuous ctMC on a discrete state space S with initial distribution μ and generator matrix Q. Let τ be a stopping time of X. Conditional on $\tau < \infty$ and on $X_\tau = x$, $(X_{\tau+t})_{t\geq 0}$ is a ctMC with initial distribution δ_x and generator matrix Q independent of $(X_s : s \leq \tau)$.*

Theorem 1.41 ([12])**.** *Let $(X_t)_{t\geq 0}$ be a ctMC on a discrete state space S with initial distribution μ and generator matrix Q. Let J_n, T_n and Y_n be as in Definition 1.36. Then, we have*

(i) $(Y_n)_{n\in\mathbb{N}_0}$ is a discrete-time Markov chain with initial distribution μ and transition matrix (also called jump matrix) $J = (j(x,y))_{x,y\in S}$ given by

$$j(x,y) := \begin{cases} \dfrac{q(x,y)}{q(x)}, & \text{if } y \neq x \text{ and } q(x) \neq 0, \\ 0, & \text{if } y = x \text{ and } q(x) \neq 0, \\ \delta_{x,y}, & q(x) = 0. \end{cases}$$

(ii) Under $\mathbb{P}_\mu$, conditional on the jump chain $(Y_n)_{n\in\mathbb{N}_0}$, the holding times $(T_n)_{n\in\mathbb{N}}$ are independent and T_n has an exponential distribution with parameter $q(Y_{n-1})$.

Remark 1.42 ([32])**.** The dynamics of a right-continuous ctMC can be summarized intuitively: Whenever the chain arrives at a state x, it waits there for an exponentially distributed time with parameter $q(x)$, and then jumps to a new state y chosen according to the probability $j(x, y)$.

Remark 1.43 ([32])**.** The exponential distribution of the holding times naturally arises from the Markov property. When predicting the future of a ctMC, the amount of time the process has already spent in its current state is irrelevant. Consequently, the holding times must be memoryless. The exponential distribution is the unique continuous distribution that satisfies this *memoryless property*, meaning:

$$\mathbb{P}(T > s + t \mid T > s) = \mathbb{P}(T > t)$$

for all $s, t \geq 0$. Furthermore, the specific parameter of this exponential holding time can be understood through a system of "competing exponential clocks". If we assign an independent clock $E_y \sim \text{Exp}(q(x, y))$ to every possible next state $y \neq x$, the overall holding time T is simply the time until the first clock rings:

$$T = \min_{y \neq x} E_y.$$

And by Theorem A.5, $T \sim \text{Exp}\big(\sum_{y \neq x} q(x, y)\big) = \text{Exp}(q(x))$.

The classification of states for a continuous-time Markov chain, including irreducibility, transience, and recurrence, is determined by its associated discrete-time jump chain [32].

Definition 1.44 (Irreducibility, [32])**.** We say that a ctMC with transition semigroup $(P_t)_{t \geq 0}$ is irreducible if for all $i, j \in S$, there exists $t > 0$ such that $p_t(i, j) > 0$.

Definition 1.45 (Stationary distribution, [10])**.** A stationary distribution of the transition semigroup $(P_t)_{t \geq 0}$ is any probability distribution π on S such that for all $t \geq 0$, we have

$$\pi P_t = \pi.$$

If a ctMC has a stationary distribution, then, if X_0 is distributed according to π, then so is X_t for all $t \geq 0$.

We can find the stationary distribution directly with the generator matrix of the ctMC.

Theorem 1.46 ([32])**.** *Let $(X_t)_{t \geq 0}$ be a ctMC with generator matrix Q. A probability distribution π is stationary if and only if*

$$\pi Q = 0.$$

1.2.1 Poisson counting processes, Birth-death processes

The *Poisson counting process* is one of the most fundamental examples of a continuous-time Markov chain. As a counting process, it is widely used to model the cumulative number of random events occurring up to a given time t, such as the arrival of buses at a station or the number of incoming calls at an exchange. The process is formally defined as follows:

Definition 1.47 (Poisson counting process, [23],[12])**.** A family $(N_t)_{t\geq 0}$ of $\mathbb{N}_0$-valued random variables on $[0,\infty]$ is called a Poisson counting process with intensity $\lambda \geq 0$ if the following properties hold:

(i) $N_0 = 0$.

(ii) The process has independent increments. This means that, for every $n \in \mathbb{N}$ and every sequence of times $0 = t_0 < t_1 < \cdots < t_n$, the family $(N_{t_i} - N_{t_{i-1}},\, i = 1,\ldots,n)$ is independent.

(iii) For all $t > s \geq 0$, the increment $N_t - N_s$ is Poisson distributed with parameter $\lambda(t-s)$, that is,

$$\mathbb{P}(\, N_t - N_s = k\,) = e^{-\lambda(t-s)}\,\frac{(\lambda(t-s))^k}{k!}, \qquad k \in \mathbb{N}_0.$$

(iv) The mapping $t \mapsto N_t$ is right-continuous with left limits.

(v) If we denote by N_{t-} the left limit of the counting process at time t, then $N_t - N_{t-} \in \{0,1\}$ for all $t > 0$.

The infinitesimal generator $Q = (q(x,y))_{x,y\in\mathbb{N}_0}$ of the Poisson counting process is given by

$$q(x,y) = \begin{cases} \lambda, & \text{if } y = x+1, \\ -\lambda, & \text{if } y = x, \\ 0, & \text{if } y \notin \{x,\, x+1\}. \end{cases}$$

Another crucial continuous-time Markov chain for our purposes is the *birth-death process*, defined as follows:

Definition 1.48 (Birth-death process, [12])**.** A birth-death process $(X_t)_{t\geq 0}$ is a ctMC whose jump steps are equal to ± 1. The generator matrix $Q = (q(x,y))_{x,y\in\mathbb{N}_0}$ for all $x \geq 1$ is given by

$$q(x,y) = \begin{cases} \lambda_x, & \text{if } y = x+1, \\ \mu_x, & \text{if } y = x-1, \\ -(\lambda_x + \mu_x), & \text{if } y = x, \\ 0, & \text{otherwise}, \end{cases}$$

where $(\lambda_x)_{x\in\mathbb{N}\setminus\{0\}}$ and $(\mu_x)_{x\in\mathbb{N}\setminus\{0\}}$ are sequences of positive real numbers and $\lambda_0 > 0$, $\mu_0 = 0$. We say that $(X_t)_{t\geq 0}$ is a pure death process if $\lambda_x = 0$ for all $x \geq 0$. We say that $(X_t)_{t\geq 0}$ is a pure birth process if $\mu_x = 0$ for all $x \geq 1$.

1.3 General Markov processes

While Chapters 1.1 and 1.2 focused on processes with discrete state spaces, the following sections extend these concepts to a more general framework. We now introduce Markov processes in continuous time with more general state spaces, specifically focusing on Brownian motion, stochastic differential equations (SDEs), and diffusion processes. To maintain the scope of this work, we will not provide a treatment of Itô calculus, but rather focus on the properties and definitions necessary for our application in later chapters.

Definition 1.49 (General Markov process, [16],[4])**.** Let $(\Omega, \mathcal{F})$ be a probability space and let $(S, \mathcal{S})$ be a measurable space. A Markov process $(X_t, \mathbb{P}_x)$ is a stochastic process and a family of probability measures $(\mathbb{P}_x)_{x\in S}$ on $(\Omega, \mathcal{F})$ satisfying the following conditions:

(i) For every $t \geq 0$, X_t is $\mathcal{F}_t$-measurable.

(ii) For every $t \geq 0$ and every $A \in \mathcal{S}$, the map $x \mapsto \mathbb{P}_x(X_t \in A)$ is Borel-measurable.

(iii) For every $s, t \geq 0$, each $A \in \mathcal{S}$, and every $x \in S$, we have the Markov property

$$\mathbb{P}_x(X_{s+t} \in A \mid \mathcal{F}_s) = \mathbb{P}_{X_s}(X_t \in A), \quad \mathbb{P}_x \text{ - a.s.}$$

Equivalently,

$$\mathbb{E}_x[f(X_{s+t}) \mid \mathcal{F}_s] = \mathbb{E}_{X_s}[f(X_t)] \quad \mathbb{P}_x \text{ - a.s.}$$

for all bounded Borel-measurable $f : S \to \mathbb{R}$.

Definition 1.50 (Transition probabilities, [4])**.** Let $(X_t, \mathbb{P}_x)$ be a Markov process with an associated transition semigroup $(P_t)_{t\geq 0}$. The transition probabilities $p_t(x, A)$ are defined for all $t \geq 0$, $x \in S$, and $A \in \mathcal{S}$ as

$$p_t(x, A) = \mathbb{P}_x(X_t \in A),$$

and satisfy the following properties:

(i) For all $t \geq 0$ and $x \in S$ $p_t(x, S) = 1$.

(ii) For all $s, t \geq 0$, $x \in S$, and $A \in \mathcal{S}$, the transition probabilities satisfy the Chapman-Kolmogorov equation

$$p_{t+s}(x, A) = \int_S p_t(y, A)\, p_s(x, dy).$$

For any Borel-measurable function $f : S \to \mathbb{R}$ that is either bounded or non-negative, we define

$$P_t f(x) = \mathbb{E}_x[f(X_t)] = \int_S f(y)\, p_t(x, dy).$$

Remark 1.51. The generator for general Markov processes, call it $\mathcal{L}$, is analogous to the infinitesimal generator Q defined for time-homogeneous ctMCs and can also be defined to be the operator

$$(\mathcal{L}f)(x) = \lim_{t \downarrow 0} \frac{\mathbb{E}_x[f(X_t)] - f(x)}{t}. \tag{1.7}$$

$\mathcal{D}(\mathcal{L})$ is the domain of functions for which such a limit exists.

Brownian motion, also known as the *Wiener process*, is a fundamental example of a continuous-time Markov process taking values in the state space $\mathbb{R}$. It plays a central role in the construction of more general diffusion processes.

Definition 1.52 (Brownian motion/Wiener process, [11])**.** A stochastic process $(B_t)_{t \geq 0}$ is a (standard) Brownian motion if and only if the following properties hold:

(i) $B_0 = 0$ a.s.

(ii) The process has independent increments.

(iii) For all $t \geq 0$ and $h > 0$, the increments $B_{t+h} - B_t$ are normally distributed with expectation 0 and variance h.

(iv) The mapping $t \mapsto B_t$ is continuous.

Another important class of Markov processes is given by *diffusion processes* (including Brownian motion as a special case). These arise as solutions to *stochastic differential equations*.

Definition 1.53 (Stochastic differential equation, [11])**.** Let $(B_t)_{t \geq 0}$ be a Brownian motion. A stochastic differential equation (SDE) on $\mathbb{R}$ is a stochastic process adapted to $(\mathcal{F}_t)_{t \geq 0}$ of the form

$$X_t = X_0 + \underbrace{\int_0^t b(X_u, u)\, du}_{\text{Lebesgue integral}} + \underbrace{\int_0^t \sigma(X_u, u)\, dB_u}_{\text{Itô integral}}, \tag{1.8}$$

where $X_0 = x \in \mathbb{R}$ is a constant. For these integrals to be well-defined, we assume the coefficients satisfy the following integrability conditions almost surely for all $t \geq 0$:

$$\int_0^t |b(X_u, u)|\, du < \infty \quad \text{and} \quad \int_0^t |\sigma(X_u, u)|^2\, du < \infty.$$

In differential form, (1.8) can be written as

$$dX_t = \underbrace{b(X_t, t)}_{\text{drift}}\, dt + \underbrace{\sigma(X_t, t)}_{\text{diffusion}}\, dB_t. \tag{1.9}$$

Definition 1.54 (Diffusion process, [11])**.** A stochastic process $(X_t)_{t \geq 0}$ satisfying (1.8) is called a diffusion process.

In the context of population dynamics, we will focus on SDEs, where the drift and diffusion coefficients depend only on the current state of the process, i.e.,

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t. \tag{1.10}$$

The infinitesimal generator $\mathcal{L}$ associated with this process, acting on a function $f \in C_b^2(\mathbb{R})$, is given by [13]:

$$\mathcal{L}f(x) = b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x). \tag{1.11}$$

The *infinitesimal mean*, or *drift coefficient*, $b(x)$ represents the deterministic trend, such as selection or mutation, while the *infinitesimal variance* $\sigma^2(x)$ characterizes the local stochastic fluctuations, which correspond to the strength of genetic drift [13].

To derive the infinitesimal generator for such diffusion processes, we utilize *Itô's formula*. This result is defined for the broader class of processes called *Itô processes*.

Remark 1.55 ([11])**.** An Itô process is defined by replacing b and σ in (1.8) with any two adapted processes $(K_t)_{t\geq 0}$ and $(H_t)_{t\geq 0}$, that satisfy the same integrability conditions.

Given an Itô process X, we can determine the stochastic evolution of a sufficiently smooth function $f(X_t, t)$ using the following result:

Theorem 1.56 (Itô's formula, [11])**.** *Let X be an Itô process and let $f \in C^{2,1}$ (f continuous with continuous partial derivatives $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial^2 f}{\partial x^2}$). Then almost surely, for each t,*

$$\begin{aligned} f(X_t, t) - f(X_0, 0) = &\int_0^t \frac{\partial f}{\partial x}(X_s, s)\, dX_s \\ &+ \int_0^t \frac{\partial f}{\partial t}(X_s, s)\, ds + \frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial x^2}(X_s, s)\,(H_s)^2\, ds. \end{aligned} \tag{1.12}$$

Derivation of Equation (1.11). By applying Itô's formula to (1.10) with $f \in C_b^2$, we obtain

$$\begin{aligned} df(X_t) &= f'(X_t)\, dX_t + \frac{1}{2}f''(X_t)\sigma^2(X_t)dt \\ &= f'(X_t)\big(b(X_t)\, dt + \sigma(X_t)\, dB_t\big) + \frac{1}{2}f''(X_t)\sigma^2(X_t)\, dt \\ &= \left(b(X_t)f'(X_t) + \frac{1}{2}\sigma^2(X_t)f''(X_t)\right) dt + \sigma(X_t)f'(X_t)\, dB_t. \end{aligned}$$

Integrating both sides from 0 to t yields the integral representation

$$f(X_t) - f(X_0) = \int_0^t \left(b(X_s)f'(X_s) + \frac{1}{2}\sigma^2(X_s)f''(X_s)\right) ds + \int_0^t \sigma(X_s)f'(X_s)\, dB_s.$$

Taking the expectation $\mathbb{E}_x[\cdot]$, we are left with the expectation of the Lebesgue integral, since the Itô integral with respect to the Brownian motion is a continuous-time martingale [11]. Hence,

$$\mathbb{E}_x[f(X_t)] - f(x) = \mathbb{E}_x\left[\int_0^t \left(b(X_s)f'(X_s) + \frac{1}{2}\sigma^2(X_s)f''(X_s)\right) ds\right].$$

Finally, by using (1.7), the Dominated Convergence Theorem and the Mean Value Theorem for integrals, we obtain

$$\begin{aligned}
\mathcal{L}f(x) &= \lim_{t\downarrow 0} \frac{\mathbb{E}_x[f(X_t)] - f(x)}{t} \\
&= \lim_{t\downarrow 0} \mathbb{E}_x\left[\frac{1}{t}\int_0^t \left(b(X_s)f'(X_s) + \frac{1}{2}\sigma^2(X_s)f''(X_s)\right) ds\right] \\
&= \mathbb{E}_x\left[\lim_{t\downarrow 0} \frac{1}{t}\int_0^t \left(b(X_s)f'(X_s) + \frac{1}{2}\sigma^2(X_s)f''(X_s)\right) ds\right] \\
&= \mathbb{E}_x\left[b(X_0)f'(X_0) + \frac{1}{2}\sigma^2(X_0)f''(X_0)\right] \\
&= b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x).
\end{aligned}$$

■

The following result establishes pathwise uniqueness for the solution to (1.10):

Theorem 1.57 (Yamada-Watanabe, [4])**.** *Suppose that b is bounded and Lipschitz continuous. Assume further that there exists a function $\rho : [0,\infty) \to [0,\infty)$ such that $\rho(0) = 0$ and*

$$\int_0^\varepsilon \rho^{-2}(u)\, du = \infty \quad \text{for all } \varepsilon > 0.$$

Let σ be bounded and satisfy

$$|\sigma(x) - \sigma(y)| \leq \rho(|x-y|)$$

for all x, y. Then the solution to (1.10) *is pathwise unique.*

Chapter 2

Exchangeable sequences

In this section, we introduce the concept of exchangeable sequences and present a fundamental result in probability theory - de Finetti's Theorem. We start by exploring the idea of mixing distributions, which lies at the core of de Finetti's framework. Understanding this notion then allows us to see how exchangeability and the independence of sequences of random variables are related.

2.1 Exchangeability and Independence

Notation 2.1 ([6])**.** Let $k \in \mathbb{N}$. We use the notation

$$[k] \coloneqq \{1, \ldots, k\}$$

for the set of the first k positive integers. In this context, we also define $[\infty] := \mathbb{N}$.

Definition 2.2 (Permutation, [6])**.** A permutation σ of $[n]$ for $n \in \mathbb{N}$ is a bijection $\sigma : [n] \to [n]$. For $n = \infty$, a bijection $\sigma : \mathbb{N} \to \mathbb{N}$ is called a permutation of $\mathbb{N}$ if $\#\{i \in \mathbb{N} : \sigma(i) \neq i\} < \infty$, i.e., if σ has finite support.

Definition 2.3 (Exchangeable sequences, [1])**.** (i) A finite sequence of random variables $(X_1, X_2, \ldots, X_n)$ is exchangeable if

$$(X_1, X_2, \ldots, X_n) \overset{d}{=} (X_{\sigma(1)}, X_{\sigma(2)}, \ldots, X_{\sigma(n)})$$

for all σ permutations of $[n]$. Equivalently, for all real numbers $x_1, \ldots, x_n$,

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \mathbb{P}(X_{\sigma(1)} \leq x_1, \ldots, X_{\sigma(n)} \leq x_n)$$

or

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_{\sigma(1)}, \ldots, X_n \leq x_{\sigma(n)}).$$

(ii) An infinite sequence of random variables $(X_1, X_2, \ldots)$ is exchangeable if

$$(X_1, X_2, \ldots) \overset{d}{=} (X_{\sigma(1)}, X_{\sigma(2)}, \ldots)$$

for all σ with finite support.

Remark 2.4 ([26])**.** If $(X_i)_{i\geq 1}$ is a sequence of i.i.d. random variables, then it is exchangeable. However, the converse does not necessarily hold.

Proof. Let $X_1, \ldots, X_n$ be i.i.d. random variables. For any permutation σ of $[n]$, we have

$$\begin{aligned}\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) &\overset{\perp\!\!\!\perp}{=} \prod_{i=1}^{n} \mathbb{P}(X_i \leq x_i) \\ &= \prod_{i=1}^{n} \mathbb{P}(X_{\sigma(i)} \leq x_i) \qquad \text{(identically distributed)} \\ &\overset{\perp\!\!\!\perp}{=} \mathbb{P}(X_{\sigma(1)} \leq x_1, \ldots, X_{\sigma(n)} \leq x_n).\end{aligned}$$

Since the joint distribution is invariant under permutations for any finite n, the infinite sequence $(X_i)_{i\geq 1}$ is exchangeable. ■

To illustrate that exchangeability is a strictly weaker condition than being i.i.d., consider the following counterexample:

Example 2.5. Suppose an urn contains three balls: one red ball and two white balls. We draw all three balls without replacement and record the color of each draw. Define

$$X_i \coloneqq \begin{cases} 1, & \text{if the } i\text{-th drawn ball is red,} \\ 0, & \text{otherwise.} \end{cases}$$

Each sequence of draws contains exactly one red ball, so

$$\mathbb{P}(X_1 = 1,\ X_2 = 0,\ X_3 = 0) = \frac{1}{3} \cdot 1 \cdot 1 = \frac{1}{3}.$$

If we permute the indices, for example by swapping X_1 and X_2, we obtain

$$\mathbb{P}(X_2 = 1,\ X_1 = 0,\ X_3 = 0) = \frac{2}{3} \cdot \frac{1}{2} \cdot 1 = \frac{1}{3},$$

and similarly

$$\mathbb{P}(X_3 = 1,\ X_1 = 0,\ X_2 = 0) = \frac{2}{3} \cdot \frac{1}{2} \cdot 1 = \frac{1}{3}.$$

The same value is obtained for all other permutations. Hence, the joint distribution of (X_1, X_2, X_3) is invariant under any permutation of the indices, so the random variables X_1, X_2 and X_3 are *exchangeable*. However, they are clearly *not independent*, since exactly one of them equals 1 in every outcome.

Remark 2.6. For an exchangeable sequence $(X_i)_{i\geq 1}$, we have $X_1 \overset{d}{=} X_i$ for all $i \geq 1$.

2.2 Concept of mixing

The concept of mixing distributions describes how a single random variable can be viewed as a random choice among several candidate distributions. In such a setting, there are two distinct sources of randomness: the selection of the distribution itself and the subsequent draw from that chosen distribution. We will clarify this idea by considering a concrete example.

Example 2.7 (Urn Mixture, [26])**.** Consider two urns, A and B, with white ball proportions $p_A = 1/3$ and $p_B = 3/4$, respectively. We select an urn by tossing a fair coin (events Λ, Γ), and then sample with replacement until the first white ball appears. The waiting times $X_A \sim \text{Geom}(\frac{1}{3})$ and $X_B \sim \text{Geom}(\frac{3}{4})$ have distribution functions F_A and F_B.

Let W be the total waiting time for a white ball. By Bayes' rule of total causes (Theorem A.14), the distribution of W is given by

$$\mathbb{P}(W \leq k) = \mathbb{P}(W \leq k \mid \Lambda)\mathbb{P}(\Lambda) + \mathbb{P}(W \leq k \mid \Gamma)\mathbb{P}(\Gamma) = \tfrac{1}{2}F_A(k) + \tfrac{1}{2}F_B(k).$$

In this setting, W follows a *mixture distribution* of two equally weighted geometric components. The mixing proportions are governed by a $\text{Ber}(\frac{1}{2})$ random variable, which acts as the *mixing variable*.

This example illustrates the two sources of randomness: first, the random selection of the urn by the coin toss, and second, the sampling from the selected urn.

This logic extends naturally to larger collections of distributions and to continuous mixers [26]:

- If X is a *discrete mixing variable* with $\mathbb{P}(X = k) = p_k$, then the mixed distribution function is given by the sum

$$F(y) = \sum_{k=1}^{\infty} p_k F_k(y).$$

  The underlying random variable is a mixture of components with distribution functions $F_1, F_2, \ldots$ and mixer X.

- Similarly, if X is a *continuous mixing variable* with density $f_X(x)$, the distribution function of the mixed random variable Y is defined by the integral

$$F(y) = \int_{-\infty}^{\infty} \mathbb{P}(Y \leq y \mid X = x) f_X(x)\, dx.$$

We now extend the idea of mixing to sequences of random variables. Our goal is to study the distribution of sequences that can be viewed as mixtures of i.i.d. sequences. The precise meaning of this notion will be clarified through the introduction of random measures, which are closely related to the concept of probability kernels.

Definition 2.8 (Probability kernel, [19])**.** Let $(E, \mathcal{E})$ and $(S, \mathcal{S})$ be two measurable spaces. We call a probability kernel from E to S a map

$$\begin{aligned} K\colon\ E \times \mathcal{S} &\longrightarrow [0, \infty], \\ (x, B) &\longmapsto K(x, B), \end{aligned}$$

such that

(i) $\forall\ x \in E$, $K(x, \cdot)$ is a probability measure on $(S, \mathcal{S})$.

(ii) $\forall\ B \in \mathcal{S}$, the map $x \mapsto K(x, B)$ is $\mathcal{E}$-measurable.

Definition 2.9 (Random probability measure, [1])**.** Let $(E, \mathcal{E})$ be a measurable space. A random probability measure on $(E, \mathcal{E})$ is a measurable mapping

$$\alpha : (\Omega, \mathcal{F}) \longrightarrow (\mathcal{P}(E), \mathcal{G}),$$

where

- $\mathcal{P}(E)$ denotes the set of all probability measures on $(E, \mathcal{E})$;
- $\mathcal{G}$ is the smallest σ-field on $\mathcal{P}(E)$ that makes all evaluation maps

  $$e_B : \mathcal{P}(E) \longrightarrow [0, 1], \qquad e_B(\theta) = \theta(B), \quad B \in \mathcal{E},$$

  measurable.

Equivalently, α satisfies:

(i) $\forall\ \omega \in \Omega$, $\alpha(\omega, \cdot)$ is a probability measure on $(E, \mathcal{E})$;

(ii) $\forall\ B \in \mathcal{E}$, the mapping $\omega \mapsto \alpha(\omega, B)$ is $\mathcal{F}$-measurable.

Intuitively, a random probability measure is nothing else, than a randomly picked probability measure on a space $(E, \mathcal{E})$. From the preceding definition, it is clear that a random probability measure can be viewed as a probability kernel from the sample space Ω to the space E.

Example 2.10. Consider choosing a normal distribution $\mathcal{N}(x, x^2)$, where the parameter x is itself drawn from a distribution with density f_X on $(0, \infty)$. Let Y be a continuous mixture of normal distributions such that, conditionally on $X = x$, Y follows a normal distribution $\mathcal{N}(x, x^2)$. For any two non-negative measurable functions f and g, the joint expectation can be expressed as:

$$\begin{aligned} \mathbb{E}[f(X)g(Y)] &= \int_0^\infty f_X(x)\, f(x) \left(\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\ x}\ \exp\left(-\frac{(y-x)^2}{2x^2} \right) g(y)\, dy \right) dx \\ &= \int_0^\infty f_X(x)\, f(x) \left(\int_{-\infty}^\infty g(y)\, K(x, dy) \right) dx \\ &= \mathbb{E}\left[f(X)\ \left(\int_{-\infty}^\infty g(y)\, K(X, dy) \right) \right]. \end{aligned}$$

The kernel $K(X(\omega), dy)$ represents the random measure α as follows:

$$\alpha(\omega, dy) = K(X(\omega), dy) = \frac{1}{\sqrt{2\pi}\ X(\omega)}\ \exp\left(-\frac{(y - X(\omega))^2}{2X(\omega)^2}\right)\ dy.$$

Definition 2.11 (Mixture of i.i.d. sequences, [1])**.** Let $(X_i)_{i\geq 1}$ be a sequence of random variables, and let α be a random probability measure on $(E, \mathcal{E})$. We say that X is a mixture of i.i.d. sequences directed by α if the following two conditions are satisfied:

(i) The variables $(X_i)_{i\geq 1}$ are conditionally independent given α. That is,

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2, \ldots X_n \in B_n \mid \alpha) = \prod_i^n \mathbb{P}(X_i \in B_i \mid \alpha), \qquad B_i \in \mathcal{E}.$$

(ii) For each $i \geq 1$ and $B \in \mathcal{E}$, the conditional distribution is given by

$$\mathbb{P}(X_i \in B \mid \alpha) = \alpha(\omega, B).$$

2.3 De Finetti's Theorem

We will start this section with an example to create an intuition for de Finetti's Theorem.

Example 2.12. Suppose you have two visually identical coins in your pocket, one fair coin and one biased coin. Let the probabilities of flipping a head be:

- Coin 1 (Fair): $\mathbb{P}(\text{Head}) = 0.5$,
- Coin 2 (Biased): $\mathbb{P}(\text{Head}) = 0.9$.

You pick one coin at random (with probability 0.5) and flip it infinitely many times, recording the results as a sequence $(X_1, X_2, \ldots)$, where $X_i \in \{0, 1\}$.

The resulting sequence is exchangeable because, from your perspective, the joint probability of any specific pattern of outcomes depends only on the total number of heads and tails, not the order in which they appear. Since you do not know which coin was chosen, the outcomes are not independent. For instance, seeing many heads early on increases your belief that you are holding the biased coin, which in turn changes your prediction for the next flip.

After flipping the coin infinitely many times, the proportion of heads will settle on a specific value, allowing you to determine which coin you have. Specifically, the success parameter will be 0.5 with probability 0.5, or 0.9 with probability 0.5. Once this parameter is known, the dependency between the tosses vanishes. The tosses then behave as i.i.d. Bernoulli random variables with that fixed parameter.

In the example above, the distribution of the success parameter was discrete. De Finetti's Theorem generalizes this idea to any infinitely exchangeable sequence of 0s and 1s.

Theorem 2.13 (De Finetti's Theorem for indicators, [26])**.** *Let $(X_i)_{i\geq 1}$ be an infinitely exchangeable sequence of indicators. Then there exits a distribution function F such that for each n and $0 \leq k \leq n$,*

$$\mathbb{P}(X_1 = 1, \ldots, X_k = 1, X_{k+1} = 0, \ldots, X_n = 0) = \int_0^1 x^k (1-x)^{n-k}\, dF(x).$$

Formulated in words, de Finetti's Theorem asserts that the partial sum of any infinitely exchangeable sequence of Bernoulli random variables possesses a representation as a mixture of Binomial random variables. The distribution function F acts as the mixing distribution that governs the success parameter of the underlying Bernoulli process. Specifically, a random parameter Θ is first drawn from the distribution F. Then, conditioned on the realization $\Theta = x$, the variables $(X_i)_{i\geq 1}$ behave as i.i.d. Bernoulli random variables with success parameter x [26].

Proof. A very intuitive proof can be found in [26]. ■

The concept seen above can be extended to the general case of infinitely exchangeable sequences of real-valued random variables.

Theorem 2.14 (De Finetti's Theorem, [6])**.** *Let $(X_i)_{i\geq 1}$ be an infinitely exchangeable sequence of real-valued random variables. Define the sequence of empirical measures as*

$$\mu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}, \tag{2.1}$$

where δ_{X_i} denotes the Dirac measure at X_i. Then there exists a random probability measure μ on $\mathbb{R}$ such that, as $n \to \infty$, the sequence μ_n converges almost surely to μ in the sense of weak convergence of probability measures on $\mathbb{R}$.
Furthermore, conditionally on μ the $X_1, X_2, \ldots$ are i.i.d. with common law μ. In other words, the infinitely exchangeable sequence $(X_i)_{i\geq 1}$ is a mixture of i.i.d. sequences directed by μ.

Proof. We refer the reader to [1] for the proof. ■

Recall ([8])**.** Let μ_n and μ be probability measures defined on $\mathbb{R}$. We say that the sequence $(\mu_n)_{n\in\mathbb{N}}$ converges weakly to μ, denoted by $\mu_n \Rightarrow \mu$, if

$$\lim_{n\to\infty} \int_{\mathbb{R}} f\, d\mu_n = \int_{\mathbb{R}} f\, d\mu$$

for $f \in C_b(\mathbb{R})$.

Remark 2.15. In the context of Theorem 2.14, since the empirical measures μ_n and the limit μ are random measures, the convergence $\mu_n \Rightarrow \mu$ is understood as follows: For almost all $\omega \in \Omega$, the sequence of realizations $\mu_n(\omega)$ converges weakly to the realization $\mu(\omega)$.

Chapter 3

Partitions

This chapter introduces the reader to random partitions, focusing particularly on exchangeability. We will see how mass-partitions can be naturally encoded by exchangeable random partitions of $\mathbb{N}$, a useful property that will serve as an important tool in the following chapters.

3.1 Definitions and basic concepts

Definition 3.1 (Partition, [6])**.** A partition π of $B \subseteq \mathbb{N}$ is a countable collection $\pi = (\pi_1, \pi_2, \ldots)$ of disjoint blocks such that $\bigcup_{i\in\mathbb{N}} \pi_i = B$.

(i) Unless stated otherwise, the blocks π_i of the partition π are ordered by the increasing order of their least element:

$$\begin{aligned} &\pi_1 \quad \text{contains the smallest element of } B, \\ &\pi_2 \quad \text{contains the smallest element of } B \setminus \pi_1, \\ &\vdots \end{aligned}$$

(ii) We write $\mathcal{P}_B$ for the set of partitions of $B \subseteq \mathbb{N}$. When $B = [k]$ for some $k \in \mathbb{N} \cup \{\infty\}$, we simply write $\mathcal{P}_k := \mathcal{P}_{[k]}$. In particular, $\mathcal{P}_\infty := \mathcal{P}_{\mathbb{N}}$.

(iii) Take $B' \subseteq B$ and $\pi \in \mathcal{P}_B$. We denote by $\pi_{|B'}$ the restriction of π to B'. This is the partition of B' induced by the sequence of blocks $(\pi_i \cap B',\ i \in \mathbb{N})$.

(iv) We write $\#\pi := \sup\{i : \pi_i \neq \emptyset\}$ for the number of non-empty blocks of a partition π.

Remark 3.2. A partition $\pi \in \mathcal{P}_B$ can be identified with an equivalence relation on B defined by

$$i \sim_\pi j \quad \Longleftrightarrow \quad i \text{ and } j \text{ belong to the same block of } \pi,$$

for all $i, j \in B$.

For any $n \in \mathbb{N}$, the set of all possible partitions of $[n]$ has finite cardinality given by the *Bell number*.

Definition 3.3 (Bell numbers, [17])**.** The Bell number, denoted by B_n, counts the total number of ways a set of n elements can be partitioned into non-empty, disjoint subsets. The sequence $(B_n)_{n\geq 0}$ is defined by the recurrence relation

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k,$$

with initial condition $B_0 = 1$.

Definition 3.4 (Compatible/Consistent, [6])**.** Let $\pi^{[k]}$ be a partition of $[k]$. A sequence of partitions $\pi^{[1]}$, $\pi^{[2]}$, ... is called compatible if for all $k' \leq k$,

$$\pi^{[k']} = \pi^{[k]}_{|[k']}. \tag{3.1}$$

Lemma 3.5. *Let $(\pi^{[n]})_{n\in\mathbb{N}}$ for $\pi^{[n]} \in \mathcal{P}_n$ be a sequence of partitions. This sequence is compatible if and only if there exists $\pi \in \mathcal{P}_\infty$ such that $\pi_{|[n]} = \pi^{[n]}$ for all $n \in \mathbb{N}$. Moreover, π is uniquely determined by the sequence of partitions.*

Proof. ($\Rightarrow$): We assume that the sequence $(\pi^{[n]})_{n\in\mathbb{N}}$ is compatible. Let $\pi_i^{[n]}$ denote the i-th block of the partition $\pi^{[n]}$. By compatibility and the ordering rule, we know that the sequence of blocks $(\pi_i^{[n]}, n \in \mathbb{N})$ is increasing in the sense that $\pi_i^{[n]} \subseteq \pi_i^{[n+1]}$. Hence, we can define

$$\pi_i := \bigcup_{n\in\mathbb{N}} \pi_i^{[n]},$$

which is a block of the partition of $\mathbb{N}$. Then $(\pi_i, i \in \mathbb{N})$ is a partition of $\mathbb{N}$, which is uniquely determined by the sequence $(\pi^{[n]})_{n\in\mathbb{N}}$.

($\Leftarrow$): This implication is immediate. By assumption, $\pi^{[n]} = \pi_{|[n]}$ for π in $\mathcal{P}_\infty$ and for all $n \in \mathbb{N}$. Since the restriction satisfies $(\pi_{|[n]})|_{[n-1]} = \pi_{|[n-1]}$, we have $\pi^{[n]}_{|[n-1]} = \pi^{[n-1]}$, which is precisely the compatibility condition. ■

Lemma 3.6 ([6])**.** *The space $\mathcal{P}_\infty$ is endowed with a distance*

$$d(\pi, \pi') = 1/\max\{k \in \mathbb{N} : \pi_{|[k]} = \pi'_{|[k]}\}$$

with the convention $1/\max\mathbb{N} = 0$. The metric space $(\mathcal{P}_\infty, d)$ is compact.

Endowed with this metric, $\mathcal{P}_\infty$ is a Polish space (separable, completely metrizable topological space, [19]). The Polish property is significant because it provides a robust framework for probability theory. In the context of random partitions, it ensures that fundamental concepts such as convergence in distribution and conditional probabilities are well-defined.

Proof. Let $(\pi^{(n)})_{n\in\mathbb{N}}$ be an arbitrary sequence of partitions in $\mathcal{P}_\infty$. We will use following procedure:

- Since $\mathcal{P}_1$ consists of a single partition, the sequence $(\pi^{(n)}_{|[1]})_{n\in\mathbb{N}}$ is trivially constant. Thus, every subsequence of $(\pi^{(n)})_{n\in\mathbb{N}}$ is stable on $[1]$.
- The set $\mathcal{P}_2$ is finite. Hence, among the values $\pi^{(1)}_{|[2]}, \pi^{(2)}_{|[2]}, \dots$ one partition of $[2]$ appears infinitely often. Choose a subsequence of $(\pi^{(n)})_{n\in\mathbb{N}}$ along which $\pi^{(n)}_{|[2]}$ is constant. This subsequence is then stable on $[1]$ and on $[2]$. We call this sequence $(\pi^{(n)}_2)_{n\in\mathbb{N}}$.
- Since $\mathcal{P}_3$ is again finite, the sequence $(\pi^{(n)}_{|[3]})_{n\in\mathbb{N}}$ takes only finitely many values. From the sequence constructed in the previous step, extract a further subsequence, on which $\pi^{(n)}_{|[3]}$ is constant. Again, rename this subsequence as $(\pi^{(n)}_3)_{n\in\mathbb{N}}$. This subsequence is stable on $[1], [2], [3]$.
- Proceeding inductively, for each $k \in \mathbb{N}$ we extract from the previous sequence a subsequence that is constant on $[k]$. After relabeling at each stage, we obtain for each k, a subsequence $(\pi^{(n)}_k)_{n\in\mathbb{N}}$ that is stable on $[1], \dots, [k]$.

Define the diagonal subsequence by $\tilde{\pi}^{(m)} := \pi^{(m)}_m$. Consequently, for a fixed k, the restriction to $[k]$ is constant for all terms of the diagonal sequence beyond the k-th term:

$$\tilde{\pi}^{(m)}_{|[k]} = \tilde{\pi}^{(m')}_{|[k]} \qquad \text{for all } m, m' \geq k.$$

We constructed a compatible sequence of partitions of $[k]$ for varying $k \in \mathbb{N}$. Hence, by Lemma 3.5, this sequence can be expressed as $\hat{\pi}_{|[k]}$ for $\hat{\pi} \in \mathcal{P}_\infty$. Therefore, we can conclude

$$d(\tilde{\pi}^{(n)}, \hat{\pi}) \leq \frac{1}{n}.$$

Since n can be taken arbitrarily large, this implies $\lim_{n\to\infty} \tilde{\pi}^{(n)} = \hat{\pi}$ in $(\mathcal{P}_\infty, d)$.

Thus, every sequence in $\mathcal{P}_\infty$ admits a convergent subsequence, proving that (P_∞, d) is compact. ■

Definition 3.7 (Asymptotic frequency)**.** Let $\pi = (\pi_1, \pi_2, \dots)$ be a partition of $\mathbb{N}$ consisting of blocks π_i. The asymptotic frequency (if it exists) of a block π_i is defined as

$$|\pi_i| := \lim_{n\to\infty} \frac{\#(\pi_i \cap [n])}{n}. \tag{3.2}$$

It can be understood as a measure of the relative size of a block and therefore it is often called the mass of a block.

Example 3.8. If π is the partition of $\mathbb{N}$ into odd and even numbers, then the partition consists of exactly two blocks, each having a mass of $1/2$.

Definition 3.9 (Permutation of a partition, [6])**.** For every permutation σ (see Definition 2.2) and partition $\pi \in \mathcal{P}_n$, we can define a new equivalence relation

$$i \sim j \iff \sigma(i) \sim_\pi \sigma(j), \qquad \text{for all } i, j \in [n]. \tag{3.3}$$

This equivalence relation can be identified as a partition denoted by $\sigma(\pi)$.

Remark 3.10 ([6])**.** In other words, (3.3) means that the blocks of $\sigma(\pi)$ are the images of the blocks of π under σ^{-1}, the inverse permutation of σ. That is,

$$\sigma(\pi) = (\sigma^{-1}(B) : B \in \pi).$$

Definition 3.11 (Exchangeable random partition, [6])**.** A random partition Π of $[n]$ for $n \in \mathbb{N} \cup \{\infty\}$ is exchangeable if its law is invariant under the action of every permutation σ of $[n]$, i.e.,

$$\Pi \overset{d}{=} \sigma(\Pi). \tag{3.4}$$

Furthermore, a random partition of $\mathbb{N}$ is exchangeable if and only if its restrictions to all finite subsets $[n]$ are exchangeable.

Intuitively, an exchangeable random partition is a partition for which labels are irrelevant; only the pattern of grouping matters. This concept becomes fundamental in Kingman's coalescent (see Chapter 4), where we trace the genealogy of a population by focusing on the structure of lineages rather than the identity of individuals.

Example 3.12 (Non-example)**.** Bond percolation on $\mathbb{Z}^2 = \{(z_1, z_2), z_i \in \mathbb{Z}\}$. Every possible edge between adjacent vertices in the grid is independently declared open with probability $p \in (0, 1)$, and closed with probability $1 - p$; see Grimmett [18]. We first enumerate all vertices of $\mathbb{Z}^2$ as $(v_1, v_2, \ldots)$ in an arbitrary order. We say that i and j are in the same block of the partition $\Pi(\omega)$ if v_i and v_j are in the same connected component of the realization ω of a bond percolation. This random partition Π is not exchangeable, since the probability that two vertices belong to the same connected component depends on their relative position in the grid.

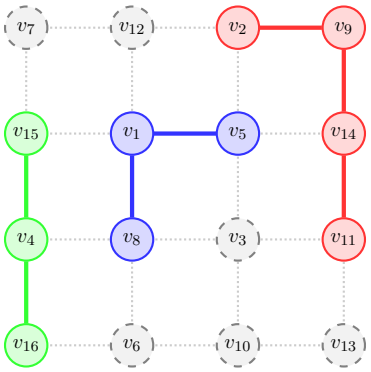


Figure 3.1: A realization of bond percolation on $\mathbb{Z}^2$. Colored lines denote the open bonds that form the connected components, while dotted lines represent closed bonds.

To study the structure of exchangeable partitions rigorously, we introduce the concept of a mass-partition.

Definition 3.13 (Space $\mathcal{S}_0$ and mass-partition, [5],[6])**.** A mass-partition is an infinite sequence

$$\mathbf{s} = (s_1, s_2, \ldots)$$

such that

$$s_1 \geq s_2 \geq \cdots \geq 0 \quad \text{and} \quad \sum_{i=1}^{\infty} s_i \leq 1.$$

The space of mass-partitions is denoted by $\mathcal{S}_0$.

Definition 3.14 (Dust, [6])**.** The dust associated with $\mathbf{s} \in \mathcal{S}_0$ is defined as

$$s_0 := 1 - \sum_{i=1}^{\infty} s_i.$$

It can be thought of as the portion of the total mass that is not contained in any of the positive components (those with $s_i > 0$).

Remark 3.15. In a later proof, it will be convenient to represent a mass-partition as the ranked sequence of the masses of the atoms of a probability measure, with the dust corresponding to the measure's continuous component.

Proposition 3.16 ([6])**.** *Consider the space $\mathcal{S}_0$ equipped with the metric*

$$d(s, s') := \sup_{i \in \mathbb{N}} |s_i - s'_i|.$$

With this metric, $\mathcal{S}_0$ forms a compact metric space. In addition, the topology generated by d agrees with the topology of pointwise convergence on $\mathcal{S}_0$.

Proof. We refer the reader to [6] for the proof. ■

Relying on the fact that any open subset of the unit open interval $(0, 1)$ admits a unique representation as a countable union of disjoint open intervals, we define an interval-partition as follows:

Definition 3.17 (Interval partition, [6])**.** Let $\vartheta \subseteq (0, 1)$ be an open set. The collection $\mathcal{J}$ of the interval-components of ϑ is called an interval-partition. The lengths of these components are referred to as spacings.

Remark 3.18 ([6])**.** Given $\mathbf{s} \in \mathcal{S}_0$, we can construct an interval-partition $\mathcal{J}$, such that the ranked sequence of spacings is given by $\mathbf{s}$. $\mathcal{J}$ is then called an interval-representation of $\mathbf{s}$. There are many different interval-representations for the same mass-partition $\mathbf{s}$.

Remark 3.19 (Notational Convention)**.** Before proceeding, we establish a crucial notational convention. By a slight abuse of notation, we will use ϑ interchangeably to refer to both the open set and its corresponding interval-partition.

To bridge the gap between mass-partitions and partitions of $\mathbb{N}$, we utilize a procedure known as the paintbox construction. This method provides a natural way to generate an exchangeable random partition of $\mathbb{N}$ directly from a fixed or random mass-partition.

Paintbox construction: Let $\mathbf{s} \in \mathcal{S}_0$ be a given mass-partition with strictly positive dust ($s_0 > 0$). Let $\vartheta = \bigcup_{k\in\mathbb{N}} J_k$ be an interval-representation of $\mathbf{s}$, where each J_k denotes the specific disjoint open interval-component of length $|J_k| = s_k$.

Let $U_1, U_2, \ldots$ be i.i.d. random variables uniformly distributed on $(0, 1)$. Let Π be a random partition induced by the equivalence relation

$$i \sim_\Pi j \quad \Longleftrightarrow \quad (i = j) \quad \text{or} \quad (U_i, U_j \text{ fall into the same interval-component of } \vartheta).$$

The first condition $(i = j)$ implies that if U_i falls into the dust, i forms a singleton block in the partition [6].

Let us define a random variable $X_i^k = \mathbb{1}_{\{U_i \in J_k\}}$ for some $k \geq 1$. Then

$$\mathbb{E}(X_i^k) = 1 \cdot \mathbb{P}(U_i \in J_k) + 0 \cdot \mathbb{P}(U_i \notin J_k) = s_k.$$

Since the X_i^k are i.i.d. Bernoulli variables with mean s_k, by the Law of Large Numbers (Theorem A.17), we obtain

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i^k = s_k \quad \text{a.s.}$$

This demonstrates that the components of the mass-partition $\mathbf{s} = (s_1, s_2, \ldots)$ define the asymptotic frequencies of the blocks of the random partition Π. Moreover, for any strictly positive mass $s_k > 0$, we have that infinitely many elements of $\mathbb{N}$ belong to that specific block almost surely.

We can apply this same reasoning to the dust part. The overall proportion of integers that fall into the dust converges to s_0. However, according to the paintbox rules, every element in this region forms its own singleton block. Consequently, while the total mass of the dust is s_0, each individual block within it has an asymptotic frequency of 0. For this reason, s_0 is referred to as the dust. We say that the partition Π has no dust if $s_0 = 0$, which occurs precisely when Π contains no singleton blocks.

Definition 3.20 (Paintbox partition)**.** The random partition Π constructed via this method is known as paintbox partition.

To visualize this construction (see Example 3.21), we can imagine that each interval-component of ϑ represents a different color. We assign integer i the color corresponding to the interval in which U_i lies. If $U_i \notin \vartheta$ (the dust), the integer i is painted with a completely new, unique color. The resulting partition Π is then formed by grouping all integers that share the same color.

Example 3.21. Suppose the black, hatched segment represents the dust component s_0. The indices are assigned to blocks according to the following color-coding:

- **Red block**: $\{6\}$, as U_6 falls within the red segment;
- **Orange block**: $\{1, 4, 8\}$, as U_1, U_4, and U_8 all fall within the orange segment;
- **Blue block**: $\{3, 7\}$, since U_3 and U_7 fall within the blue segment;
- **Singletons**: $\{2\}$ and $\{5\}$ each form an independent block, as U_2 and U_6 fall within the dust component s_0.

Consequently, we obtain the paintbox partition $\Pi_{|[8]} = (\{2\}, \{6\}, \{5\}, \{1, 4, 8\}, \{3, 7\})$.

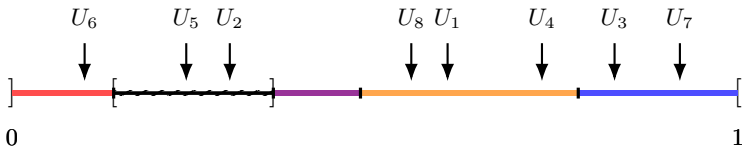


Figure 3.2: Paintbox partition $\Pi_{|[8]}$.

Lemma 3.22. *The paintbox partition* Π *based on* $\mathbf{s} \in \mathcal{S}_0$ *is a random exchangeable partition and its law does not depend on the interval-representation of* $\boldsymbol{s}$*. We will denote this law as* $\rho_{\mathbf{s}}$*.*

Proof. Let Π be a paintbox partition based on a mass-partition $\mathbf{s}$. First, we want to show that Π is exchangeable. For that, let σ be a permutation with finite support and take ϑ to be an interval-representation of $\mathbf{s}$. The partition $\sigma(\Pi)$ is specified by the following equivalence relation:

$$\begin{aligned} i \sim_{\sigma(\Pi)} j &\iff \sigma(i) \sim_{\Pi} \sigma(j) \\ &\iff (i = j) \text{ or } (U_{\sigma(i)}, U_{\sigma(j)} \text{ belong to the same interval-component of } \vartheta). \end{aligned}$$

The random variables $U_{\sigma(i)}, U_{\sigma(j)}, \ldots$ are still independent and uniformly distributed on $(0, 1)$. Therefore,

$$(U_1, U_2, \ldots) \overset{d}{=} (U_{\sigma(1)}, U_{\sigma(2)}, \ldots).$$

Hence, we can conclude that Π is exchangeable.

Next, let ϑ' be another interval-representation of $\mathbf{s}$. Let

$$\vartheta = \bigcup_{k \in \mathbb{N}} J_k \qquad \text{and} \qquad \vartheta' = \bigcup_{k \in \mathbb{N}} J'_k$$

where $J_k = (a_k, a_k + s_k)$ and $J'_k = (a'_k, a'_k + s_k)$. By construction, $|J_k| = |J'_k| = s_k$. We can define a bijection f by matching the corresponding interval-components linearly by

$$f(x) := \begin{cases} a'_k + (x - a_k), & \text{if } x \in J_k, \text{ for some } k \in \mathbb{N}, \\ g(x), & \text{if } x \in (0, 1) \setminus \vartheta, \end{cases}$$

where g is an arbitrary measure-preserving map from the dust of ϑ to the dust of ϑ'. By the definition of f, it is a measure-preserving bijection. Therefore, $f(U_1)$, $f(U_2)$,... are again i.i.d. and uniformly distributed on $(0,1)$ (see Lemma A.8). Thus,

$$\begin{aligned} i \sim_\Pi j &\iff (i=j) \text{ or } (U_i, U_j \text{ belong to the same interval-component of } \vartheta) \\ &\iff (i=j) \text{ or } (f(U_i), f(U_j) \text{ belong to the same interval-component of } \vartheta'). \end{aligned}$$

Hence, i and j are in the same interval-component of ϑ if and only if they are in the same interval-component of ϑ'. ■

In the paintbox construction above, the mass-partition $\mathbf{s}$ was taken to be fixed. However, we also obtain a random exchangeable partition if $\mathbf{s} \in \mathcal{S}_0$ is chosen at random, provided that the random variables $(U_i)_{i\geq 1}$ are chosen independently of $\mathbf{s}$. The following theorem states that this is the most general form of an exchangeable partition.

Theorem 3.23 (Kingman's Representation Theorem)**.** *Let Π be any exchangeable random partition of $\mathbb{N}$. Then we can find a probability distribution $\mu(d\mathbf{s})$ on $\mathcal{S}_0$ such that*

$$\mathbb{P}(\Pi \in \cdot) = \int_{S_0} \rho_{\mathbf{s}}(\cdot)\, \mu(d\mathbf{s}). \tag{3.5}$$

Remark 3.24. In essence, Theorem 3.23 states that every exchangeable partition is a mixture of paintbox partitions. If $\mathbf{S}$ is a random mass-partition with distribution μ, then the conditional distribution of Π given $\mathbf{S}$ is exactly the paintbox distribution:

$$\Pi \mid \mathbf{S} \overset{d}{=} \rho_{\mathbf{S}}. \tag{3.6}$$

Proof of Remark. We want to show that (3.5) equivalent to (3.6) by utilizing the probability kernel. Let $f : \mathcal{S}_0 \to \mathbb{R}_+$ and $h : \mathcal{P}_\infty \to \mathbb{R}_+$ be measurable functions. Then

$$\mathbb{E}[f(\mathbf{S})\, h(\Pi)] = \int_{\mathcal{S}_0} f(\mathbf{s}) \left(\int_{\mathcal{P}_\infty} h(\pi)\, K(\mathbf{s}, d\pi) \right) \mu(d\mathbf{s}) = \mathbb{E}\left[f(\mathbf{S}) \int_{\mathcal{P}_\infty} h(\pi)\, K(\mathbf{S}, d\pi) \right],$$

where $K : \mathcal{S}_0 \times \mathcal{P}_\infty \to [0,1]$ is a probability kernel.

On the other hand, by Kingman's Representation Theorem,

$$\mathbb{E}[f(\mathbf{S})\, h(\Pi)] = \int_{\mathcal{S}_0} f(\mathbf{s}) \left(\int_{\mathcal{P}_\infty} h(\pi)\, \rho_{\mathbf{s}}(d\pi) \right) \mu(d\mathbf{s}) = \mathbb{E}\left[f(\mathbf{S}) \int_{\mathcal{P}_\infty} h(\pi)\, \rho_{\mathbf{S}}(d\pi) \right].$$

By equating these two representations, it follows that

$$K(\mathbf{S}, \cdot) = \rho_{\mathbf{S}}(\cdot).$$

Thus, the conditional distribution of Π given $\mathbf{S}$ is exactly the paintbox distribution determined by $\mathbf{S}$. ■

Proof of Theorem 3.23. Let Π be an exchangeable partition of $\mathbb{N}$. Let $b : \mathbb{N} \to \mathbb{N}$ be a choice function for Π, defined by

$$b(i) \coloneqq \min\{\, j \in \mathbb{N} : j \sim_\Pi i \,\}.$$

Take $U_1, U_2, \ldots$ to be a sequence of i.i.d. Uniform$(0,1)$ random variables, independent of Π and of the choice function b. Define

$$X_i \coloneqq U_{b(i)}.$$

Then, for $i, j \in \mathbb{N}$,

$$i \sim_\Pi j \quad \iff \quad b(i) = b(j) \quad \iff \quad X_i = X_j.$$

Hence, the partition Π is determined by equality relations among the sequence $X = (X_1, X_2, \ldots)$.

Claim 1: The law of the sequence X does not depend on b.
Proof of Claim 1: Let γ be a partition of $[n]$ with blocks $B_1, \ldots, B_k$. For each block B_r, there is a unique representative index $b_r = b(i)$ for all $i \in B_r$. Thus, while the sequence $(U_{b(1)}, \ldots, U_{b(n)})$ has n terms, it only contains k distinct values. Each term $U_{b(i)}$ simply inherits the value of U_{b_r} assigned to its specific block B_r. Hence, conditional on the event $\{\Pi_{|[n]} = \gamma\}$ we have

$$\prod_{i=1}^{n} f_i(U_{b(i)}) = \prod_{r=1}^{k} \prod_{i \in B_r} f_i(U_{b_r})$$

where $f_1, \ldots, f_n : (0,1) \to \mathbb{R}_+$ are measurable functions. Therefore,

$$\mathbb{E}\left[\prod_{i=1}^{n} f_i(U_{b(i)}) \,\middle|\, \Pi_{|[n]} = \gamma\right] = \mathbb{E}\left[\prod_{r=1}^{k} \prod_{i \in B_r} f_i(U_{b_r}) \,\middle|\, \Pi_{|[n]} = \gamma\right] \overset{\perp\!\!\!\perp}{=} \prod_{r=1}^{k} \mathbb{E}\left[\prod_{i \in B_r} f_i(U_{b_r}) \,\middle|\, \Pi_{|[n]} = \gamma\right].$$

Since each U_{b_r} is independent of Π, we may drop the conditioning

$$\prod_{r=1}^{k} \mathbb{E}\left[\prod_{i \in B_r} f_i(U_{b_r}) \,\middle|\, \Pi_{|[n]} = \gamma\right] = \prod_{r=1}^{k} \mathbb{E}\left[\prod_{i \in B_r} f_i(U_{b_r})\right] = \prod_{r=1}^{k} \left(\int_0^1 \prod_{i \in B_r} f_i(u)\, du\right).$$

Claim 2: The sequence X is exchangeable.
Proof of Claim 2: Let σ be a permutation of $\mathbb{N}$. Define $U'_i \coloneqq U_{\sigma(i)}$. Then,

$$X_{\sigma(i)} = U_{b(\sigma(i))} = U'_{b'(i)},$$

where $b' \coloneqq \sigma^{-1} \circ b \circ \sigma$. We observe that b' acts as a valid choice function for the permuted partition $\sigma(\Pi)$. Indeed, for $i, j \in \mathbb{N}$,

$$i \sim_{\sigma(\Pi)} j \iff \sigma(i) \sim_\Pi \sigma(j)$$

$$\begin{aligned}
&\iff b(\sigma(i)) = b(\sigma(j)) \\
&\iff \sigma^{-1}(b(\sigma(i))) = \sigma^{-1}(b(\sigma(j))) \\
&\iff b'(i) = b'(j).
\end{aligned}$$

Hence,

$$i \sim_{\sigma(\Pi)} j \Leftrightarrow X_{\sigma(i)} = X_{\sigma(j)}.$$

The permutation σ only reorders the uniforms. Thus, $U'_i = U_{\sigma(i)}$ are still i.i.d. Uniform$(0,1)$ random variables and independent of $\sigma(\Pi)$. Moreover, $\sigma(\Pi) \overset{d}{=} \Pi$, since Π was assumed to be exchangeable. Hence, we have that $(U, \Pi) \overset{d}{=} (U', \sigma(\Pi))$. Since X is built from (U, Π) using b, and the permuted sequence X_σ is built from $(U', \sigma(\Pi))$ using b', and from Claim 1 we know that the specific choice function used does not matter, we can conclude that X is exchangeable.

Since X is an infinitely exchangeable sequence, de Finetti's Theorem (Theorem 2.14) implies the existence of a random probability measure $\tilde{\mu}$ on $(0,1)$ such that, conditionally on $\tilde{\mu}$, the random variables $(X_i)_{i\geq 1}$ are i.i.d. with common law $\tilde{\mu}$.

Conditioning on $\tilde{\mu}$, we define F as the cumulative distribution function of $\tilde{\mu}$. Let $q(u)$ denote its (generalized) inverse function:

$$q(u) := \inf\{\, y \in \mathbb{R} : F(y) > u \,\}, \qquad u \in (0,1).$$

Let $V_1, V_2, \ldots$ be i.i.d. Uniform$(0,1)$ random variables. By the Inverse Transform Sampling (see Theorem A.9) we have that $q(V_i) \sim \tilde{\mu}$. Moreover, $(q(V_i))_{i\geq 1}$ are independent random variables. Hence,

$$(X_1, X_2, \ldots) \overset{d}{=} (q(V_1), q(V_2), \ldots).$$

Define a random partition Π' of $\mathbb{N}$ by the following equivalence relation

$$i \sim_{\Pi'} j \quad \iff \quad q(V_i) = q(V_j).$$

Let $\tilde{\mu}$ have an atom at x_k with mass $s_k > 0$, meaning that $\tilde{\mu}(\{x_k\}) = s_k$. The cumulative distribution function F will exhibit a jump at x_k. Therefore, we have the following:

$$\begin{aligned}
F(x_k^-) &= \tilde{\mu}((0, x_k)), \\
F(x_k) &= F(x_k^-) + s_k.
\end{aligned}$$

We can analyze the behavior of $q(u)$ around this jump:

- If $u < F(x_k^-)$, then $q(u) \leq x_k^-$.
- If $F(x_k^-) \leq u < F(x_k)$, then $q(u) = x_k$.

Consequently, for all u in the interval $I_k = [F(x_k^-), F(x_k))$, we have $q(u) = x_k$. Since $|I_k| = F(x_k) - F(x_k^-) = s_k$, we conclude that q is constant on an interval of length s_k. Thus, for $i, j \in \mathbb{N}$,

$$i \sim_{\Pi'} j \iff \exists\, k \text{ such that } q(V_i) = q(V_j) = x_k \iff V_i, V_j \in I_k.$$

That is, V_i and V_j fall in the same interval-component of $(0,1)$ of length s_k.

Thus, conditionally on $\tilde{\mu}$, $\Pi' \sim \rho_{\mathbf{s}}$ where $\mathbf{s} = (s_1, s_2, \ldots) \in \mathcal{S}_0$ are the atom masses of $\tilde{\mu}$, and $s_0 = 1 - \sum_{i\geq 1} s_i$ is the continuous component of the measure. Furthermore, since the partition Π is determined by the equality relations of the sequence X, and conditionally on $\tilde{\mu}$, $(X_1, X_2, \ldots) \overset{d}{=} (q(V_1), q(V_2), \ldots)$, it follows that

$$\Pi \mid \tilde{\mu} \overset{d}{=} \Pi' \mid \tilde{\mu} \sim \rho_{\mathbf{s}}.$$

■

Kingman's Theorem has several important consequences. Almost surely, the only blocks with finitely many elements are the singletons. All other blocks contain infinitely many elements and possess positive asymptotic frequencies. Furthermore, the random mass-partition $\mathbf{S} \in \mathcal{S}_0$ can be completely recovered from the random exchangeable partition Π as the sequence of its ranked asymptotic frequencies. In particular, if $\Pi = (\Pi_1, \Pi_2, \ldots)$, then

$$\sum_{i=1}^{\infty} |\Pi_i| = 1 - S_0 \quad \text{a.s.}$$

This relationship is formalized by the following correspondence:

Corollary 3.25 (Kingman's correspondence)**.** *There exists a one-to-one correspondence between the law of an exchangeable random partitions* Π *and probability measures* μ *on the space* S_0 *of mass-partitions. This bijection is known as Kingman's correspondence.*

Moreover, even though the map from a partition to its corresponding ranked asymptotic frequencies is not continuous, Kingman's correspondence itself is a continuous map, which is a highly useful structural result.

Theorem 3.26 (Continuity of Kingman's correspondence, [6])**.** *Let* $\Pi^{(n)}$, $n \in \mathbb{N} \cup \{\infty\}$*, be random exchangeable partitions, and let*

$$\mathbf{S}^{(n)} := (S_1^{(n)}, S_2^{(n)}, \ldots)$$

denote their ranked frequency sequences. The following statements are equivalent:

(i) $\mathbf{S}^{(n)} \overset{d}{\Rightarrow} \mathbf{S}^{(\infty)}$ *in* $\mathcal{S}_0$*;*

(ii) $\Pi^{(n)} \overset{d}{\Rightarrow} \Pi^{(\infty)}$ *in* $\mathcal{P}_\infty$*.*

Proof. See [6] for the proof. ■

3.1.1 Size-biased sampling

Until now, we have seen that the terms of a mass-partition $\mathbf{s} \in \mathcal{S}_0$ are arranged from left to right in decreasing order with respect to $\geq$. In this section, we are going to discuss a procedure that that rearranges these elements randomly, but with the bias that larger blocks are more likely to appear first.

We may interpret a mass-partition $\mathbf{s} \in \mathcal{S}_0$ with no dust as a discrete probability measure. In particular, we can define an integer-valued random variable 1^* such that

$$\mathbb{P}(1^* = i) = s_i \quad \text{for } i \in \mathbb{N}.$$

Every random variable X with the same distribution as s_{1^*} is called a *size-biased sample of* $\mathbf{s}$ [6].

For a mass-partition with dust ($s_0 > 0$), we extend 1^* to take values in $\mathbb{N} \cup \{\infty\}$ such that

$$\mathbb{P}(1^* = i) = s_i \quad \text{for } i \in \mathbb{N}, \qquad \text{and} \quad \mathbb{P}(1^* = \infty) = s_0.$$

We clearly have $s_\infty = 0$. As in the non-dust case, a random variable X distributed according to s_{1^*} is called a *size-biased sample of* $\mathbf{s}$ [6].

Now consider an exchangeable random partition Π corresponding to a random mass-partition $\mathbf{S} = (S_1, S_2, \ldots) \in \mathcal{S}_0$ with no dust almost surely. Let 1^* be the random index of the block containing the integer 1. By Kingman's Representation Theorem, conditionally on $\mathbf{S}$, Π follows the paintbox distribution $\rho_{\mathbf{S}}$. Under this construction, the integer 1 is assigned to the i-th block with asymptotic frequency S_i if and only if the independent uniform random variable $U_1 \sim \text{Uniform}(0, 1)$ falls into the interval-component with length S_i. Consequently, it immediately follows that

$$\mathbb{P}(1^* = i \mid \mathbf{S}) = S_i.$$

Hence, the mass of the block containing the integer 1, given by $X = S_{1^*}$, is a size-biased sample of the random mass-partition $\mathbf{S}$. X is also called the *mass of a size-biased picked block*.

Theorem 3.27. *Let Π be an exchangeable partition with ranked random frequencies $P = (P_1, P_2, \ldots)$ with no dust almost surely. Let f be a non-negative function. Then*

$$\mathbb{E}\left(\sum_i f(P_i)\right) = \int_0^1 \frac{f(x)}{x}\, \mu(dx), \tag{3.7}$$

where μ defines the law of the mass of a size-biased picked block X.

Proof. Let $g(x) = \frac{f(x)}{x}$ for f a non-negative function. By Kingman's correspondence, we have that $(P_1, P_2, \dots) \stackrel{d}{=} (S_1, S_2, \dots) \in \mathcal{S}_0$. Hence, by applying this fact and the by Tower property, we obtain

$$\mathbb{E}\left[g(X)\right] = \mathbb{E}\left[\sum_{i\geq 1} \mathbb{1}_{\{1^*=i\}}\, g(P_i)\right] = \mathbb{E}\left[\sum_{i\geq 1} \mathbb{E}\left[\mathbb{1}_{\{1^*=i\}} \mid P\right]\, g(P_i)\right] = \mathbb{E}\left[\sum_{i\geq 1} P_i\, g(P_i)\right].$$

Since μ is the mass of X, we have

$$\mathbb{E}\left[g(X)\right] = \int_0^1 g(x)\mu(dx).$$

By substituting $g(x)$ with $\frac{f(x)}{x}$, we can conclude that

$$\mathbb{E}\left[\frac{f(X)}{X}\right] = \mathbb{E}\left[\sum_{i\geq 1} f(P_i)\right] = \int_0^1 \frac{f(x)}{x}\mu(dx).$$

■

Let $f(x) = x^{n+1}$. By using Equation (3.7), we obtain

$$\mathbb{E}[X^n] = \mathbb{E}\left[\sum_i P_i^{n+1}\right],$$

which can be interpreted as the expected value that $n+1$ randomly chosen points fall into the same segment.

Theorem 3.28. *Let* Π *be a random exchangeable partition with* N *blocks. Then we have the formula*

$$\mathbb{E}\left[N\right] = \mathbb{E}\left[\frac{1}{X}\right].$$

Proof. First, assume that Π has dust a.s., corresponding to a random mass-partition $P = (P_1, P_2, \dots)$ for $P_0 > 0$. Conditional on P, the partition Π follows the paintbox distribution. By the Law of Large Numbers, the proportion of integers that fall into the dust region is P_0. Since each such integer forms its own singleton block, $P_0 > 0$ implies that the number of blocks N is infinite almost surely. Thus,

$$\mathbb{E}[N \mid P] = \infty.$$

Let X be the mass of the block containing 1 and let 1^* be the random index of this block. Since the integer 1 falls into the dust with probability P_0, we have

$$\mathbb{P}(X = 0 \mid P) = \mathbb{P}(1^* = \infty \mid P) = P_0 > 0.$$

Because $X = 0$ with strictly positive probability, its conditional expectation diverges:

$$\mathbb{E}\left[\frac{1}{X} \;\middle|\; P\right] = \infty.$$

Hence, for the dust case we can conclude

$$\mathbb{E}[N] = \infty = \mathbb{E}\left[\frac{1}{X}\right].$$

Now, assume that Π has no dust almost surely. We work conditionally on a mass-partition $P = (P_1, P_2, \dots)$ and use the paintbox construction. Let N_n be the number of blocks of the restricted partition $\Pi_{|[n]}$. Let I_i be the disjoint interval-components in $(0,1)$ with lengths $|I_i| = P_i$. We define the set of indices

$$A_i = \{j \in \mathbb{N} : U_j \in I_i\}$$

and the indicator variable $I_{i,n}$ as

$$I_{i,n} := \mathbb{1}_{\{A_i \cap [n] \neq \emptyset\}}.$$

The total number of blocks given for the restricted partition is then given by

$$N_n = \sum_i I_{i,n}.$$

Moreover, for every $i \geq 1$,

$$\mathbb{E}[I_{i,n} \mid P] = \mathbb{P}(A_i \cap [n] \neq \emptyset \mid P) = \mathbb{P}(\exists j \in [n] : U_j \in I_i \mid P) = 1 - (1 - P_i)^n.$$

By the linearity of expectation, we can sum over all blocks i to find the total expected number of blocks in the restriction $\Pi_{|[n]}$:

$$\mathbb{E}[N_n \mid P] = \sum_i \mathbb{E}[I_{i,n} \mid P] = \sum_i (1 - (1 - P_i)^n).$$

By the Tower property

$$\mathbb{E}[N_n] = \mathbb{E}[\mathbb{E}[N_n \mid P]] = \sum_i \mathbb{E}[1 - (1 - P_i)^n].$$

Let $f_n(x) = 1 - (1 - x)^n$ for $x \in (0, 1]$. Then, by (3.7), for X the mass of a size-biased picked block, we have

$$\mathbb{E}[N_n] = \mathbb{E}\left[\sum_i (1 - (1 - P_i)^n)\right] = \mathbb{E}\left[\frac{f_n(X)}{X}\right] = \mathbb{E}\left[\frac{1 - (1 - X)^n}{X}\right]. \tag{3.8}$$

Moreover, we have

$$0 < N_1 \leq N_2 \leq N_3 \leq \cdots \quad \text{and} \quad \lim_{n \to \infty} N_n = N.$$

By the Monotone Convergence Theorem, it follows that

$$\mathbb{E}(N) = \lim_{n \to \infty} \mathbb{E}(N_n).$$

Let $g_n(x) = \frac{1-(1-x)^n}{x}$ for $x \in (0,1]$. Since we have no dust $X > 0$ a.s. and therefore $g_n(X) \geq 0$. Moreover, $g_1(X) \leq g_2(X) \leq \cdots$ and $\lim_{n\to\infty} g_n(X) = \frac{1}{X}$ a.s. A second application of the Monotone Convergence Theorem yields

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{1-(1-X)^n}{X}\right] = \lim_{n\to\infty} \mathbb{E}[g_n(X)] = \mathbb{E}\left[\frac{1}{X}\right].$$

Thus, by (3.8) and by taking $n \to \infty$,

$$\mathbb{E}[N] = \mathbb{E}\left[\frac{1}{X}\right].$$

■

By applying size-biased sampling without replacement recursively to the frequencies of a mass partition $\mathbf{s} \in \mathcal{S}_0$, we obtain a size-biased reordering of the terms of the initial mass partition $\mathbf{s}$. Focusing on mass-partitions with no dust, we arrive at the following definition:

Definition 3.29 (Size-biased reordering)**.** Let $\mathbf{s} = (s_1, s_2, \ldots) \in \mathcal{S}_0$ with $s_0 = 0$ almost surely. Let $\sigma : \mathbb{N} \to \mathbb{N}$ be a random permutation whose distribution satisfy

$$\mathbb{P}(\sigma(1) = j) = s_j.$$

Moreover, conditional on the previously chosen indices $\sigma(1), \ldots, \sigma(i-1)$, we have

$$\mathbb{P}(\sigma(i) = j \mid \sigma(1), \ldots, \sigma(i-1)) = \frac{s_j}{1 - s_{\sigma(1)} - \cdots - s_{\sigma(i-1)}}.$$

The random sequence $\mathbf{s}^* = (s_{\sigma(1)}, s_{\sigma(2)}, \ldots)$ is called a size-biased reordering of $\mathbf{s}$.

Now, let Π be an exchangeable random partition with no dust almost surely, corresponding to a random mass-partition $\mathbf{S} = (S_1, S_2, \ldots) \in \mathcal{S}_0$. We order the blocks of Π according to their least elements. Thus, Π_1 is the block containing the integer 1 with asymptotic frequency $|\Pi_1|$. Conditioning on $\mathbf{S}$, the uniform random variable U_1 falls into a segment of length S_i with probability S_i. Consequently, the conditional probability that the first block has mass S_i is

$$\mathbb{P}(|\Pi_1| = S_i \mid \mathbf{S}) = S_i.$$

Proceeding recursively, Π_2 is the block containing the smallest integer $k \notin \Pi_1$, with mass $|\Pi_2|$. Conditional on $\mathbf{S}$ and on the first selection $|\Pi_1| = S_i$, the probability that U_k falls into a remaining segment S_j, with $j \neq i$, is the ratio of its mass to the total remaining mass

$$\mathbb{P}(|\Pi_2| = S_j \mid \mathbf{S}, |\Pi_1| = S_i) = \frac{S_j}{1 - S_i}.$$

Iterating this argument over all blocks shows that the sequence of asymptotic frequencies ordered by their least elements $(|\Pi_1|, |\Pi_2|, \ldots)$ is a size-biased reordering of $\mathbf{S}$.

3.2 Poisson-Dirichlet random partitions

The Poisson-Dirichlet (PD) distribution defines an important two-parameter family of random partitions that arise in a wide range of fields. We denote the two parameters by α and θ. In this section, we analyze the cases $\alpha = 0$ or when $\theta = 0$. We explore three distinct methods for constructing these partitions:

- The stick-breaking construction;
- The Chinese restaurant process (CRP);
- The Poisson random measure construction.

The following continuous distribution will play an important role in the constructions of Poisson-Dirichlet random variables:

Definition 3.30 (Beta distribution)**.** A random variable X is said to follow the beta distribution with parameters $a, b > 0$, denoted by $X \sim \text{Beta}(a, b)$, if

$$\frac{\mathbb{P}(X \in dx)}{dx} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, x^{a-1}(1-x)^{b-1}, \qquad \text{for} \quad 0 < x < 1, \tag{3.9}$$

where Γ is the gamma function (Definition A.27).

3.2.1 $\alpha = 0,\ \theta > 0$

Definition 3.31 (Stick-breaking construction)**.** Let $(W_i)_{i \geq 1}$ be an i.i.d. sequence of random variables with distribution

$$W_i \sim \text{Beta}(1, \theta).$$

Using these variables, define a sequence of random proportions by

$$P_1 = W_1, \qquad P_2 = (1 - P_1)W_2, \qquad \ldots, \qquad P_{n+1} = \left(1 - \sum_{j=1}^{n} P_j\right) W_{n+1}. \tag{3.10}$$

This construction is known as the stick-breaking construction.

The name of this construction comes from a simple analogy. Imagine starting with a stick of unit length. We cut off a random portion W_1, where $W_1 \sim \text{Beta}(1, \theta)$. We get the first piece of length $P_1 = W_1$. From the remaining stick, which is now of length $1 - P_1$, we break off another random fraction, W_2, which is independent of the first fraction and has the same distribution. This produces the second piece, which has a length of $P_2 = (1 - P_1)W_2$. We can continue like this.

To see that a reordering of these proportions yields a mass-partition with no dust almost surely, we can show by induction that the sum of the first n proportions is given by

$$\sum_{i=1}^{n} P_i = 1 - \prod_{j=1}^{n}(1 - W_j).$$

Because $W_j \sim \text{Beta}(1, \theta)$, a straightforward calculation gives $\mathbb{E}[\ln(1 - W_j)] = -1/\theta$. By the Strong Law of Large Numbers, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \ln(1 - W_j) = -\frac{1}{\theta} \quad \text{a.s.}$$

Hence, $\sum_{j=1}^{n} \ln(1 - W_j)$ must diverge to $-\infty$ almost surely. Consequently, taking the exponential gives

$$\lim_{n\to\infty} \prod_{j=1}^{n}(1 - W_j) = \lim_{n\to\infty} \exp\left(\sum_{j=1}^{n} \ln(1 - W_j)\right) = 0 \quad \text{a.s.}$$

This guarantees that $\sum_{i=1}^{\infty} P_i = 1 - \lim_{n\to\infty} \prod_{j=1}^{n}(1 - W_j) = 1$ almost surely.

Definition 3.32 (PD random variables, PD random partitions, [5],[6])**.** Let $(S_i)_{i\geq 1}$ be a random mass-partition with $\text{PD}(0, \theta)$ law. Let $(S_i^*)_{i\geq 1}$ be its size-biased reordering, often called a Griffiths-Engen-McCloskey sequence with parameter θ, denoted as $\text{GEM}(\theta)$. Then,

$$(P_i)_{i\geq 1} \overset{d}{=} (S_i^*)_{i\geq 1}. \tag{3.11}$$

By applying the paintbox construction to the random mass-partition $(S_i)_{i\geq 1}$, we obtain a Poisson-Dirichlet partition Π. We write

$$\Pi \sim \text{PD}(0, \theta).$$

The stick-breaking construction gives us the asymptotic frequencies of the blocks of a PD partition. However, the next construction, known as the *Chinese restaurant process*, provides a more visual and intuitive explanation of this partition.

Chinese restaurant process [29]: Imagine a Chinese restaurant with an infinite number of round tables, each with an infinite number of seats. Initially, all the tables are empty. People enter the restaurant one by one and join either an occupied table or a new one. Each customer is indexed according to their chronological order of entry.

Consider a configuration of N customers already seated in the restaurant. For each individual $i \in \{1, \ldots, N\}$, let z_i denote the label of the table to which they are assigned. The seating arrangement is then represented by the assignment vector $\mathbf{z} = (z_1, \ldots, z_N)$. By convention, the first customer, labeled 1, occupied table 1, thereby setting the initial state as $z_1 = 1$. Furthermore, each time a new table is

opened, it is assigned the next available integer, keeping the tables also numbered chronologically.

Let n_j be the number of visitors sitting at table j for $1 \leq j \leq k$ for k the number of non-empty tables. The seating arrangement is thus characterized by the occupancy vector $\mathbf{n} = (n_1, n_2, \ldots, n_k)$, where $\sum_{j=1}^{k} n_j = N$. The probability that the customer $N+1$ is sitting at table j is

$$\mathbb{P}(z_{N+1} = j \mid \mathbf{n}) = \frac{n_j}{N+\theta}, \tag{3.12}$$

where $\theta > 0$ is called *dispersion parameter*. Alternatively, the $(N+1)$-th customer can start a new table with label $k+1$. This happens with probability

$$\mathbb{P}(z_{N+1} = k+1 \mid \mathbf{n}) = \frac{\theta}{N+\theta}. \tag{3.13}$$

The induced random partition may be described by the equivalence relation:

$$i \sim j \iff \text{customers } i \text{ and } j \text{ are seated at the same table.}$$

This procedure clearly gives us a consistent family of partitions $(\Pi_N)_{N\geq 1}$. Π_1 is the trivial partition $(\{1\})$. Given a partition Π_N, the next partition Π_{N+1} is obtained by assigning customer $N+1$ either to one of the existing blocks or to a new block. The restriction of Π_{N+1} to $[N]$ then recovers Π_N. By Lemma 3.5, there exists a unique random partition $\Pi \in \mathcal{P}_\infty$ such that $\Pi_{|[N]} = \Pi_N$ for all $N \geq 1$.

Theorem 3.33. *The random partition Π generated by the CRP with dispersion parameter $\theta > 0$ is a Poisson-Dirichlet random partition with parameters $(0, \theta)$. In particular, Π is exchangeable and the size-biased reordering of its asymptotic block frequencies follows the stick-breaking representation (3.10).*

The proof of this theorem relies on a construction involving a classical urn model and the application of a preliminary result stated below as Theorem 3.35.

Definition 3.34 (Pólya-Eggenberger urn, [26])**.** Consider an urn containing balls of two colors, red and black. Initially, the urn contains x_0 red ball and b_0 black balls. At each step, a ball is drawn uniformly at random from the urn and then returned together with s additional balls of the same color.

Theorem 3.35 ([26])**.** *Let $Y_n = (X_n, B_n)$ represent the composition of the Pólya-Eggenberger urn after the n-th draw, where X_n is the number of red balls and B_n the number of black balls in the urn. The initial composition of the urn is given by $Y_0 = (x_0, b_0) \in \mathbb{N}^2$. Then the proportion of red balls in the urn converges in distribution to a Beta distribution:*

$$\frac{X_n}{n} \xrightarrow{d} \text{Beta}(x_0, b_0). \tag{3.14}$$

Proof of Theorem 3.33. Consider the Pólya-Eggenberger urn with initially, at stage $n = 1$, one red balls and θ black balls and take $s = 1$ (see Definition 3.34).

In the CRP consider the block containing 1 (table at which customer 1 is sitting), say $1 \in B_1$. We assign to each customer $1 \leq i \leq n$ a ball in an urn, and i is red if $i \sim 1$, that is, if i is in the same block as 1, and black otherwise. If we have already r integers ($r \geq 1$) in B_1 at stage n, then as the new integer $n+1$ is added to the partition it

$$\begin{cases} \text{joins } B_1 \text{ with probability: } \dfrac{r}{n+\theta}, \\ \text{does not join } B_1 \text{ with probability: } 1 - \dfrac{r}{n+\theta} = \frac{n+\theta-r}{n+\theta}. \end{cases}$$

This is the same as saying that we have r red balls and $n+\theta-r$ black balls after having drawn $n-1$ times from our urn. We then choose a ball from the urn uniformly at random. This new ball either is red (is in B_1) with probability $r/(n+\theta)$ or black (is not B_1) with probability $(n+\theta-r)/(n+\theta)$.

Let us define $X_n(1)$ to be the number of red balls at stage n. By Theorem 3.35,

$$|B_1| = \frac{X_n(1)}{n} \overset{d}{\Rightarrow} W_1 \quad \text{for } W_1 \sim \text{Beta}(1, \theta).$$

Next, we exclude block B_1. Among the remaining customers, the first not being in B_1 starts B_2, say customer i_2. We apply again the Pólya-urn model with 1 red ball and θ black balls at stage $n = 1$ for B_2. Letting $X_n(2)$ be the number of red balls at stage n (customers at table B_2), Theorem 3.35 implies

$$\frac{X_n(2)}{n} \overset{d}{\Rightarrow} W_2 \quad \text{for } W_2 \sim \text{Beta}(1, \theta).$$

Thus, $W_2 \sim \text{Beta}(1, \theta)$ represents the limiting frequency of B_2 within the complement of B_1. Scaling this by the measure of the remaining customers, we find the total mass of the second block to be

$$|B_2| = (1 - W_1)W_2.$$

We can continue in the same way for B_3, obtaining

$$\frac{X_n(3)}{n} \overset{d}{\Rightarrow} W_3 \quad \text{for } W_3 \sim \text{Beta}(1, \theta).$$

Hence,

$$|B_3| = (1 - |B_1| - |B_2|)\, W_3 = (1 - W_1 - (1 - W_1)W_2)\, W_3 = (1 - W_1)(1 - W_2)\, W_3.$$

We can continue inductively.

It remains to prove the exchangeability of the random partition Π induced by the CRP. To this end, we define

$$\xi_n \coloneqq \mathbb{1}_{\{n \in B_1\}}, \quad n \in \mathbb{N},$$

that is,

$$\xi_n = \begin{cases} 1, & \text{if customer } n \text{ is in the block } B_1, \\ 0, & \text{otherwise.} \end{cases}$$

We have seen that there exists a random variable $W \sim \mathrm{Beta}(1, \theta)$, which describes the asymptotic frequency of customers seated at the first table B_1. Therefore, conditional on W, the random variables $(\xi_n)_{n \geq 0}$ are i.i.d. $\mathrm{Ber}(W)$. Thus, for any $x_1, \ldots, x_n \in \{0, 1\}$ and any permutation σ of $[n]$,

$$\begin{aligned} \mathbb{P}(\xi_1 = x_1, \ldots, \xi_n = x_n \mid W) &= W^{\sum_i x_i}(1 - W)^{n - \sum_i x_i} \\ &= \mathbb{P}(\xi_{\sigma(1)} = x_1, \ldots, \xi_{\sigma(n)} = x_n \mid W). \end{aligned}$$

Taking expectations over W preserves this equality, and therefore

$$\mathbb{P}(\xi_1 = x_1, \ldots, \xi_n = x_n) = \mathbb{P}(\xi_{\sigma(1)} = x_1, \ldots, \xi_{\sigma(n)} = x_n).$$

Since the joint distribution is invariant under permutations for any finite n, the infinite sequence is exchangeable. The preceding argument can be applied to every block. Hence, we conclude that the full CRP partition is exchangeable.

Moreover, as discussed in the previous chapter on size-biased sampling, ordering the blocks of an exchangeable random partition according to their least elements and considering the corresponding frequencies yields a size-biased reordering of the associated random mass-partition. Therefore, the size-biased reordering of the frequencies of Π coincides with the stick-breaking representation (3.10).

To conclude, we have all the requirements to use Kingman's correspondence. Indeed, Π constructed by the CRP is exchangeable. Moreover, we have shown that $(|B_i|)_{i \geq 1} \overset{d}{=} (P_i)_{i \geq 1} \overset{d}{=} (S_i^*)_{i \geq 0}$, where $(S_i^*)_{i \geq 1}$ is the size-biased reordering of a random mass-partition $(S_i)_{i \geq 1}$ following the $\mathrm{PD}(0, \theta)$ law. By the uniqueness, it follows that Π is a Poisson-Dirichlet random partition with parameter θ. ■

The construction of the CRP leads directly to an expression for the distribution of the random partition Π_n.

Theorem 3.36 (Ewens' sampling formula, $\alpha = 0$)**.** *Let π be any partition of $[n]$ with block sizes $n_1, n_2, \ldots, n_k$. Then,*

$$\mathbb{P}(\Pi_n = \pi) = \frac{\theta^k}{(\theta) \cdots (\theta + n - 1)} \prod_{i=1}^{k} (n_i - 1)!. \tag{3.15}$$

Proof. We proceed by induction on n.

Base case: $n = 1$**:**
We have $\Pi_1 = (\{1\})$. In this case $k = 1$ and $n_k = 1$. Hence,

$$\mathbb{P}(\Pi_1 = \pi) = \frac{\theta^1}{\theta}(1-1)! = 1.$$

Induction assumption (IA):
Assume the formula holds for n. We now prove that it holds for $n+1$.

Induction step:
Suppose π' is a partition of $[n+1]$ with block sizes $n'_1, n'_2, \ldots, n'_{k'}$, which is obtained from a partition π of $[n]$ with block sizes $n_1, \ldots, n_k$. There are two ways in which $n+1$ can be added to the partition π.

- $n+1$ starts a new block:
  The partition of π' is formed by taking π and adding $n+1$ as a singleton block.
  Then, $k' = k+1$, $n'_i = n_i$ for $1 \le i \le k$ and $n'_{k+1} = 1$. Accordingly,

$$\begin{aligned}
\mathbb{P}(\Pi_{n+1} = \pi') &= \mathbb{P}(\Pi_n = \pi)\ \mathbb{P}((n+1) \text{ starts new block} \mid \Pi_n = \pi) \\
&= \mathbb{P}(\Pi_n = \pi)\ \frac{\theta}{n+\theta} \\
&\overset{(IA)}{=} \left[\frac{\theta^k}{\theta(\theta+1)\cdots(\theta+n-1)} \prod_{i=1}^{k}(n_i-1)!\right] \frac{\theta}{n+\theta} \\
&= \frac{\theta^{k+1}}{\theta(\theta+1)\cdots(\theta+n)} \prod_{i=1}^{k}(n_i-1)!\ (1-1)!,
\end{aligned}$$

  which gives us exactly what we want since $(1-1)! = 0! = 1$.

- $n+1$ joins an existing block:
  Suppose $n+1$ joins block j, which has size n_j in π. In the updated partition π', the number of blocks remains $k' = k$, but the size of the j-th block increases to $n'_j = n_j + 1$. The resulting block sizes are: $n_1, \ldots, n_j+1, \ldots, n_k$.
  Therefore,

$$\prod_{i=1}^{k}(n'_i-1)! = (n'_j-1)! \prod_{i\neq j, i=1}^{k}(n_i-1)! = (n_j)! \prod_{i\neq j, i=1}^{k}(n_i-1)!.$$

Hence, we can express the probability as

$$\begin{aligned}
\mathbb{P}(\Pi_{n+1} = \pi') &= \mathbb{P}(\Pi_n = \pi)\, \mathbb{P}\big(n+1 \text{ joins block } j \mid \Pi_n = \pi\big) \\
&= \mathbb{P}(\Pi_n = \pi)\, \frac{n_j}{n+\theta} \\
&\overset{\text{(IA)}}{=} \left[\frac{\theta^k}{\theta(\theta+1)\cdots(\theta+n-1)} \prod_{i=1}^{k} (n_i - 1)!\right] \frac{n_j}{n+\theta} \\
&= \frac{\theta^k}{\theta(\theta+1)\cdots(\theta+n)} \left[(n_j) \prod_{i=1}^{k} (n_i - 1)!\right] \\
&= \frac{\theta^k}{\theta(\theta+1)\cdots(\theta+n)} \left[(n_j)(n_j - 1)! \prod_{i\neq j, i=1}^{k} (n_i - 1)!\right] \\
&= \frac{\theta^k}{\theta(\theta+1)\cdots(\theta+n)} \prod_{i=1}^{k} (n_i' - 1)!,
\end{aligned}$$

which matches the desired form.

■

3.2.2 $0 < \alpha < 1$, $\theta = 0$

Definition 3.37 (Stick-breaking construction, $\theta = 0$)**.** Let $(W_i)_{i\geq 1}$ be an i.i.d. sequence of random variables with distribution

$$W_i \sim \text{Beta}(1-\alpha, i\alpha).$$

Using these variables, define a sequence of random proportions by

$$P_1 = W_1, \qquad P_2 = (1-P_1)W_2, \qquad \ldots, \qquad P_{n+1} = \left(1 - \sum_{j=1}^{n} P_j\right) W_{n+1}. \quad (3.16)$$

This sequence is distributed as the size-biased reordering of a mass-partition $(S_i)_{i\geq 1}$ with $\text{PD}(\alpha, 0)$ law. By applying the paintbox construction to the random mass-partition $(S_i)_{i\geq 1}$, we obtain a Poisson-Dirichlet partition Π. We write

$$\Pi \sim \text{PD}(\alpha, 0).$$

As in the case $\alpha = 0$, there is again a *Chinese restaurant process*, that describes and visualizes the previos definition. However, some modifications are required. The assumptions of the restaurant stay the same and we will use the same notations.

Chinese restaurant process construction: Let Π_N have k blocks with block size $n_1, \ldots, n_k$. Customer $N+1$ has two possibilities. Either join an already occupied table or sit at an unoccupied table. These two possibilities occur with the following probabilities:

$$\begin{cases} \text{join a table with } n_j \text{ customers:} & \mathbb{P}(z_{N+1} = j \mid \mathbf{n}) = \dfrac{n_j - \alpha}{N}, \quad 1 \leq j \leq k, \\ \text{start a new table:} & \mathbb{P}(z_{N+1} = k+1 \mid \mathbf{n}) = \dfrac{k\alpha}{N}. \end{cases} \tag{3.17}$$

Again, as a consequence of the construction, Ewens' sampling formula takes the following form in this setting:

Theorem 3.38 (Ewens' sampling formula, $\theta = 0$)**.** *Let π be any partition of $[n]$ with block sizes $n_1, n_2, \ldots, n_k$. Then,*

$$\mathbb{P}(\Pi_n = \pi) = \frac{\alpha^{k-1}(k-1)!}{(n-1)!} \prod_{i=1}^{k} (1-\alpha) \cdots (n_i - \alpha). \tag{3.18}$$

3.2.3 Poisson random measures and Poisson-Dirichlet laws

We want to introduce a third construction of a Poisson-Dirichlet random variable based on Poisson random measures. To establish this framework, we begin with the necessary formal definitions and results, assuming a standard underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 3.39 ([24])**.** Let $(E, \mathcal{E})$ be a measurable space, where E is a Polish space and $\mathcal{E}$ denotes the Borel σ-algebra on E. Let $\mathsf{N}_{<\infty}(E) \equiv \mathsf{N}_{<\infty}$ denote the space of all measures μ on E such that $\mu(B) \in \mathbb{N}_0$ for all $B \in \mathcal{E}$. We denote by $\mathsf{N}(E) \equiv \mathsf{N}$ the space of all measures that can be written as a countable sum of measures from $\mathsf{N}_{<\infty}$.

Definition 3.40 ([24])**.** Define $\mathcal{N}(E) \equiv \mathcal{N}$ to be the σ-algebra generated by the sets of the form $\{\mu \in \mathsf{N} : \mu(B) = k\}$ for $B \in \mathcal{E}$, $k \in \mathbb{N}_0$. This means, that $\mathcal{N}$ is the smallest σ-algebra on N, that makes the maps

$$\mu \mapsto \mu(B)$$

measurable for all $B \in \mathcal{E}$.

Definition 3.41 (Point process/Point random measure, [24])**.** A point process (PP) on E is a measurable map

$$\mathcal{M} : (\Omega, \mathcal{F}) \to (\mathsf{N}, \mathcal{N}).$$

For $\mathcal{M}$ a PP on E and $B \in \mathcal{E}$, we denote by $\mathcal{M}(B)$ the mapping $\omega \mapsto \mathcal{M}(\omega, B) \coloneqq \mathcal{M}(\omega)(B)$. By the definition of $\mathcal{M}$ and the σ-algebra, we have that this are random variables with values in $\mathbb{N}_0 \cup \{\infty\}$, that is,

$$\{\mathcal{M}(B) = k\} = \{w \in \Omega : \mathcal{M}(\omega, B) = k\} \in \mathcal{F}, \quad B \in \mathcal{E}, k \in \mathbb{N}_0 \cup \{\infty\}. \tag{3.19}$$

On the other hand, if (3.19) is satisfies for a mapping $\mathcal{M} : (\Omega, \mathcal{F}) \to (\mathsf{N}, \mathcal{N})$, $\mathcal{M}$ is a PP. $\mathcal{M}(B)$ is called number of points of $\mathcal{M}$ in B.

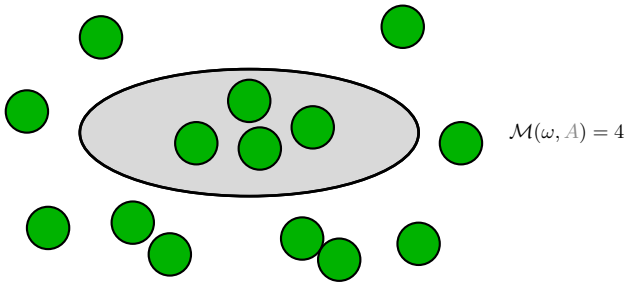


Figure 3.3: Point process on $\mathbb{R}^2$ with $A \subset \mathbb{R}^2$.

Definition 3.42 (Distribution of a point process, [9])**.** The distribution (or law) of $\mathcal{M}$ is the probability measure $\mathbb{P}_{\mathcal{M}}$ on $(\mathsf{N}, \mathcal{N})$ defined by

$$\mathbb{P}_{\mathcal{M}}(\Gamma) \coloneqq \mathbb{P}(\mathcal{M} \in \Gamma), \qquad \Gamma \in \mathcal{N}.$$

Definition 3.43 (Proper PP, [24])**.** A PP $\mathcal{M}$ is called proper if there exist random elements $X_1, X_2, \ldots$ in E and a random variable k in $\mathbb{N}_0 \cup \{\infty\}$ such that

$$\mathcal{M} = \sum_{i=1}^{\kappa} \delta_{X_i}.$$

These random variables $X_1, X_2, \ldots$ are often called the atoms of $\mathcal{M}$. Moreover, if $\kappa = 0$, $\mathcal{M}$ is the zero measure on E.

Definition 3.44 (Intensity measure, [24])**.** The intensity measure μ of a PP $\mathcal{M}$ on E is defined by

$$\mu(B) \coloneqq \mathbb{E}[\mathcal{M}(B)].$$

for $B \in \mathcal{E}$.

Definition 3.45 (Poisson random measure/ Poisson point process, [6])**.** Let μ be a σ-finite measure on E. We call a PP $\mathcal{M}$ on E a Poisson random measure (PRM) with intensity μ if it fulfills the following two properties:

(i) For every Borel subset B of E with $\mu(B) < \infty \Rightarrow \mathcal{M}(B) \sim \text{Poisson}(\mu(B))$.

(ii) If $B_1, \ldots, B_n$ are disjoint Borel sets $\Rightarrow \mathcal{M}(B_1), \ldots, \mathcal{M}(B_n)$ are independent.

Recall ([25])**.** A measure μ on E is called σ-finite if there exists a sequence $(E_n)_{n\in\mathbb{N}}$ of measurable sets such that $E = \bigcup_{n\in\mathbb{N}} E_n$ and $\mu(E_n) < \infty$ for every n.

Remark 3.46 ([2])**.** There is often terminological confusion regarding the concepts of a *Poisson random measure* and a *Poisson counting process*, as they are closely related but not identical. A Poisson counting process $(N_t)_{t\geq 0}$ can be formally defined through a Poisson random measure $\mathcal{M}$ on $E = \mathbb{R}_+$. Specifically, the process at time t is given by the measure of the interval from the origin to t:

$$N_t = \mathcal{M}([0, t]).$$

Remark 3.47 (Representation as a proper process, [24])**.** Let $\mathcal{M}$ be PRM with intensity measure μ on $(E, \mathcal{E})$. Then there exists a random variable $\kappa \in \mathbb{N} \cup \{\infty\}$ and atoms $(X_n)_{n\geq 1}$ with $X_n \in E$ such that

$$\mathcal{M} = \sum_{n=1}^{\kappa} \delta_{X_n}.$$

Lemma 3.48 (Disjointness Lemma, [22])**.** *Let $\mathcal{M}$ and $\mathcal{M}'$ be two independent PRM on E. $\mathcal{M}$ has intensity μ and $\mathcal{M}'$ has intensity μ'. Let A be a Borel set such that $\mu(A) < \infty$ and $\mu'(A) < \infty$. Then $\mathcal{M}(A)$ and $\mathcal{M}'(A)$ are disjoint almost surely on A.*

Proof. For the proof see [22] on page 15. ■

Definition 3.49 (Homogeneity, [22])**.** Let $E = \mathbb{R}^d$ for $d \geq 1$. A PRM $\mathcal{M}$ is said to be homogeneous (or stationary) if its intensity measure μ is a multiple of the Lebesgue measure on $\mathbb{R}^d$, that is, there exists a $\lambda \in (0, \infty)$ such that

$$\mu(B) = \lambda|B|$$

for every $B \in \mathcal{B}(\mathbb{R}^d)$.

Theorem 3.50 (Superposition Theorem, [6])**.** *Let μ and μ' be two σ-finite measures on E. Let $\mathcal{M}$ be a PRM with intensity μ, and let $\mathcal{M}'$ be a PRM with intensity μ' independent of $\mathcal{M}$. Then $\mathcal{M} + \mathcal{M}'$ is a PRM with intensity $\mu + \mu'$.*

Proof. Let $\mathcal{M}(A)$ be the number of atoms of $\mathcal{M}$ in the Borel set A. The same holds for $\mathcal{M}'(A)$. If $\mu(A) < \infty$ and $\mu'(A) < \infty$, by the Disjointness Lemma, $\mathcal{M}$ and $\mathcal{M}'$ are disjoint on the set A almost surely. Therefore,

$$(\mathcal{M} + \mathcal{M}')(A) = \mathcal{M}(A) + \mathcal{M}'(A).$$

Then by Lemma A.2, we have $(\mathcal{M} + \mathcal{M}')(A) \sim \text{Poisson}(\mu(A) + \mu'(A))$.

Now, we are left to show the independence of $(\mathcal{M}+\mathcal{M}')(A_1), \ldots, (\mathcal{M}+\mathcal{M}')(A_k)$ for $A_1, \ldots, A_k$ disjoint Borel sets. Since $\mathcal{M}$ is a PRM, $\mathcal{M}(A_1), \ldots, \mathcal{M}(A_k)$ are independent. The same is true for $\mathcal{M}'$. Additionally, $\mathcal{M}$ and $\mathcal{M}'$ are assumed to be independent of each other. Hence, $\mathcal{M}(A_j)$ is independent of $\mathcal{M}'(A_l)$ for every $1 \leq j \leq k$ and $1 \leq l \leq k$. Therefore, $(\mathcal{M} + \mathcal{M}')(A_j)$ is independent of $(\mathcal{M} + \mathcal{M}')(A_i)$ for $i \neq j$. And we can conclude that $(\mathcal{M} + \mathcal{M}')(A_1), \ldots, (\mathcal{M} + \mathcal{M}')(A_k)$ are independent. ■

Remark 3.51 ([24])**.** The Superposition Theorem is also true for a countable number of independent PRMs.

Theorem 3.52 (Existence Theorem, [24])**.** *Let μ be a σ-finite measure on E. Then there exists a PRM on E with intensity μ.*

Proof. If $\mu(E) < \infty$, we have finitely many atoms almost surely. In this case, the number of atoms $\kappa := \mathcal{M}(E)$ can be chosen to be Poisson distributed with parameter $\mu(E)$. We can choose the atoms X_i for $i \in I = \{1, 2, \dots, \kappa\}$ taking values in E as i.i.d. random variables with common distribution $\mu(\cdot)/\mu(E)$ independent of κ. Take $\mathcal{M}(B) = \sum_{i=1}^{N} \mathbb{1}_{\{X_i \in B\}}$ for B a Borel subset of E.

Let $B_1, \dots, B_m$ be disjoint Borel sets in E. Without loss of generality, we may assume $\bigcup_{i=1}^{m} B_i = E$. If this is not the case, we simply adjoin the complement of the union to cover E. Let $k_1, \dots, k_m \in \mathbb{N}_0$, and set $l := k_1 + \cdots + k_m$. Then,

$$
\begin{aligned}
&\mathbb{P}(\mathcal{M}(B_1) = k_1, \dots, \mathcal{M}(B_m) = k_m) \\
&\qquad = \mathbb{P}(\kappa = l)\, \mathbb{P}(\mathcal{M}(B_1) = k_1, \dots, \mathcal{M}(B_m) = k_m \mid \kappa = l) \\
&\qquad \overset{(A.1)}{=} e^{-\mu(E)} \frac{\mu(E)^l}{l!} \frac{l!}{k_1! \cdots k_m!} \left(\frac{\mu(B_1)}{\mu(E)}\right)^{k_1} \cdots \left(\frac{\mu(B_m)}{\mu(E)}\right)^{k_m} \\
&\qquad = \prod_{j=1}^{m} \frac{\left(\mu(E)\frac{\mu(B_j)}{\mu(E)}\right)^{k_j}}{k_j!} e^{-\mu(E)\frac{\mu(B_j)}{\mu(E)}} \\
&\qquad = \prod_{j=1}^{m} \frac{(\mu(B_j))^{k_j}}{k_j!} e^{-\mu(B_j)}.
\end{aligned}
$$

If we sum over $k_2, \dots, k_m$, we get that $\mathcal{M}(B_1)$ is Poisson distributed with parameter $\mu(B_1)$. The same statement applies to $\mathcal{M}(B_2), \dots, \mathcal{M}(B_m)$. Therefore, we can conclude that $\mathcal{M}(B_1), \dots, \mathcal{M}(B_m)$ are independent.

If $\mu(E) = \infty$, there are almost surely infinitely many atoms. We can write $E = \bigcup_{i=1}^{\infty} E_i$ as a disjoint union where $\mu(E_i) < \infty$. In this way, we are able to construct a PRM as we did for the case $\mu(E) < \infty$. We take $\mathcal{M}_i$ to be the PRM with intensity μ restricted on E_i, i.e., $\mu_{|E_i}(A) := \mu(A \cap E_i)$ for A a Borel set. By making the $\mathcal{M}_i's$ independent we can use the Superposition Theorem and conclude that $\mathcal{M} = \sum_{i \geq 1} \mathcal{M}_i$ is a PRM on E with intensity μ. ■

Theorem 3.53 (Mapping Theorem, [24])**.** *Let E and E' be two Polish spaces and let $f : E \to E'$ be a measurable map. Let $\mathcal{M}$ be a PRM on E with intensity μ. Then $f(\mathcal{M}) = \mathcal{M} \circ f^{-1}$ is a PRM with intensity $\mu \circ f^{-1}$ (also called the pushforward measure of μ).*

Remark 3.54 ([24])**.** For $\mathcal{M} = \sum_{i=1}^{\kappa} \delta_{X_i}$, $f(\mathcal{M}) = \sum_{i=1}^{\kappa} \delta_{f(X_i)}$.

Proposition 3.55 (Uniqueness, [24]). *Let $\mathcal{M}$ and $\mathcal{M}'$ be two PRMs on E with the same σ-finite intensity measure μ. Then,*

$$\mathcal{M} \overset{d}{=} \mathcal{M}'.$$

Theorem 3.56. *(Restriction Theorem, [24]) Let $\mathcal{M}$ be a PRM with intensity measure μ on E. Let $C_1, C_2, \ldots \in \mathcal{E}$ be disjoint measurable sets. The restricted random measures*

$$\mathcal{M}_{|C_1},\ \mathcal{M}_{|C_2}, \ldots$$

are independent Poisson random measures with respective intensity measures

$$\mu_{|C_1},\ \mu_{|C_2}, \ldots,$$

where $\mu_{|C_i} \coloneqq \mu(\cdot \cap C_i)$.

Construction of a PRM on the real line [22]: We now consider a particularly important case where the state space is $E = \mathbb{R}$. The linear structure of the real line is advantageous because it allows for a natural ordering of the atoms. Throughout, we restrict ourselves to the homogeneous case. Therefore, let $\mathcal{M}$ be a Poisson random measure on $\mathbb{R}$ with intensity measure

$$\mu(dx) = \lambda\, dx,$$

where $\lambda \in (0, \infty)$ is a constant. In this case, we say that $\mathcal{M}$ is a *PRM with a constant rate* λ. $\mathcal{M}$ has infinitely many points in $(0, \infty)$ almost surely. Consequently, since we are working on $\mathbb{R}$, the atoms of $\mathcal{M}$ in $(0, \infty)$ can be indexed in increasing order as

$$0 < X_1 < X_2 < \cdots.$$

Similarly, we can look at atoms in $(-\infty, 0)$

$$\cdots < X_{-2} < X_{-1} < 0.$$

The point 0 is just taken as a reference point for ordering and will not create problems, since $\mu(\{0\}) = 0$, which implies that $\mathbb{P}(\mathcal{M}(\{0\}) = 0) = 1$. Therefore, we can represent the measure $\mathcal{M}$ as the sum of Dirac masses at these ordered atoms:

$$\mathcal{M} = \sum_{i \in \mathbb{Z}\setminus\{0\}} \delta_{X_i}.$$

Moreover, $\mathcal{M}$ can be decomposed as

$$\mathcal{M} = \mathcal{M}_- + \mathcal{M}_+,$$

where

$$\mathcal{M}_+ \coloneqq \mathcal{M}_{|(0,\infty)}, \qquad \mathcal{M}_- \coloneqq \mathcal{M}_{|(-\infty,0)}$$

are two independent Poisson random measures by the Restriction Theorem.

Consider $f(x) = -x$. By the Mapping Theorem, $f(\mathcal{M})$ is also a PRM. Since the intensity measure of the original process is $\mu(dx) = \lambda\, dx$, and the Lebesgue measure is invariant under reflection ($dx = d(-x)$), the intensity of the transformed process is still $\lambda\, dx$. By the existence and uniqueness of PRMs with a given intensity, we have $f(\mathcal{M}) \overset{d}{=} \mathcal{M}$. This distributional invariance implies that $f(\mathcal{M}_-)$ is indistinguishable from the positive process $\mathcal{M}_+$.

Notation 3.57 ([6])**.** Consider $f : E \to \mathbb{R}$ a measurable function and $\mathcal{M}$ a PRM. We write

$$\langle \mathcal{M}, f \rangle \coloneqq \int_E f(x)\, \mathcal{M}(dx) = \sum_{i \in I} f(X_i).$$

The representation as a sum follows from:

$$\int_E f(x)\, \mathcal{M}(dx) = \int_E f(x) \left(\sum_{i \in I} \delta_{X_i}(dx) \right) = \sum_{i \in I} \int_E f(x)\, \delta_{X_i}(dx) = \sum_{i \in I} f(X_i).$$

Lemma 3.58 ([6])**.** *Let $\mathcal{M}$ be a PRM with intensity μ.*

(i) ***First moment formula:*** *Let $f \geq 0$ or $f \in L^1(\mu)$. Then we have*

$$\mathbb{E}\left[\langle \mathcal{M}, f \rangle\right] = \int_E f(x)\, \mu(dx).$$

(ii) ***Campbell formula:*** *Let $f \geq 0$ or $1 - e^{-f} \in L^1(\mu)$. Then we have*

$$\mathbb{E}\Big[e^{-\langle \mathcal{M}, f \rangle}\Big] = e^{\left(-\int_E \left(1 - e^{-f(x)}\right) \mu(dx)\right)}.$$

This gives us the Laplace functional of a PRM.

(iii) ***Palm formula:*** *Let $G : E \times \mathsf{N} \to \mathbb{R}_+$ be a measurable functional and $f \geq 0$ a measurable function. Then we have*

$$\mathbb{E}[\langle \mathcal{M}, fG(\cdot, \mathcal{M}) \rangle] \coloneqq \mathbb{E}\left[\sum_{i \in I} f(X_i) G(X_i, \mathcal{M}) \right] = \int_E \mathbb{E}[G(x, \delta_x + \mathcal{M})]\, f(x) \mu(dx).$$

Intuitive explanation of the Palm formula [6]: Suppose $\mu(\{x\}) = 0$ for some x in E. Then the conditional distribution of $\mathcal{M}$, given that the process has an atom at x, has the same distribution as the point process $\mathcal{M} + \delta_x$.

However, since $\mu(\{x\}) = 0$, the event that $\mathcal{M}$ has an atom at the fixed point x has probability zero. Thus, this conditioning cannot be understood in the usual elementary sense. A more rigorous mathematical framework is required to clarify the role of the function f and to make the conditioning precise.

Mathematical explanation: Let $\mathcal{M}$ be a PRM on a state space E with intensity measure μ. We write

$$\mathcal{M} = \sum_{i \in I} \delta_{X_i},$$

where the X_i are the atoms of the process and $\mathbb{P}$ denotes the law of $\mathcal{M}$.

To "mark" a point of $\mathcal{M}$, we introduce a measurable function $f : E \to \mathbb{R}_+$. To simplify the probabilistic interpretation, we take f such that

$$f\mu(E) \coloneqq \int_E f(x)\mu(dx) = 1.$$

Therefore, $f\mu$ defines a valid probability measure on the state space E. We "mark" an atom X_* of $\mathcal{M}$ proportionally to this f, meaning that

$$\mathbb{P}(X_* = X_i \mid \mathcal{M}) = \frac{f(X_i)}{\langle \mathcal{M}, f \rangle} = \frac{f(X_i)}{\sum_{i \in I} f(X_j)}.$$

We can now define a new probability measure $\mathbb{P}_*$, which is the law of $(X_*, \mathcal{M})$. This represents the law of the process conditioned on having an atom at X_*. However, the random factor $\langle \mathcal{M}, f \rangle$ complicates the calculations. We circumvent this difficulty by introducing a probability measure on E, defined as

$$\mathbb{Q} \coloneqq \langle \mathcal{M}, f \rangle \mathbb{P}_*.$$

Under $\mathbb{Q}$, the condition that X_* is an atom of $\mathcal{M}$ still holds.

Now consider, independently of the previous construction, a PRM $\mathcal{M}'$ with intensity measure μ. Let ξ be a random atom with law $f\mu$, independent of $\mathcal{M}'$, and denote by $\mathbb{K}$ the law of $(\xi, \mathcal{M}' + \delta_\xi)$.

The Palm formula connects this two constructions and states the fundamental identity:

$$\boxed{\mathbb{Q} = \mathbb{K}.}$$

Proof of Lemma 3.58. We provide the proof for (ii) by assuming $\mu(E) < \infty$; the proof for (i) is omitted as it follows a nearly identical logic, while the proof for (iii) can be found in [6].

(ii) : We can utilize the explicit i.i.d. construction of a finite PRM. As we have seen, the atoms X_i are i.i.d. random variables with common distribution $\mu(\cdot)/\mu(E)$ and $\kappa \coloneqq \mathcal{M}(E)$ is Poisson distributed with parameter $\mu(E)$ independent of $(X_i)_{i \geq 1}$. Then,

$$\begin{aligned}
\mathbb{E}\left[e^{-\langle \mathcal{M}, f \rangle}\right] &= \mathbb{E}\left[e^{-\sum_{i=1}^{\kappa} f(X_i)}\right] \\
&= \sum_{k=0}^{\infty} \mathbb{P}(\kappa = k)\, \mathbb{E}\left[e^{-f(X_1)} \cdots e^{-f(X_k)}\right]
\end{aligned}$$

$$\overset{\text{ll}}{=} \sum_{k=0}^{\infty} e^{-\mu(E)} \frac{\mu(E)^k}{k!} \left(\int_E e^{-f(x)} \frac{\mu(dx)}{\mu(E)} \right)^k$$

$$= e^{-\mu(E)} \sum_{k=0}^{\infty} \frac{1}{k!} \left(\int_E e^{-f(x)} \, \mu(dx) \right)^k$$

$$= \exp\Big(-\mu(E) + \int_E e^{-f(x)} \, \mu(dx) \Big) \qquad \left(\text{using: } e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!} \right)$$

$$= \exp\Big(- \int_E (1 - e^{-f(x)}) \, \mu(dx) \Big) .$$

■

Construction Poisson-Dirichlet random variable

Equipped with the necessary theoretical foundations, we now proceed to construct a Poisson-Dirichlet random variable with parameters $0 < \alpha < 1$ and $\theta = 0$ utilizing a Poisson random measure on $(0, \infty)$.

Let $\mathcal{M}$ be a PRM with intensity $\mu(dx) = x^{-\alpha-1} \, dx$ on $(0, \infty)$. To show that μ is σ-finite, observe that for any $0 < \varepsilon < t < \infty$,

$$\mu\big((\varepsilon, t)\big) = \int_\varepsilon^t x^{-\alpha-1} \, dx = \left[\frac{x^{-\alpha}}{-\alpha} \right]_\varepsilon^t = \frac{\varepsilon^{-\alpha} - t^{-\alpha}}{\alpha},$$

which is finite because $0 < \alpha < 1$ and $0 < \varepsilon < t < \infty$. Since

$$(0, \infty) = \bigcup_{n=1}^{\infty} \left(\tfrac{1}{n}, \, n \right),$$

and each interval $(\frac{1}{n}, \, n)$ has finite μ-measure, the measure μ is σ-finite on $(0, \infty)$.

A simple calculation shows that $\mu((0, \infty)) = \infty$. Therefore, $\mathcal{M}$ has almost surely infinitely many atoms in $(0, \infty)$. However, for any $\varepsilon > 0$

$$\mu\big((\varepsilon, \infty)\big) = \int_\varepsilon^\infty x^{-\alpha-1} \, dx = \left[\frac{x^{-\alpha}}{-\alpha} \right]_\varepsilon^\infty = \frac{\varepsilon^{-\alpha}}{\alpha}, \tag{3.20}$$

which is finite for all $\alpha > 0$. Hence, above any threshold $\varepsilon > 0$ there are almost surely only finitely many points of $\mathcal{M}$. Therefore, for $\mathcal{M} = \sum_{i \geq 1} \delta_{X_i}$, the atoms can be ordered in decreasing order and we take X_1 to be the largest one, X_2 the second largest, and so on.

We now claim that the sum of the atoms of $\mathcal{M}$ is finite almost surely, that is,

$$\sum_{i=1}^{\infty} X_i < \infty \qquad \text{a.s.}$$

Proof of the Claim. Let $f(x) = x\,\mathbb{1}_{\{x\leq 1\}}$ for $x \in (0,\infty)$. Using the First moment formula and $0 < \alpha < 1$, we obtain

$$\mathbb{E}\left[\sum_{i=1}^{\infty} X_i \mathbb{1}_{\{X_i\leq 1\}}\right] = \int_0^{\infty} x\,\mathbb{1}_{\{x\leq 1\}}\,\mu(dx) = \int_0^1 x^{-\alpha}\,dx = \frac{1}{1-\alpha} < \infty.$$

This implies

$$\sum_{i=1}^{\infty} X_i \mathbb{1}_{\{X_i\leq 1\}} < \infty \quad \text{a.s.} \tag{3.21}$$

Moreover, by (3.20) we know that $\mu((1,\infty)) < \infty$. Hence, there are only finitely many points outside $(0,1)$ almost surely. By combining this fact with (3.21), we conclude that

$$\sum_{i=1}^{\infty} X_i < \infty \quad \text{a.s.}$$

■

In such a setting, there is a natural way to build a random mass-partition $\mathbf{S}$ from the atoms of the PRM $\mathcal{M}$. This random mass-partition is given by

$$\mathbf{S} = \left(\frac{X_n}{\sum_{i=1}^{\infty} X_i},\, n \in \mathbb{N}\right), \tag{3.22}$$

and is known as the *Poisson-Kingman partition*. The correspondence between this normalized sequence and the Poisson-Dirichlet distribution is established by the following theorem.

Theorem 3.59. *The random mass-partition* $\mathbf{S}$ *defined in* (3.22) *follows the Poisson-Dirichlet distribution with parameters* $0 < \alpha < 1$ *and* $\theta = 0$.

Proof. We have a PRM $\mathcal{M} = \sum_{i=1}^{\infty} \delta_{X_i}$ with intensity measure $\mu(dx) = x^{-\alpha-1}\,dx$ and law $\mathbb{P}$. Set $T = \sum_{i=1}^{\infty} X_i = \int_0^{\infty} y\,\mathcal{M}(dy)$, and hence $\mathbf{S} = (X_1/T, X_2/T, \ldots)$.

Let N be a random index such that

$$\mathbb{P}(N = i \mid \mathcal{M}) = \frac{X_i}{T} =: S_i.$$

After removing the size-biased mass S_N from the mass-partition $\mathbf{S}$, or equivalently removing the corresponding atom X_N from the PRM $\mathcal{M}$, we are left with

$$\mathcal{M}' = \sum_{i\neq N} \delta_{X_i}.$$

The remaining random mass-partition after cutting off S_N is given by **S'** $= (S_j/(1 - S_N),\ j \neq N)$.

The goal of this proof is to show that S_N and **S'** are independent and that $S_N \sim \text{Beta}(1-\alpha, \alpha)$. The final result then follows by iteration, applying the same procedure to the remaining mass.

(i): First, we compute $\mathbb{E}[T^{-a}]$.

For $q \geq 0$, let $f(x) = qx$. Then

$$\langle \mathcal{M}, f \rangle = \sum_{i \in I} f(X_i) = q \sum_{i=1}^{\infty} X_i = qT.$$

By Campbell's formula,

$$\mathbb{E}\left[e^{-qT}\right] = \exp\left(-\int_0^\infty (1 - e^{-qx})\, \mu(dx)\right) = \exp(-\phi(q))\,,$$

where we take

$$\begin{aligned}\phi(q) &:= \int_0^\infty (1 - e^{-qx})\, \mu(dx), \\ &= \int_0^\infty (1 - e^{-qx}) x^{-\alpha-1}\, dx.\end{aligned}$$

Performing a change of variables with $u = qx$, we obtain

$$\begin{aligned}\phi(q) &= \int_0^\infty (1 - e^{-u}) \left(\frac{u}{q}\right)^{-\alpha-1} \frac{du}{q} \\ &= q^\alpha \int_0^\infty (1 - e^{-u}) u^{-1-\alpha}\, du.\end{aligned}$$

Using integration by parts, the preceding integral is equal to

$$q^\alpha \left[(1 - e^{-u}) \left(-\frac{1}{\alpha} u^{-\alpha}\right)\right]_0^\infty + q^\alpha \frac{1}{\alpha} \int_0^\infty e^{-u} u^{-\alpha}\, du.$$

The first term vanishes, hence

$$\phi(q) = q^\alpha \int_0^\infty (1 - e^{-u}) u^{-1-\alpha}\, du = q^\alpha \frac{1}{\alpha} \int_0^\infty e^{-u} u^{-\alpha}\, du = q^\alpha \frac{1}{\alpha} \Gamma(1 - \alpha).$$

Set $c := \Gamma(1 - \alpha)/\alpha$. Then, we can write

$$\mathbb{E}\left[e^{-qT}\right] = \exp(-\phi(q)) = \exp(-cq^\alpha). \tag{3.23}$$

Moreover, for $a > 0$, we have the following identity [6]:

$$x^{-a} = \frac{1}{\Gamma(a)} \int_0^\infty e^{-xt} t^{a-1} dt. \tag{3.24}$$

Hence,

$$\mathbb{E}[T^{-a}] \overset{(3.24)}{=} \mathbb{E}\Big[\frac{1}{\Gamma(a)}\int_0^\infty e^{-tT}t^{a-1}\,dt\Big]$$

$$\overset{\substack{\text{Fubini-}\\\text{Tonelli}}}{=} \frac{1}{\Gamma(a)}\int_0^\infty \mathbb{E}[e^{-tT}]t^{a-1}\,dt$$

$$\overset{(3.23)}{=} \frac{1}{\Gamma(a)}\int_0^\infty e^{-ct^\alpha}t^{a-1}\,dt.$$

By performing the change of variables $s = ct^\alpha$, the preceding quantity equals

$$\frac{1}{\alpha c^{a/\alpha}\Gamma(a)}\int_0^\infty e^{-s}s^{\frac{a}{\alpha}-1}ds = \frac{\Gamma(a/\alpha)}{\alpha c^{a/\alpha}\Gamma(a)}.$$

Therefore, by setting $a = \alpha$, we can conclude that

$$\mathbb{E}[T^{-\alpha}] = \frac{1}{\Gamma(1-\alpha)\Gamma(\alpha)}. \tag{3.25}$$

(ii): Now, take $f : [0,1] \to [0,\infty)$ measurable, and let G to be a non-negative functional. Then, we have

$$\mathbb{E}[f(S_N)G(\mathcal{M}')] = \mathbb{E}\left[\sum_{i=1}^\infty \frac{X_i}{T}\,f\Big(\frac{X_i}{T}\Big)\,G\left(\sum_{j\neq i}\delta_{X_j}\right)\right]$$

$$= \mathbb{E}\left[\int_0^\infty G(x,\mathcal{M}-\delta_x)f\left(\frac{x}{\int_0^\infty y\mathcal{M}(dy)}\right)\frac{x}{\int_0^\infty y\mathcal{M}(dy)}\,\mathcal{M}(dx)\right].$$

Define

$$g(x) = x,$$

$$H(x,\mathcal{M}) = G(x,\mathcal{M}-\delta_x)\,f\left(\frac{x}{\int_0^\infty y\,\mathcal{M}(dy)}\right)\frac{1}{\int_0^\infty y\,\mathcal{M}(dy)}.$$

By the definition of H,

$$H(x,\mathcal{M}+\delta_x) = G(x,\mathcal{M})\,f\left(\frac{x}{\int_0^\infty y\,(\mathcal{M}+\delta_x)(dy)}\right)\frac{1}{\int_0^\infty y\,(\mathcal{M}+\delta_x)(dy)}$$

$$= G(x,\mathcal{M})\,f\left(\frac{x}{x+\int_0^\infty y\,\mathcal{M}(dy)}\right)\frac{1}{x+\int_0^\infty y\,\mathcal{M}(dy)}.$$

Using the Palm formula with H and g, we arrive at

$$\mathbb{E}\left[\int_0^\infty G(x, \mathcal{M} - \delta_x) f\left(\frac{x}{\int_0^\infty y\mathcal{M}(dy)}\right) \frac{x}{\int_0^\infty y\mathcal{M}(dy)}\, \mathcal{M}(dx)\right]$$
$$= \int_0^\infty \mathbb{E}\left[G(\mathcal{M}) f\left(\frac{x}{x+T}\right) \frac{x}{x+T}\right] x^{-\alpha-1}\, dx.$$

Then, by Fubini-Tonelli and by performing a change of variables $y = \frac{x}{x+T}$, we can continue as follows

$$\mathbb{E}\left[\int_0^1 G(\mathcal{M})\, f(y)\, \left(\frac{yT}{1-y}\right)^{-\alpha} \frac{1-y}{T} \frac{T}{(1-y)^2}\, dy\right]$$
$$= \mathbb{E}\big[G(\mathcal{M})\, T^{-\alpha}\big] \int_0^1 f(y) \left(\frac{y}{1-y}\right)^{-\alpha} \frac{1}{1-y}\, dy$$
$$= \frac{\mathbb{E}[T^{-\alpha}]}{\mathbb{E}[T^{-\alpha}]}\, \mathbb{E}\big[G(\mathcal{M})\, T^{-\alpha}\big] \int_0^1 f(y)\, (1-y)^{\alpha-1}\, y^{-\alpha}\, dy$$
$$= \mathbb{E}[T^{-\alpha}]\, \mathbb{E}\left[G(\mathcal{M})\, \frac{T^{-\alpha}}{\mathbb{E}[T^{-\alpha}]}\right] \int_0^1 f(y)\, (1-y)^{\alpha-1}\, y^{-\alpha}\, dy$$
$$\overset{(3.25)}{=} \frac{1}{\Gamma(1-\alpha)\Gamma(\alpha)}\, \mathbb{E}\left[G(\mathcal{M})\, \frac{T^{-\alpha}}{\mathbb{E}[T^{-\alpha}]}\right] \int_0^1 f(y)\, (1-y)^{\alpha-1}\, y^{-\alpha}\, dy$$
$$= \frac{1}{\Gamma(1-\alpha)\Gamma(\alpha)}\, \mathbb{E}_\alpha[G(\mathcal{M})] \int_0^1 f(y)\, (1-y)^{\alpha-1}\, y^{-\alpha}\, dy.$$

In the last equality, we used the Radon-Nicodym derivative (see Definition A.23)

$$\frac{d\mathbb{P}_\alpha}{d\mathbb{P}} = \frac{T^{-\alpha}}{\mathbb{E}[T^{-\alpha}]},$$

and Proposition A.26. Therefore, we can conclude that $S_N \sim \text{Beta}(1-\alpha, \alpha)$ and that S_N is independent of the remaining process $\mathcal{M}'$ under the law $\mathbb{P}_\alpha$.

■

Remark 3.60. There exists a similar construction for $\alpha = 0$ and $\theta > 0$. The intensity measure for the PRM should then be chosen as $\mu(dx) = \theta x^{-1} e^{-x}\, dx$.

3.2.4 Example

As a motivating example of where the Poisson-Dirichlet distribution arises in practice, we will explore the cycle decomposition of random permutations.

Let S_n be the set of permutations of $S = \{1, \ldots, n\}$. Every permutation $\sigma \in S_n$ can be expressed as a product of disjoint cycles. This decomposition is unique up to the ordering of the cycles and the cyclic rotation of elements within each cycle. To fix a unique representation, we order the disjoint cycles according to their smallest elements and start each cycle with its smallest element.

Example 3.61. Consider $S = \{1, 2, 3, 4, 5, 6, 7\}$ and a permutation $\sigma \in S_7$ defined in two-line notation as

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 2 & 4 & 1 & 7 & 5 & 6 \end{pmatrix}.$$

The cycle decomposition of σ is obtained as follows:

- Start at 1: $1 \to 3,\ 3 \to 4,\ 4 \to 1$ giving the cycle $(1\,3\,4)$.
- 2 maps to itself: cycle (2) (often omitted).
- $5 \to 7,\ 7 \to 6,\ 6 \to 5$ gives $(5\,7\,6)$.

Thus, the cycle decomposition is

$$\sigma = (1\,3\,4)\,(2)\,(5\,7\,6).$$

Now, let $\mathbf{L}^{(n)} = (L_1, L_2, \ldots, L_s)$ denote the cycle lengths of σ for s cycles, ordered by their least elements. We define the normalized vector as

$$\mathbf{X}^{(n)} = (L_1/n, L_2/n, \ldots, L_s/n)\,,$$

which creates a tiling of the unit interval.

Theorem 3.62. *Let $\mathbf{X}^{(n)}$ and $\mathbf{L}^{(n)}$ be as above. Then*

$$\mathbf{X}^{(n)} \overset{d}{\Rightarrow} (P_1, P_2, \ldots),$$

where $(P_1, P_2, \ldots)$ are the asymptotic frequencies of a $\mathrm{PD}(0, 1)$ *random partition in size-biased order.*

Proof. To proof this theorem we will use a construction due to Feller, which is a convenient way to construct a random permutation. We begin a cycle with the smallest unused element. At each subsequent step we either close the current cycle or choose one of the remaining unused elements to extend it. Once a cycle is closed, a new cycle is started again with the smallest unused element. Every choice is made uniformly at random among all admissible options.

To illustrate this construction, let us return to the permutation from Example 3.61. We start with the cycle containing 1. At this stage the permutation looks like:

$$\sigma = (1$$

Next, we choose our following action uniformly at random from n available options. We may either extend the cycle by selecting one of the remaining elements in $\{2, \ldots, n\}$, or we may close the current cycle. Let's say that we choose 5 as the next element. This would lead σ to be:

$$\sigma = (1 \quad 5$$

For the next step, we can choose an element either in the set $\{1, 2, \ldots, n\} \setminus \{1, 5\}$ or the symbol ")", which closes the cycle. Again, one chooses uniformly at random among all possibilities. Let's say we choose the element 2 and then close the cycle. This gives:

$$\sigma = (1 \quad 5 \quad 2)$$

Proceeding further, we will open a new cycle by the smallest element not yet used. In our case, this would be 3. Thus,

$$\sigma = (1 \quad 5 \quad 2)(3 \qquad .$$

At this point we again, choose uniformly at random among the elements that doesn't appear in the previous list or to close the cycle. We continue like this, until all elements of S are used.

With this construction, the cycle decomposition itself becomes a random object. For instance, when $n = 7$ and $\sigma_0 = (1\,3\,4)(2)(5\,6\,7)$, the construction selects exactly the sequence of choices that produces σ_0 with probability

$$\mathbb{P}(\sigma = \sigma_0) = \frac{1}{6} \cdot \frac{1}{5} \cdot \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{1} = \frac{1}{7!}.$$

Following Feller's construction, we now turn our attention to the distribution of the cycle lengths. We first determine the probability $\mathbb{P}(L_1 = k)$ that the initial cycle has length k, where $1 \leq k \leq n$. Because each choice is made uniformly at random from the remaining available options, we can trace the probabilities at each step:

- Step 1: There are n total options available, namely the elements of the set $\{2, \ldots, n\} \cup \{\text{“)”}\}$. To keep the cycle open, we must choose one of the $n - 1$ unused numbers. The probability of this choice is $(n - 1)/n$.
- Step 2: There are now $n - 1$ total options available. We must choose one of the $n - 2$ remaining numbers. The probability is $(n - 2)/(n - 1)$.

$\vdots$

- Step k: There are $n-k+1$ total options available. We must now choose the unique symbol ")" to close the cycle. The probability of making this specific choice is $1/(n-k+1)$.

Hence,

$$\mathbb{P}(L_1 = k) = \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdots \frac{n-k+1}{n-k+2} \cdot \frac{1}{n-k+1} = \frac{1}{n}.$$

Therefore, L_1 is uniformly distributed on $\{1, \ldots, n\}$.

For the second cycle, we condition on the event $\{L_1 = x\}$ for some $x < n$. We then consider the remaining $n - x$ elements and determine the distribution of the second cycle length, $\mathbb{P}(L_2 = k)$, for $1 \le k \le n - x$. By arguing as for the first cycle, we obtain

$$\mathbb{P}(L_2 = k) = \frac{1}{n-x},$$

and therefore $L_2 \sim$ Uniform on $\{1, \ldots, n - L_1\}$.

Similarly, given $(L_1, \ldots, L_{j-1})$ and given that $L_1 + \cdots + L_{j-1} < n$, we have

$$L_j \sim \text{Uniform on } \{1, \ldots, n - (L_1 + \cdots + L_{j-1})\}.$$

We now consider the normalized cycle lengths

$$\mathbf{X}^{(n)} = \left(\frac{L_1}{n}, \frac{L_2}{n}, \ldots\right) = (X_1^{(n)}, X_2^{(n)}, \ldots).$$

Define

$$S_0 \coloneqq n, \qquad S_i \coloneqq n - \sum_{j=1}^{i} L_j.$$

By the preceding computations, for all i such that $S_{i-1} > 0$,

$$\mathbb{P}(L_i = k \mid L_1, \ldots, L_{i-1}) = \frac{1}{S_{i-1}}, \qquad k = 1, \ldots, S_{i-1}.$$

Thus $L_i \mid (L_1, \ldots, L_{i-1})$ is uniformly distributed on $\{1, \ldots, S_{i-1}\}$. Consequently,

$$W_i^{(n)} \coloneqq \frac{L_i}{S_{i-1}} \;\Big|\; (L_1, \ldots, L_{i-1}) \sim \text{Uniform on } \left\{\frac{1}{S_{i-1}}, \frac{2}{S_{i-1}}, \ldots, 1\right\}. \tag{3.26}$$

We may now express the normalized cycle lengths as

$$\begin{aligned} X_1^{(n)} &= W_1^{(n)}, \\ X_2^{(n)} &= (1 - X_1^{(n)})\, W_2^{(n)}, \\ X_3^{(n)} &= (1 - X_1^{(n)} - X_2^{(n)})\, W_3^{(n)}, \\ &\vdots \end{aligned} \tag{3.27}$$

Furthermore, Equation (3.26) shows that the previous cycle lengths only determine the size of the grid on which $W_i^{(n)}$ is uniformly distributed. Moreover, for every fixed i, the mesh size of this grid converges to 0 as $n \to \infty$. Hence, the dependence on the past disappears in the limit. Therefore, for every fixed $s \in \mathbb{N}$, as $n \to \infty$,

$$(W_1^{(n)}, \ldots, W_s^{(n)}) \stackrel{d}{\Rightarrow} (W_1, \ldots, W_s),$$

where $W_1, \ldots, W_s$ are independent Uniform$(0,1)$ = Beta$(1,1)$ random variables. ■

Chapter 4

Kingman's coalescent

In this chapter, we introduce Kingman's coalescent. After defining the stochastic process and exploring its properties, particularly the special phenomenon of *coming down from infinity* and the *consistency property*, we present a formal construction based on Aldous' work. This construction allows us to view the coalescent as a multidimensional version of the paintbox construction. We then connect this framework to classical population genetics by studying its relationship to the well-known Wright-Fisher and Moran model, which naturally leads us to establish its duality relation with the Wright-Fisher diffusion. Finally, we consider extensions of the model incorporating mutations.

4.1 Introduction to the model

Definition 4.1 (Coagulation, [6])**.** Let $\pi = (\pi_1, \pi_2, \dots)$ and $\pi' = (\pi'_1, \pi'_2, \dots)$ be two partitions of $B \subseteq \mathbb{N}$. We say that π' is obtained from π by the coagulation of exactly two of its (non-empty) blocks if there exist integers $1 \leq i < j$ such that

$$\pi_i \neq \emptyset, \qquad \pi_j \neq \emptyset, \qquad \pi'_i = \pi_i \cup \pi_j,$$

and for every $n' \neq i$, there exists some $n \in \mathbb{N} \setminus \{i, j\}$ such that $\pi'_{n'} = \pi_n$.

Definition 4.2 (Kingman's n-coalescent, [5],[6])**.** Let $n \geq 1$. Kingman's n-coalescent is a continuous-time Markov chain $\Pi^{(n)} = (\Pi_t^{(n)})_{t \geq 0}$ taking values in $\mathcal{P}_n$, the space of partitions of $[n] = \{1, \dots, n\}$, characterized by:

(i) **Initial condition:** $\Pi_0^{(n)} = (\{1\}, \{2\}, \dots, \{n\})$ is the partition into singletons.

(ii) **Transition rates:** For any two distinct partitions $\pi, \pi' \in \mathcal{P}_n$, the transition rate $q(\pi, \pi')$ is given by

$$q(\pi, \pi') = \begin{cases} 1, & \text{if } \pi' \text{ is obtained by the coagulation of exactly two blocks of } \pi, \\ 0, & \text{otherwise.} \end{cases}$$

(iii) **Absorbing state:** The partition $([n], \emptyset, \dots)$ is the absorbing state.

We now examine the jump structure and holding times of Kingman's n-coalescent. Suppose the current state of the chain is given by a partition π with $\#\pi = k \geq 2$. Since any pair of blocks can coalesce with rate 1, and there are $\binom{k}{2}$ such pairs, the *total exit rate* from state π is

$$q(\pi) = \sum_{\pi' \neq \pi} q(\pi, \pi') = \underbrace{1 + \cdots + 1}_{\binom{k}{2} \text{ times}} = \binom{k}{2}.$$

It follows that the *holding time* T_k, during which the process stays at π, is exponentially distributed with parameter $q(\pi)$, i.e.,

$$T_k \sim \text{Exp}\left(\binom{k}{2}\right).$$

Crucially, the holding time depends only on the total number of blocks, rather than the specific composition of the partition π.

After time T_k, the process jumps to one of the $\binom{k}{2}$ partitions obtained by the coagulation of exactly two blocks of π, according to the uniform probability. This follows from the fact that the transition probabilities of the jump chain are given by

$$j(\pi, \pi') = \begin{cases} \dfrac{1}{\binom{k}{2}}, & \text{if } \pi' \text{ is obtained by the coagulation of exactly two blocks of } \pi, \\ 0, & \text{otherwise.} \end{cases}$$

A direct consequence of this structure is that the jump chain

$$(\Pi_k^{(n)})_{n \geq k \geq 1},$$

where $\Pi_k^{(n)}$ denotes the state of the n-coalescent when it comprises k blocks, is independent of the sequence of holding times

$$(T_n, T_{n-1}, \ldots, T_2).$$

Remark 4.3 ([6])**.** The dynamics of an n-coalescent can be described by considering the process of the number of non-empty blocks

$$\#\Pi^{(n)} = (\#\Pi_t^{(n)})_{t \geq 0}.$$

This process is a *pure death process* on the state space $\{n, n-1, \ldots, 1\}$. When the process is in state k, it transitions to $k-1$ at rate

$$\mu_k = \binom{k}{2}, \quad k \in \{2, \ldots, n\}.$$

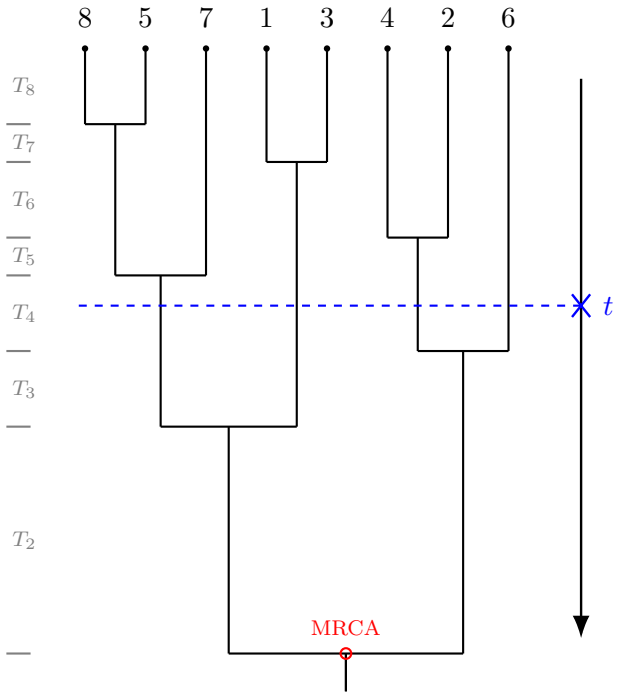


Figure 4.1: Realization of Kingman's n-coalescent ($n = 8$). T_i denotes the sequence of holding times for i ancestral blocks, leading in the Most Recent Common Ancestor (MRCA). At the indicated time t, the process consists of four lineages: $\Pi_t^{(8)} = (\{1,3\},\{2,4\},\{6\},\{5,7,8\})$.

We now arrive at a fundamental property of an n-coalescent, namely the following *consistency* result.

Lemma 4.4 (Consistency)**.** *Let $\Pi^{(n)} = (\Pi_t^{(n)})_{t\geq 0}$ be an n-coalescent. For any $m < n$, the restriction of $\Pi^{(n)}$ to the subset $[m]$, denoted by $\Pi^{(n)}_{|[m]}$, has the same law as the m-coalescent $\Pi^{(m)}$.*

The following proof follows the idea of the argument in [6].

Proof. We aim to show that $\Pi^{(n)}_{|[n-1]} \overset{d}{=} \Pi^{(n-1)}$. To demonstrate this, fix a partition $\gamma \in \mathcal{P}_{n-1}$ and consider an arbitrary partition $\pi \in \mathcal{P}_n$ such that its restriction satisfies $\pi_{|[n-1]} = \gamma$. Assume the n-coalescent $\Pi^{(n)}$ has initial state π.

It is necessary to distinguish the case in which n belongs to a non-singleton block of π from the case in which $\{n\}$ is a singleton block.

Case 1: First, assume n is part of a non-singleton block, and let $i < n$ be the smallest integer in that block. In this case, $\Pi_t^{(n)}$ is recovered by simply adding n to the block of $\Pi^{(n)}_{t|[n-1]}$ that contains the integer i for every $t \geq 0$. Consequently, we can directly conclude that $\Pi^{(n)}_{|[n-1]}$ is a ctMC with the same transition rates as an $(n-1)$-coalescent with initial state γ.

Case 2: Suppose that $\{n\}$ is a singleton block of π such that $\pi = (\{n\}, B_2, \ldots, B_k)$ with $\#\pi = k \geq 3$. Let T be the first jump time of $\Pi^{(n)}$. It follows that

$$T \sim \text{Exp}\left(\frac{k(k-1)}{2}\right)$$

and that T is independent of

$$\pi' \coloneqq \Pi^{(n)}_T.$$

Now define the two events

$$A = \{\text{the block } \{n\} \text{ is not involved in the coagulation at time } T\},$$
$$A^c = \{\text{the block } \{n\} \text{ is involved in the coagulation at time } T\}.$$

The number of pairs involving $\{n\}$ is $k-1$. Hence,

$$\mathbb{P}(A^c) = \frac{k-1}{\binom{k}{2}} = \frac{k-1}{\frac{k(k-1)}{2}} = \frac{2}{k},$$

$$\mathbb{P}(A) = 1 - \frac{2}{k}.$$

Conditionally on A, the restriction $\pi'_{|[n-1]}$ has uniform distribution on the set of partitions obtained by the coagulation of two blocks of $\pi_{|[n-1]} = \gamma$. Indeed, for any distinct $i, j \in \{2, \ldots, k\}$ the probability that blocks B_i and B_j coalesce is given by

$$\mathbb{P}(B_i \text{ and } B_j \text{ coalesce} \mid A) = \frac{\mathbb{P}(B_i, B_j \text{ coalesce } \cap A)}{\mathbb{P}(A)} = \frac{\frac{1}{\binom{k}{2}}}{\frac{k-2}{k}} = \frac{2}{(k-1)(k-2)}.$$

Conditionally on A^c, denote by T' the holding time for the second jump of $\Pi^{(n)}$. Then,

$$T' \sim \text{Exp}\left(\frac{(k-1)(k-2)}{2}\right).$$

Let

$$\pi'' \coloneqq \Pi^{(n)}_{(T+T')}.$$

By the strong Markov property, the shifted process $(\Pi^{(n)}_{T+t})_{t\geq 0}$ behaves as an n-coalescent starting from π' with $\pi'_{|[n-1]} = \gamma$. Furthermore, since we are working conditionally on A^c, the integer n belongs to a non-singleton block of π'. Therefore, we are in Case 1 of our proof. Hence, restricting the process to $[n-1]$ gives us a ctMC with the same transition probabilities as an $(n-1)$-coalescent with initial state γ. It follows that the state $\pi''_{|[n-1]}$ is independent of both the previous holding time T and the current holding time T', and is uniformly distributed over the $(k-1)(k-2)/2$ partitions obtained from the coagulation of two blocks of $\pi'_{|[n-1]} = \gamma$.

Therefore, the holding time of the process $\Pi^{(n)}_{|[n-1]}$ in state $\pi_{|[n-1]} = \gamma$ depends on whether the singleton $\{n\}$ is involved in the coagulation. This time is given by

$$T + \mathbf{1}_{A^c} T',$$

after which the process jumps to $\pi'_{|[n-1]}$ on the event A, or to $\pi''_{|[n-1]}$ on the event A^c. Moreover, since T and T' are independent, the density of their sum $T + T'$ is given by the convolution $f_T * f_{T'}$ (see [25], Chapter 9.5). Defining the random variable $X = T + \mathbf{1}_{A^c} T'$, the density of it can be expressed as

$$f_X(t) = \mathbb{P}(A) f_T(t) + \mathbb{P}(A^c)(f_T * f_{T'})(t).$$

A direct calculation of this expression shows that $X \sim \text{Exp}\left(\frac{(k-1)(k-2)}{2}\right)$. Moreover, X is independent of $\pi'_{|[n-1]}$ on A and $\pi''_{|[n-1]}$ on A^c.

All in all, we conclude that $\Pi^{(n)}_{|[n-1]}$ is a continuous-time Markov chain with the characteristic of an $(n-1)$-coalescent. ■

Applying the Kolmogorov Extension Theorem (Theorem A.31) to the consistent family of n-coalescents yields the construction of a coalescent process on the space of partitions of $\mathbb{N}$, denoted by $\mathcal{P}_\infty$. More precisely, we have the following definition:

Definition 4.5 (Kingman's coalescent, [6])**.** There exists a unique (in law) process $\Pi^K = (\Pi^K_t)_{t \geq 0}$ with values in $\mathcal{P}_\infty$ such that the restriction of Π^K to $[n]$ is an n-coalescent. Π^K is called Kingman's coalescent.

Càdlàg. Note that the n-coalescent is a càdlàg process and therefore is an element of the Skorokhod space $D(\mathbb{R}_+, \mathcal{P}_n)$. The left limit at any jump time always exists and corresponds to the partition immediately before the coalescence event. Right-continuity holds by definition, as the state of the process at the exact moment of a jump is defined to be the newly formed partition. Therefore, the càdlàg property of Kingman's coalescent is a direct consequence of the sample path continuity of its finite-dimensional restrictions.

Exchangeability. The exchangeability of the n-coalescent follows from its definition, since the coalescence rate is identical for every possible pair of blocks. By the consistency property, this implies that $\Pi^K_{|[n]} \overset{d}{=} \sigma(\Pi^K_{|[n]})$ for any permutation σ of $[n]$. Hence, Π^K is an exchangeable process.

Another special property of Kingman's coalescent is that it *comes down from infinity*. This property states that, although Kingman's coalescent is initially made up of an infinite number of singletons, after any positive amount of time the partition consists of a finite number of blocks almost surely.

Theorem 4.6. *Let $N_t = \#\Pi^K_t$. Let F be the event that for all $t > 0$, $N_t < \infty$. Then $\mathbb{P}(F) = 1$.*

Before presenting the formal proof, we first provide an intuitive explanation of why this "big-bang" phenomenon occurs. We have already seen that the waiting time for a transition from a state with n blocks to a state with $n-1$ blocks is exponentially distributed with parameter $\lambda = n(n-1)/2$. For large n, this rate is approximately $n^2/2$. Motivated by this observation, we introduce a deterministic function $u(t)$ that approximates the number of blocks at time t. This suggests that the evolution of the number of blocks should approximately satisfy the differential equation

$$\begin{cases} u'(t) = -\dfrac{u(t)^2}{2}, \\ u(0) = +\infty. \end{cases} \tag{4.1}$$

The solution of (4.1) is $u(t) = 2/t$. In particular, $u(t)$ is finite for every $t > 0$, while it diverges as $t \to 0$. This therefore suggests that the number of blocks N_t is almost surely finite for every $t > 0$.

Proof. We consider the restrictions of Π^K to $[n]$ and denote the number of blocks by $N_t^n = \#\Pi^K_{|[n]}$. The sequence of random variables $(N_t^n)_{n\in\mathbb{N}}$ is non-decreasing in n for every fixed t. For a fixed $M > 0$ the sequence of events $\{N_t^n > M\}$ is increasing and nested with respect to n. Consequently, by the continuity of probability measures from below

$$\mathbb{P}(N_t > M) = \lim_{n\to\infty} \mathbb{P}(N_t^n > M).$$

The event $\{N_t^n > M\}$ means that by time t, the process has not yet reached M blocks. Consequently, fewer than $n - M$ coalescence events have occurred. The total waiting time to go from n to M blocks is

$$T_{tot} = T_n + T_{n-1} + \cdots + T_{M+1} = \sum_{k=M+1}^{n} T_k,$$

where $T_k \sim \mathrm{Exp}\left(\frac{k(k-1)}{2}\right)$.

By combining these facts with Markov's inequality (A.2), it follows that

$$\begin{aligned} \mathbb{P}(N_t^n > M) &= \mathbb{P}\left(\sum_{k=M+1}^{n} T_k > t\right) \\ &\leq \frac{1}{t}\,\mathbb{E}\left[\sum_{k=M+1}^{n} T_k\right] \\ &= \frac{1}{t}\sum_{k=M+1}^{n} \mathbb{E}[T_k] \\ &= \frac{1}{t}\sum_{k=M+1}^{n} \frac{2}{k(k-1)} \end{aligned}$$

$$\leq \frac{1}{t} \sum_{k=M+1}^{\infty} \frac{2}{k(k-1)} = \frac{1}{t} \left(\frac{2}{M} \right).$$

Letting $n \to \infty$, we obtain

$$\mathbb{P}(N_t > M) \leq \frac{2}{tM}. \tag{4.2}$$

Since

$$\{N_t = \infty\} = \bigcap_{M=1}^{\infty} \{N_t > M\},$$

and the events $\{N_t > M\}$ form a decreasing nested sequence with respect to M, continuity of probability measures from above yields

$$\mathbb{P}(N_t = \infty) = \mathbb{P}\left(\bigcap_{M=1}^{\infty} \{N_t > M\} \right) = \lim_{M \to \infty} \mathbb{P}(N_t > M).$$

Therefore, by using inequality (4.2),

$$\mathbb{P}(N_t = \infty) = \lim_{M \to \infty} \mathbb{P}(N_t > M) \leq \lim_{M \to \infty} \frac{2}{tM} = 0.$$

Since probabilities are non-negative, it follows that

$$\mathbb{P}(N_t = \infty) = 0,$$

which implies

$$\mathbb{P}(N_t < \infty) = 1.$$

■

Remark 4.7 (Shape of the Kingman's coalescent, [13])**.** The expected total time of the coalescent process, which represents the time needed to trace a sample back to its Most Recent Common Ancestor (MRCA), is finite and is given by

$$\mathbb{E}[T_{\text{MRCA}}] = \sum_{k=2}^{n} \mathbb{E}[T_k] = \sum_{k=2}^{n} \frac{2}{k(k-1)} = 2 \sum_{k=2}^{n} \left(\frac{1}{k-1} - \frac{1}{k} \right) = 2 \left(1 - \frac{1}{n} \right).$$

As $n \to \infty$, this quantity converges to 2. Note also that $\mathbb{E}[T_2] = 1$. This implies that the expected time spent waiting for the final coalescence event (when only two lineages remain) accounts for exactly half of the total expected coalescent time. This confirms that the "deep" ancestral history is the most time-consuming part of the process, whereas the "recent" history of a large population is compressed into a very short interval near $t = 0$.

4.2 Aldous' construction

Let $(U_j)_{j\in\mathbb{N}}$ be a collection of i.i.d. uniform random variables on $(0,1)$. For $k \geq 2$, let E_k be independent exponential random variables with rate $k(k-1)/2$ and define the *random times*

$$\tau_j = \sum_{k=j+1}^{\infty} E_k < \infty.$$

By comparing two consecutive random times,

$$\tau_{j-1} = E_j + E_{j+1} + \cdots \quad \text{and} \quad \tau_j = E_{j+1} + E_{j+2} + \cdots,$$

it is evident that $\tau_{j-1} - \tau_j = E_j > 0$ a.s., which implies the strict ordering

$$\tau_{j-1} > \tau_j \quad \text{a.s.}$$

Next, we define a function $f : (0,1) \to \mathbb{R}$ that assigns *heights* to our uniform random variables:

$$f(u) = \begin{cases} \tau_j, & \text{if } u = U_j \text{ for some } j \geq 1, \\ 0, & \text{otherwise.} \end{cases}$$

This construction is represented graphically in Figure 4.2 below, where the horizontal axis corresponds to the unit interval $(0,1)$ and the vertical axis represents time t. The heights are visualized by drawing a vertical stick from $(U_j, 0)$ to $(U_j, f(U_j))$ for each $j \geq 1$.

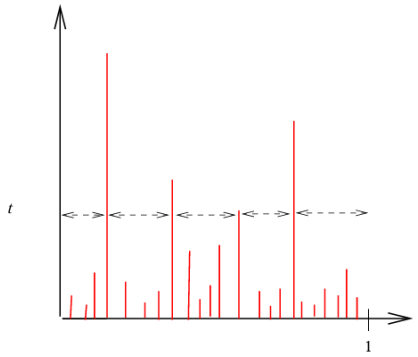


Figure 4.2: Aldous' construction. Source: [5].

For any $t \geq 0$, we define a tiling S_t of $(0,1)$ as the collection of open connected components of the set $\{u \in (0,1) : f(u) > t\}$. Moreover, note that under the condition

$$\tau_{k+1} < t < \tau_k,$$

only the points $U_1, \ldots, U_k$ satisfy $f(U_j) > t$. Consequently, the interval $(0,1)$ is cut at exactly k points, producing $k+1$ intervals. At the random time $t = \tau_k$, the set of cut points becomes $\{U_1, \ldots, U_{k-1}\}$, resulting in exactly k intervals.

Theorem 4.8. *$(S_t)_{t\geq 0}$ has the distribution of the asymptotic frequencies of Kingman's coalescent.*

To prove this theorem, we need to introduce the notion of *rooted segments* and *broken rooted segments*.

Definition 4.9 (Rooted segment)**.** A rooted segment on k points $i_1, \ldots, i_k$ is one of the $k!$ possible linear orderings of these k points. Think of them as being ordered from left to the right. The left-most point is called the root of the segment.

Example 4.10. Let $k = 3$ and $i_1 = 2, i_2 = 5, i_3 = 7$. For example, two possible rooted segments are

$$5 \to 2 \to 7 \quad \text{and} \quad 2 \to 5 \to 7,$$

where in the first case the root is 5, and in the second case the root is 2.

Definition 4.11 (Broken rooted segment)**.** Let $n \geq 1$ and $1 \leq k \leq n$. A broken rooted segment of $[n]$ with k distinct connected components is a collection of k disjoint rooted segments whose union is $[n]$. The relative order of the k segments themselves is irrelevant. We denote the set of all such broken rooted segments by $\mathcal{R}_{n,k}$.

Remark 4.12. Each broken rooted segment naturally induces a partition by disregarding the linear ordering within each segment and treating the components as blocks.

Example 4.13. Let $n = 5$ and $k = 2$. One possible element of $\mathcal{R}_{5,2}$ is

$$(4 \to 1 \to 5)\ (3 \to 2).$$

The associated partition of [5] is

$$\pi = (\{1, 4, 5\}, \{2, 3\}).$$

Lemma 4.14. *The random partition induced by a uniform element of $\mathcal{R}_{n,k}$ has the same distribution as $\Pi_k^{(n)}$, the state of the jump chain of Kingman's n-coalescent when it comprises k blocks.*

Proof. Let Σ_n be a random element of $\mathcal{R}_{n,n}$. The partition associated with Σ_n has the same structure as $\Pi_n^{(n)}$, since in both cases we have only singletons almost surely.

Assume that for $k \leq n$ and Σ a uniformly chosen element of $\mathcal{R}_{n,k}$, we have that the associated partition to Σ has the same distribution as $\Pi_k^{(n)}$. Let Σ' be obtained from Σ by:

- Choosing an unordered pair of segments of Σ uniformly at random.
- Merging the two segments into a single component.
- Choosing one of the two possible ways to concatenate the segments (preserving their internal relative order) uniformly at random.

Let $\xi' \in \mathcal{R}_{n,k-1}$. We denote the relation that ξ' can be obtained from $\xi \in \mathcal{R}_{n,k}$ by merging two parts by

$$\xi \preceq \xi'.$$

By the Bayes' rule of total causes (see Theorem A.14),

$$\mathbb{P}(\Sigma' = \xi') = \sum_{\xi \in \mathcal{R}_{n,k}, \xi \preceq \xi'} \mathbb{P}(\Sigma' = \xi' \mid \Sigma = \xi)\, \mathbb{P}(\Sigma = \xi).$$

Since Σ is uniform on $\mathcal{R}_{n,k}$, we have

$$\mathbb{P}(\Sigma = \xi) = \frac{1}{\#\mathcal{R}_{n,k}}.$$

Given $\Sigma = \xi$, the probability that $\Sigma' = \xi'$ is

$$\mathbb{P}(\Sigma' = \xi' \mid \Sigma = \xi) = \frac{1}{\binom{k}{2}} \cdot \frac{1}{2} = \frac{1}{k(k-1)},$$

since there are $\binom{k}{2}$ unordered pairs of segments to merge and two possible orders for the merged segments. Substituting these into the sum yields

$$\mathbb{P}(\Sigma' = \xi') = \sum_{\xi \preceq \xi'} \frac{1}{\#\mathcal{R}_{n,k}} \, \frac{1}{k(k-1)}.$$

To determine the number of predecessors ξ for a fixed ξ', we observe that to recover ξ from ξ', we must split one of the $k-1$ segments of ξ' into two. If a segment has size m, there are $m-1$ possible "cut points" that preserve the internal linear order of the elements.
Let $m_1, \ldots, m_{k-1}$ be the segment sizes of the $k-1$ segments of ξ'. Then, the total number of possible predecessors is given by

$$(m_1 - 1) + \cdots + (m_{k-1} - 1) = \left(\sum_{i=1}^{k-1} m_i \right) - (k-1) = n - k + 1.$$

Consequently,

$$\mathbb{P}(\Sigma' = \xi') = \frac{1}{\#\mathcal{R}_{n,k}} \, \frac{n-k+1}{k(k-1)}.$$

Since this probability is constant for all $\xi' \in \mathcal{R}_{n,k-1}$, we conclude that Σ' is uniformly distributed on $\mathcal{R}_{n,k-1}$.

By proving that the broken rooted segments remain uniformly distributed after each merger, we establish that the induced partition transitions exactly like the jump chain of Kingman's n-coalescent. ■

Proof of Theorem 4.8. For an n-coalescent $(\Pi_t^{(n)})_{t\geq 0}$ the successive states visited by this process are

$$(\Pi_n^{(n)}, \Pi_{n-1}^{(n)}, \ldots, \Pi_1^{(n)}),$$

where $\Pi_k^{(n)}$ denotes the partition of $[n]$ with k blocks. As already seen, the jump chain is independent of the sequence of holding times $(T_n, T_{n-1}, \ldots, T_2)$. Letting $n \to \infty$, and consider the two processes backward in time, we obtain

$$(\Pi_1, \Pi_2, \ldots),\ (T_2, T_3, \ldots).$$

The holding times and the reversed jump chain are still independent.

In Aldous' construction, τ_k denotes the time at which the tiling S_t consists of exactly k intervals. We define a sequence of partitions $(\Pi'_k)_{k\geq 1}$ via a multidimensional paintbox construction. Let $(V_j)_{j\in\mathbb{N}}$ be i.i.d. uniform variables on $(0,1)$, independent of S_t for every $t \geq 0$. The partition Π'_k is defined, for every $k \geq 1$, by the following equivalence relation:

$$i \sim j \text{ in } \Pi'_k \iff V_i, V_j \text{ fall within the same interval-component of } S_{\tau_k}.$$

Moreover, by construction of the process $(S_t)_{t\geq 0}$, the holding time in a state consisting of k intervals is exponentially distributed with rate $k(k-1)/2$. Hence, the holding times of $(\Pi_1, \Pi_2, \ldots)$ and $(\Pi'_1, \Pi'_2, \ldots)$ have the same distribution. Consequently, it remains to verify that

$$\Pi'_k \overset{d}{=} \Pi_k$$

for all $k \geq 1$. Once this is established, Kingman's correspondence ensures that the process $(S_t)_{t\geq 0}$ correctly describes the evolution of the asymptotic frequencies of Kingman's coalescent.

A random uniform element of $\mathcal{R}_{n,k}$ can be selected by first choosing a rooted segment on $[n]$ uniformly at random and then break it at $k-1$ uniformly chosen links. The cut points can be naturally identified with positions in the interval $(0,n)$, corresponding to the left-to-right ordering along the line. These $k-1$ cuts $J_1, \ldots, J_{k-1}$ are uniform distributed on $\{1, \ldots, n-1\}$. Then, by rescaling $(0,n)$ by a factor $1/n$, the cut points become

$$U_i^{(n)} = \frac{J_i}{n} \quad \text{for} \quad i \in \{1, \ldots, k-1\}.$$

By taking $n \to \infty$,

$$(U_1^{(n)}, \ldots, U_{k-1}^{(n)}) \overset{d}{\Rightarrow} (U_1, \ldots, U_{k-1}),$$

where U_i are i.i.d. uniform random variables on $(0,1)$.

Furthermore, Lemma 4.14 ensures that a random partition associated with a uniform element of $\mathcal{R}_{n,k}$ has the same law as $\Pi_k^{(n)}$. As the rescaled block sizes of $\mathcal{R}_{n,k}$ converge in distribution to the interval lengths of the tiling S_{τ_k}, the continuity of Kingman's correspondence ensures that the convergence of these asymptotic frequencies implies the convergence in distribution of the associated partitions in $\mathcal{P}_\infty$. Thus,

$$\Pi_k^{(n)} \overset{d}{\Rightarrow} \Pi'_k.$$

By the consistency property of Kingman's coalescent it follows that

$$\Pi_k^{(n)} \overset{d}{\Rightarrow} \Pi_k.$$

Finally, the uniqueness of the limit establishes the distributional identity

$$\Pi'_k \overset{d}{=} \Pi_k.$$

■

A direct consequence of Theorem 4.8 is the following corollary.

Corollary 4.15. *Let τ_k be the first time at which Kingman's coalescent has k blocks, and let S_{τ_k} denote the asymptotic frequencies at time τ_k, ranked in non-increasing order. Then S_{τ_k} is distributed uniformly over the $(k-1)$-dimensional simplex*

$$\Delta_k = \{x_1 \geq \cdots \geq x_k \geq 0 : \sum_{i=1}^{k} x_i = 1\}.$$

Furthermore, by applying Lemma 4.14 directly, we obtain the explicit probability distribution for any given state of Kingman's n-coalescent.

Corollary 4.16. *Let $1 \leq k \leq n$. Then, for any partition of $[n]$ with exactly k blocks, say $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$, we have*

$$\mathbb{P}(\Pi_k^{(n)} = \pi) = \frac{(n-k)!\,k!\,(k-1)!}{n!\,(n-1)!} \prod_{i=1}^{k} (\#\pi_i)!. \tag{4.3}$$

Proof. A rooted segment on $[n]$ is simply a linear ordering of the n elements, and there are $n!$ possible choices. Such a rooted segment contains $n-1$ links between consecutive elements. Selecting $k-1$ of these links determines a decomposition into k rooted segments, and there are

$$\binom{n-1}{k-1}$$

possible choices of cut locations. Since, by definition of $\mathcal{R}_{n,k}$ the order of the rooted segments does not matter, it finally follows that

$$\#\mathcal{R}_{n,k} = \frac{n!}{k!}\binom{n-1}{k-1}. \tag{4.4}$$

Let ϕ be the map that sends a broken rooted segment to the associated partition. Note that,

$$\#\{\xi \in \mathcal{R}_{n,k} : \phi(\xi) = \pi\} = \prod_{i=1}^{k} (\#\pi_i)!.$$

Hence, if Σ is uniformly distributed on $\mathcal{R}_{n,k}$, then

$$\begin{aligned}
\mathbb{P}(\phi(\Sigma) = \pi) &= \sum_{\xi:\, \phi(\xi)=\pi} \mathbb{P}(\Sigma = \xi) \\
&= \sum_{\xi:\, \phi(\xi)=\pi} \frac{1}{\#\mathcal{R}_{n,k}} \\
&= \frac{\#\{\xi \in \mathcal{R}_{n,k} : \phi(\xi) = \pi\}}{\#\mathcal{R}_{n,k}} \\
&\overset{(4.4)}{=} \frac{1}{\binom{n-1}{k-1}\frac{n!}{k!}} \prod_{i=1}^{k} (\#\pi_i)! \\
&= \frac{(n-k)!k!(k-1)!}{n!(n-1)!} \prod_{i=1}^{k} (\#\pi_i)!.
\end{aligned}$$

Finally, from Lemma 4.14, $\phi(\Sigma)$ has the same distribution as $\Pi_k^{(n)}$. Therefore,

$$\mathbb{P}(\Pi_k^{(n)} = \pi) = \frac{(n-k)!k!(k-1)!}{n!(n-1)!} \prod_{i=1}^{k} (\#\pi_i)!.$$

■

Let $\Sigma = (\Sigma_1, \ldots, \Sigma_n)$ be a process such that $\Sigma_k \in \mathcal{R}_{n,k}$ for all $1 \leq k \leq n$. We define this process as follows: Let Σ_1 be a uniform rooted segment on $[n]$. Given Σ_i with $1 \leq i \leq n-1$, we get to Σ_{i+1} by cutting a randomly chosen link from Σ_i. (See Figure 4.3).

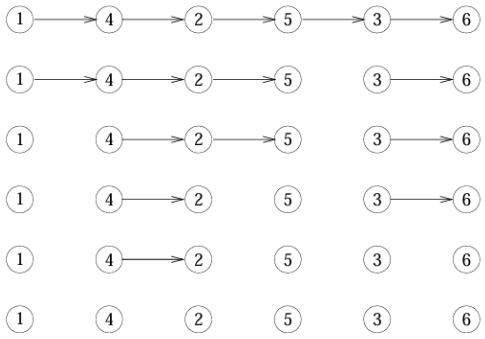


Figure 4.3: Cutting a rooted segment. Source: [5].

Corollary 4.17. *The partition-valued process associated with the time-reversal of* Σ*, denoted by* $(\Sigma_n, \Sigma_{n-1}, \dots, \Sigma_1)$*, is distributed as the jump chain of Kingman's* n*-coalescent* $(\Pi_k^{(n)})_{n \geq k \geq 1}$.

Proof. Although this is a direct consequence of Lemma 4.14, we provide an independent proof using the theory of discrete-time Markov chains.

By construction, the process Σ is a discrete-time Markov chain, since its future evolution depends only on its current state and not on its past. For a state $\xi' \in \mathcal{R}_{n,k}$, there are exactly $n-k$ links remaining that can be cut to reach a state in $\mathcal{R}_{n,k+1}$. Thus, the transition probabilities are

$$p(\xi', \xi) = \begin{cases} \frac{1}{n-k}, & \text{if } \xi \text{ is obtained by cutting a link from } \xi' \\ 0, & \text{otherwise.} \end{cases}$$

The transition probabilities of the reversed chain $(\Sigma_n, \dots, \Sigma_1)$ are given by

$$p_r(\xi, \xi') = \mathbb{P}\big(\Sigma_k = \xi' \mid \Sigma_{k+1} = \xi\big) = \frac{\mathbb{P}(\Sigma_k = \xi')}{\mathbb{P}(\Sigma_{k+1} = \xi)}\, p(\xi', \xi).$$

Σ_k is uniformly distributed on $\mathcal{R}_{n,k}$ (holds for $k = 1$ and propagates by induction). Moreover, by Equation (4.4), we have

$$\frac{\mathbb{P}(\Sigma_k = \xi')}{\mathbb{P}(\Sigma_{k+1} = \xi)}\, p(\xi', \xi) = \frac{\#\mathcal{R}_{n,k+1}}{\#\mathcal{R}_{n,k}} = \frac{\frac{n!}{(k+1)!}\binom{n-1}{k}}{\frac{n!}{k!}\binom{n-1}{k-1}}\, \frac{1}{n-k} = \frac{1}{k(k+1)}.$$

However, to get the associated partition of ξ' where two specific blocks are merged, we have two possibilities: concatenating the first segment before the second, or the second before the first. Adding the probabilities of these two equivalent transitions gives $2/k(k+1)$. This perfectly matches the transition rate of the jump chain of Kingman's n-coalescent. ∎

4.3 Genealogy of populations

The power of Kingman's coalescent lies in its role as *universal limit*. Kingman's coalescent serves as the limit for a wide array of population genetic models. This approximation is established by considering the scaling limit as the population size $N \to \infty$, while keeping the sample size n fixed at an arbitrary but finite value. For this approximation to work, we need the following *assumptions*:

Model Assumptions

(A1) Constant population size: The total number of individuals remains stable over time and individuals typically have few offspring.

(A2) Mean-field mixing: Everyone is liable to interact with anyone.

(A3) Selection neutrality: Selection is absent from the model. Therefore, all individuals are fitness-equivalent.

We want to analyze the ancestry of a sample. If we consider the ancestry of individuals, the genealogy becomes complex because every individual has two parents. However, analyzing the ancestry of a gene is much simpler. In diploid populations, such as humans, each individual carries two copies of every gene, one inherited from each parent. Therefore, the genealogical lineage of a single gene copy is always a single path. This means that if we sample n individuals from a diploid population, we have $2n$ haploid genetic lineages to track. Consequently, from now on, we will make no distinction between a *gene* and an *individual*.

To analyze the ancestry of a sample, we introduce the *ancestral partition process.* Suppose we have a population model of constant size N defined over a time interval $I = [-T, 0]$. We sample n individuals (without replacement) at time 0 for $n < N$. Consider the random partition $\Pi_t^{N,n}$, where $i \sim j$ (i and j are in the same block) if and only if individuals x_i and x_j share the same ancestor at time $-t$. This process $(\Pi_t^{N,n})_{0 \le t \le T}$ is a *coalescent process.*

It is important to emphasize that time in the coalescent is measured *backward* into the past. This runs opposite to the natural forward-time direction of biological evolution and traditional population genetic models, such as the Wright-Fisher and Moran model.

4.4.1 Moran model

The Moran model is a continuous-time population model. We consider a population of constant size N, where each individual lives for an exponentially distributed amount of time with rate 1. Since birth and death occur simultaneously, whenever an individual dies, it is immediately replaced. To generate this replacement, a parent is chosen uniformly at random from the entire population (which may include the

individual that just died). The offspring then inherits the genetic lineage of this chosen parent [13].

In the simplest case [13], the population is made up of N genes, each of either type A or type a. Let X_t be the number of type a individuals at time $t \geq 0$. The process $(X_t)_{t\geq 0}$ is a ctMC. Specifically, it is a birth-death process on the state space $S = \{0, 1, \ldots, N\}$ characterized by the following transition rates:

$$q(i, i+1) = (N-i)\,\frac{i}{N}, \quad q(i, i-1) = i\,\frac{(N-i)}{N}. \tag{4.5}$$

To explain the transition rates, consider the first one. Suppose there are i individuals of type a and $N - i$ individuals of type A. An individual of type A is chosen to be replaced. This occurs at a total rate $N - i$. The number of individuals of type a increases if the reproducing individual is of type a. Since a parent is chosen uniformly from the population, this happens with probability i/N.

The infinitesimal generator, which will play a fundamental role in Chapter 4.4.4, is given by

$$\begin{aligned}(\mathcal{L}f)(i) &= \sum_{j\neq i} q(i,j)(f(j) - f(i)) \\ &= q(i, i+1)(f(i+1) - f(i)) + q(i, i-1)(f(i-1) - f(i)) \\ &= \frac{i}{N}(N-i)(f(i+1) + f(i-1) - 2f(i)).\end{aligned} \tag{4.6}$$

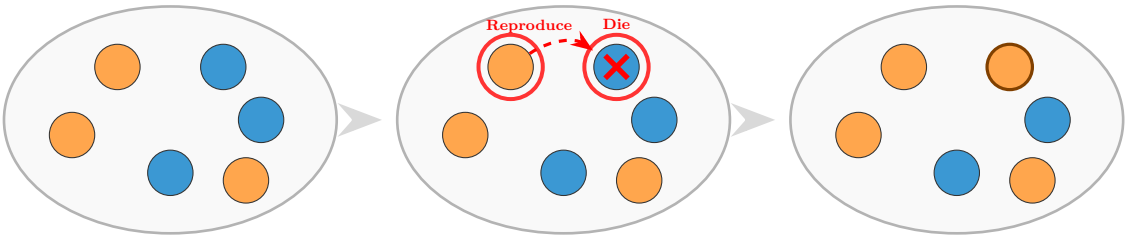


Figure 4.4: A simple visualization of the Moran model with two alleles.

We now change perspective and explore the connection between the ancestral partition of this model and Kingman's n-coalescent $(\Pi_t^{(n)})_{t\geq 0}$.

Theorem 4.18. *Let $n \geq 1$ be a fixed number. Let $x_1, \ldots, x_n$ be n individuals sampled without replacement from a population of size N at time $t = 0$. Let $(\Pi_t^{N,n})_{t\geq 0}$ denote the corresponding ancestral partition process. If we accelerate time by a factor of $N/2$, it follows that:*

$$\left(\Pi^{N,n}_{\frac{Nt}{2}}\right)_{t\geq 0} \overset{d}{=} \left(\Pi_t^{(n)}\right)_{t\geq 0}. \tag{4.7}$$

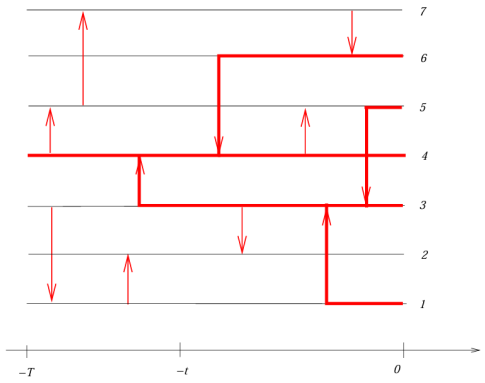


Figure 4.5: Ancestral partition process associated with the Moran model. $N = 7$ and the sample consists of individuals $1, 3, 4, 5, 6$. $\Pi_t^{7,5} = (\{1, 3, 5\}, \{4, 6\})$ and $\Pi_T^{7,5} = (\{1, 3, 5, 4, 6\})$. Source: [5].

Proof. We consider N independent homogeneous PRMs on $\mathbb{R}$ with rate 1 (intensity measure $\mu(dx) = 1\,dx$). For individual $i \in \{1, \ldots, N\}$, write

$$\Pi^{(i)} \;=\; \sum_{\substack{k=-\infty,\\ k\neq 0}}^{\infty} \delta_{T_k^{(i)}},$$

where the atoms are ordered as

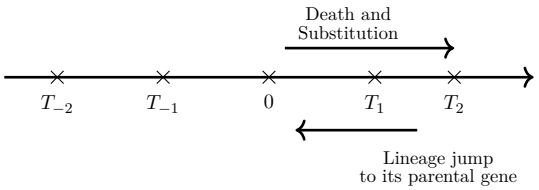
$$\cdots < T_{-2}^{(i)} < T_{-1}^{(i)} < 0 < T_1^{(i)} < T_2^{(i)} < \cdots .$$

Each atom represents a time/event at which individual i dies and is replaced by an offspring of an individual in the population, chosen uniformly at random.

As shown in the Construction of a PRM on the real line, we have that

$$\widetilde{\Pi}^{(i)} \;=\; \sum_{\substack{k=-\infty,\\ k\neq 0}}^{\infty} \delta_{-T_k^{(i)}}$$

has the same distribution as $\Pi^{(i)}$, since the intensity measure is unchanged. Therefore, looking backward in time does not change the distribution of the events.



Assume now that at time $-t$ there are $k \leq n$ lineages that have not coalesced yet. Define the process

$$Z_k \;=\; \widetilde{\Pi}^{(1)} + \cdots + \widetilde{\Pi}^{(k)}.$$

By the Superposition Theorem (Theorem 3.50), Z_k is a PRM with intensity

$$\mu_k(dx) \;=\; k\,dx.$$

Hence, k is the total rate at which any of these k lineages hits an event.

Now, fix one of these k lineages and look for a parental gene of this lineage. There are N other individuals in the population who could be that parental gene. At an event, the probability that the parent is one of the other $(k-1)$ lineages is

$$\mathbb{P}(\text{Coalescence} \mid \text{Event}) \;=\; \frac{\text{number of other lineages}}{\text{number of potential parents}} \;=\; \frac{k-1}{N}.$$

Therefore, the total coalescence rate is

$$k\,\frac{k-1}{N} \;=\; \frac{k(k-1)}{N}.$$

By speeding up time by $N/2$, we obtain

$$\frac{k(k-1)}{N}\,\frac{N}{2} \;=\; \frac{k(k-1)}{2},$$

which corresponds exactly to the coalescent rate of Kingman's n-coalescent. ■

4.4.2 Wright-Fisher model

The Wright-Fisher model is a discrete-time population model. Like the Moran model, the population size N remains fixed. However, it assumes non-overlapping generations. Consider a population at generation $t \in \mathbb{Z}$ consisting of individuals $x_1, \ldots, x_N$. The generation at time $t+1$ consists of individuals $y_1, \ldots, y_N$, where each y_i independently selects a parent uniformly at random from the set $\{x_1, \ldots, x_N\}$.

Let's look at the simplest version of the Wright-Fisher model [32], where the population consists of N genes, each of either type A or type a. Let X_t denote the number of type a genes in the population at generation t. It follows from the model construction that the process $(X_t)_{t\in\mathbb{N}_0}$ is a discrete-time Markov chain on the state space $S = \{0, 1, \ldots, N\}$ with transition probabilities given by

$$p(i,j) = \mathbb{P}(X_{t+1} = j \mid X_t = i) = \binom{N}{j}\left(\frac{i}{N}\right)^j\left(1-\frac{i}{N}\right)^{N-j} \tag{4.8}$$

for all $i, j \in S$. Note that once all individuals are of the same type, this will stay forever. Consequently, the states 0 and N are absorbing states.

We now adopt a different perspective and examine how the ancestral partition in this model is related to Kingman's coalescent.

Theorem 4.19. *Fix $n \geq 1$. Let $x_1, \ldots, x_n$ be n individuals sampled without replacement sampled from a population of size N at time $t = 0$. Let $(\Pi_t^{N,n})_{t\geq 0}$ denote the corresponding ancestral partition process. Then, as $N \to \infty$, and keeping n fixed, speeding up time by a factor N:*

$$\left(\Pi_{Nt}^{N,n}\right)_{t\geq 0} \overset{d}{\Rightarrow} \left(\Pi_t^{(n)}\right)_{t\geq 0} \tag{4.9}$$

under the Skorokhod topology of $D(\mathbb{R}_+, \mathcal{P}_n)$.

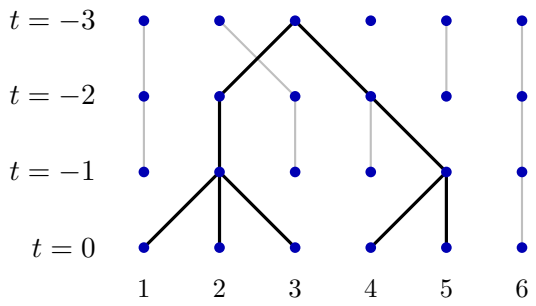


Figure 4.6: Ancestral partition process associated with the Wright-Fisher model. $N = 6$ and the sample consists of individuals $1, 2, 3, 4, 5$. $\Pi_2^{6,5} = \big(\{1, 2, 3\}, \{4, 5\}\big)$ and $\Pi_3^{6,5} = \big(\{1, 2, 3, 4, 5\}\big)$.

Proof. We provide a brief sketch of the derivation under the standard assumptions of the Wright-Fisher model. Consider two individuals, x and y, sampled at random from the current generation. Let P_x and P_y be random variables representing the indices of their respective parents in the preceding generation, where $P_x, P_y \in \{1, 2, \ldots, N\}$.
The probability that x and y share the same parent, is the probability that they choose the same index i out of N possibilities:

$$\mathbb{P}(P_x = P_y) = \sum_{i=1}^{N} \mathbb{P}(P_x = i)\, \mathbb{P}(P_y = i) = N \cdot \frac{1}{N} \cdot \frac{1}{N} = \frac{1}{N}.$$

It follows that the waiting time T_N until the lineages of x and y coalesce is geometrically distributed with success probability $1/N$, i.e.,

$$T_N \sim \text{Geom}\left(\frac{1}{N}\right).$$

For $N \to \infty$ (see Theorem A.6),

$$\frac{1}{N}\, T_N \overset{d}{\Rightarrow} \text{Exp}(1).$$

This implies, that by speeding up time by a factor N and by letting $N \to \infty$, the coalescent rate of x and y is equal to 1. This rate applies to any arbitrary pair, identifying the limit as the pairwise coalescence rate of Kingman's n-coalescent. ■

4.4.3 Möhle's Lemma

A broader class of discrete-time population models, which encompasses the Wright-Fisher model, is known as Cannings models. These models assume a constant population size N in each generation, with individuals indexed as $1, \dots, N$. The reproductive mechanism is described by an exchangeable sequence of random variables $(\nu_1, \dots, \nu_N)$, where $\nu_i \in \mathbb{N}_0$ represents the number of offspring produced by individual i. By the assumption of constant population size, this offspring vector must satisfy the constraint

$$\sum_{i=1}^{N} \nu_i = N. \tag{4.10}$$

Once we have specified the law of $(\nu_i)_{i=1}^N$, we can define our population model for $t \in \mathbb{Z}$ by using i.i.d. copies $\{(\nu_i(t))_{i=1}^N, t \in \mathbb{Z}\}$. The exchangeability of these vectors ensures that individuals are indistinguishable in their reproductive potential, representing a neutral model with no selective advantages or spatial structure.

Define c_N as the probability that two distinct individuals, sampled without replacement from the current generation ($t = 0$), share a common parent in the previous generation ($t = -1$). This probability is called the *coalescence probability* [28]. In a Cannings model with offspring vector $(\nu_1, \dots, \nu_N)$, c_N is given by

$$\begin{aligned} c_N &= \mathbb{E}\left[\sum_{i=1}^{N} \frac{\nu_i}{N} \frac{\nu_i - 1}{N-1}\right] \\ &= \sum_{i=1}^{N} \mathbb{E}\left[\frac{\nu_i(\nu_i - 1)}{N(N-1)}\right] \\ &= \mathbb{E}\left[\frac{\nu_1(\nu_1 - 1)}{N-1}\right], \end{aligned} \tag{4.11}$$

where, in the last step, we used $\nu_i \overset{d}{=} \nu_1$ for all $2 \le i \le N$ by exchangeability.

Remark 4.20 ([13],[28])**.** Note that if $c_N = 0$ we would have the reproduction law $\mathbb{P}(\nu_1 = 1, \dots, \nu_N = 1) = 1$, which corresponds to a population where every individual is replaced by exactly one offspring. Indeed,

$$\mathbb{E}[\nu_1(\nu_1 - 1)] = \mathbb{E}[\nu_1^2] - \mathbb{E}[\nu_1].$$

Given the exchangeability and the constraint (4.10), we have $\mathbb{E}[\nu_1] = 1$. Thus,

$$\mathbb{E}[\nu_1(\nu_1 - 1)] = \mathbb{E}[\nu_1^2] - 1^2 = \mathrm{Var}(\nu_1).$$

If $c_N = 0$, then $\mathrm{Var}(\nu_1) = 0$, implying that ν_1 is a constant $k \in \mathbb{N}_0$ almost surely. Since $\mathbb{E}[\nu_1] = 1$, this constant must be equal to 1. By exchangeability, the same holds for every ν_i for $i \in \{1, \dots, N\}$. Since in the case $c_N = 0$ no coalescence can occur, it is natural to assume $c_N > 0$.

The *coalescent probability* c_N is of fundamental interest in coalescent theory, as $1/c_N$ provides the appropriate time-scale for the convergence of the ancestral partition process of a Cannings model to Kingman's n-coalescent as $N \to \infty$.

Remark 4.21. In the Wright-Fisher model, the joint distribution of the numbers of offspring $\nu_1, \nu_2, \ldots, \nu_N$ for the N individuals in a generation is assumed to be a multinomial distribution [21]:

$$\mathbb{P}(\nu_1 = n_1, \nu_2 = n_2, \ldots, \nu_N = n_N) = \frac{N!}{n_1! n_2! \cdots n_N!} \left(\frac{1}{N}\right)^N.$$

This formulation arises because every offspring chooses its parent independently, and each parent $i \in \{1, \ldots, N\}$ has an identical probability $p = 1/N$ of being chosen. This formula clearly demonstrates the exchangeability of the offspring vector for the Wright-Fisher model.

Theorem 4.22 (Möhle's Lemma)**.** *Consider a Cannings model defined by i.i.d. copies* $\{(\nu_i(t))_{i=1}^N, t \in \mathbb{Z}\}$. *Let* $(\Pi_t^{n,N})_{t\geq 0}$ *denote the ancestral partition process of a sample of size* n, *and let* c_N *be as in* (4.11). *If*

$$\lim_{N\to\infty} \frac{\mathbb{E}[\nu_1(\nu_1 - 1)(\nu_1 - 2)]}{N^2 c_N} = 0 \tag{4.12}$$

is satisfied, then $\lim_{N\to\infty} c_N = 0$ *and the time-rescaled process*

$$\left(\Pi_{t/c_N}^{n,N}\right)_{t\geq 0} \overset{d}{\Rightarrow} \left(\Pi_t^{(n)}\right)_{t\geq 0} \tag{4.13}$$

converges in distribution to Kingman's n*-coalescent* $(\Pi_t^{(n)})_{t\geq 0}$ *in the Skorokhod space* $D(\mathbb{R}_+, \mathcal{P}_n)$.

What does condition (4.12) signify? Let d_N denote the probability that three specific lineages coalesce into a single common ancestor in the previous generation. This probability is given by

$$\begin{aligned} d_N &= \mathbb{E}\left[\sum_{i=1}^N \frac{\nu_i}{N} \frac{\nu_i - 1}{N-1} \frac{\nu_i - 2}{N-2}\right] \\ &= \frac{\mathbb{E}[\nu_1(\nu_1 - 1)(\nu_1 - 2)]}{(N-1)(N-2)} \end{aligned}$$

For large N, the denominator behaves asymptotically like N^2 and therefore

$$\frac{\mathbb{E}[\nu_1(\nu_1 - 1)(\nu_1 - 2)]}{(N-1)(N-2)} \sim \frac{\mathbb{E}[\nu_1(\nu_1 - 1)(\nu_1 - 2)]}{N^2}.$$

This implies,

$$\frac{d_N}{c_N} \sim \frac{\mathbb{E}[\nu_1(\nu_1 - 1)(\nu_1 - 2)]}{N^2 c_N}.$$

Since this value converges to zero as N increases, the probability of three lineages coalescing simultaneously becomes insignificant when compared to the rate of pairwise coalescence. This logic applies to any higher-order merger. This is fundamental to Kingman's model, which assumes that ancestral lines merge exactly two at a time.

Remark 4.23. As previously established, for the Wright-Fisher model, we have $c_N = 1/N$. Furthermore, the probability of observing a triple coalescence event in one generation is given by $d_N = 1/N^2$. Consequently, we have

$$\frac{d_N}{c_N} = \frac{1}{N^2} \cdot N = \frac{1}{N},$$

which vanishes in the limit as $N \to \infty$. Hence, we can use Möhle's Lemma.

4.4.4 Wright-Fisher diffusion and duality

In nature, biological populations are typically very large. It is therefore mathematically convenient to study the limiting behavior of evolutionary models as the population size tends to infinity. Under an appropriate rescaling of time and of the state space, both the Wright-Fisher model and the Moran model converge to the same limiting diffusion process, known as the Wright-Fisher diffusion. This diffusion approximation allows us to address many important questions more effectively, as numerous quantities admit explicit formulas in this continuous framework, whereas they are often much harder to compute in the corresponding discrete-state models.

Recall. A diffusion process $(X_t)_{t\geq 0}$ solving the SDE $dX_t = b(X_t)dt + \sigma(X_t)dB_t$ has infinitesimal generator

$$\mathcal{L}f(x) = b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x), \quad f \in C_b^2(\mathbb{R}).$$

Following the description of the Moran model in Section 4.4.1, we consider a population of size N with two allelic types, a and A. Let $X_t^{(N)}$ denote the number of allele a at time $t \geq 0$, where the superscript indicates the dependence on the total population size N.

Theorem 4.24. *Assume $X_0^{(N)} = pN$ for some $0 < p < 1$. Define the space-time rescaled process as*

$$Y_t^{(N)} := \frac{1}{N}\, X^{(N)}_{\frac{N}{2}t}.$$

Then

$$(Y_t^{(N)})_{t\geq 0} \overset{d}{\Rightarrow} (X_t)_{t\geq 0} \tag{4.14}$$

in the Skorokhod space $D(\mathbb{R}_+, \mathbb{R})$. $(X_t)_{t\geq 0}$ is called the (neutral) Wright-Fisher diffusion and is the unique solution in law of the SDE

$$dX_t = \underbrace{\sqrt{X_t(1-X_t)}}_{\sigma(X_t)}\, dB_t, \quad X_0 = p \tag{4.15}$$

for $(B_t)_{t\geq 0}$ a standard Brownian motion.

Remark 4.25. For the Wright-Fisher diffusion X, the drift coefficient $b(x) = 0$ for all $x \in [0,1]$. Consequently, the process X provides a continuous-time characterization of how allele frequencies fluctuate forward in time only under the influence of genetic drift, in the absence of selection and mutation.

Moreover, the diffusion coefficient $\sqrt{x(1-x)}$ vanishes at the points 0 and 1. Hence, once the process reaches either boundary, it is absorbed there. For functions $f \in C_0^2([0,1]) \coloneqq \{f \in C^2([0,1]) : f(0) = f(1) = 0\}$, the infinitesimal generator of X is given by

$$\mathcal{L}f(x) = \frac{1}{2}x(1-x)f''(x). \tag{4.16}$$

Proof of Theorem 4.24. To prove the convergence of the scaled Moran process to the Wright-Fisher diffusion, we apply the Diffusion Approximation Theorem from Section 7.4 of [16]. According to this result, it suffices to verify:

(i) The initial distributions satisfy $Y_0^{(N)} \overset{d}{\Rightarrow} X_0$ as $N \to \infty$.

(ii) Convergence of the infinitesimal generators.

(iii) For every $\epsilon > 0$

$$\lim_{N\to\infty} \sup_{x\in[0,1]} N\, p_N(x, \{y : |y-x| > \epsilon\}) = 0,$$

where $p_N(x,\cdot)$ is the one-step transition probability measure of the space-time rescaled Moran model.

(iv) Pathwise uniqueness of solution to the SDE (4.15).

We now verify each condition in turn:

Verification of (i): The initial distributions of $Y^{(N)}$ and X coincide, since both processes start at the same constant value p.

Verification of (ii): Let $f \in C_0^2([0,1])$. For the normalized process $(X_t^{(N)}/N)_{t\geq 0}$, the infinitesimal generator $\mathcal{L}_N$ is derived from (4.6) and is given by

$$\mathcal{L}_N f(x) = (Nx(1-x))\Big(f\Big(x+\frac{1}{N}\Big) + f\Big(x-\frac{1}{N}\Big) - 2f(x)\Big).$$

We define $\tilde{\mathcal{L}}_N$ as the infinitesimal generator associated with the process $Y^{(N)}$. Accounting for the chosen time-scaling, it follows that

$$\tilde{\mathcal{L}}_N f(x) = \frac{N}{2}\mathcal{L}_N f(x).$$

Hence,

$$\tilde{\mathcal{L}}_N f(x) = \frac{N}{2}(Nx(1-x))\Big(f(x+\tfrac{1}{N}) + f(x-\tfrac{1}{N}) - 2f(x)\Big).$$

Using Taylor expansion of f around x,

$$f(x+\tfrac{1}{N})+f(x-\tfrac{1}{N})-2f(x)=\frac{1}{N^2}f''(x)+R_N(x),$$

where R_N consists of third- and higher-order terms that vanish when $N\to\infty$. Therefore,

$$\tilde{\mathcal{L}}_N f(x)=\frac{1}{2}x(1-x)f''(x)+\hat{R}_N(x).$$

Let $\mathcal{L}f(x)$ be the infinitesimal generator of the Wright-Fisher diffusion. Then, for $N\to\infty$,

$$\sup_{x\in[0,1]}|\tilde{\mathcal{L}}_N f(x)-\mathcal{L}f(x)|\to 0.$$

Verification of (iii): We know $Y^{(N)}$ is càdlàg by its definition, hence $Y^{(N)}\in D(\mathbb{R}_+,\mathbb{R})$. The jump sizes of this process are bounded by $1/N$. For any fixed $\epsilon>0$, we can choose N large enough such that $1/N<\epsilon$. Thus,

$$p_N\left(x,\left\{y:|y-x|>\frac{1}{N}\right\}\right)=0.$$

Verification of (iv): Next, we use Yamada-Watanabe Theorem (see Theorem 1.57) by taking $\rho(u)=2\sqrt{u}$. The first condition of the Theorem is satisfied since

$$\int_0^\varepsilon \rho(u)^{-2}du=\int_0^\varepsilon\left(\frac{1}{2\sqrt{u}}\right)^2du=\frac{1}{4}\int_0^\varepsilon\frac{1}{u}\,du=\infty$$

Next, we consider the bounded diffusion coefficient $\sigma(x)=\sqrt{x(1-x)}$ for $x\in[0,1]$. To establish the condition $|\sigma(u)-\sigma(v)|\le\rho(|u-v|)$, we apply the algebraic identity $ab-cd=a(b-d)+d(a-c)$ together with the inequality $|\sqrt{u}-\sqrt{v}|\le\sqrt{|u-v|}$ for all $u,v\ge 0$. For any $u,v\in[0,1]$, we have

$$\begin{aligned}
|\sigma(u)-\sigma(v)| &= |\underbrace{\sqrt{u}}_{a}\underbrace{\sqrt{1-u}}_{b}-\underbrace{\sqrt{v}}_{c}\underbrace{\sqrt{1-v}}_{d}|\\
&\le |\underbrace{\sqrt{u}}_{\le 1}||\sqrt{1-u}-\sqrt{1-v}|+|\underbrace{\sqrt{1-v}}_{\le 1}||\sqrt{u}-\sqrt{v}|\\
&\le |\sqrt{1-u}-\sqrt{1-v}|+|\sqrt{u}-\sqrt{v}|\\
&\le \sqrt{|1-u-1+v|}+\sqrt{|u-v|}\\
&\le 2\sqrt{|u-v|}
\end{aligned}$$

Hence, the assumptions of the Yamada-Watanabe Theorem are satisfied, and therefore the solution of (4.15) is pathwise unique.

Finally, since all the required conditions are satisfied, we can conclude that

$$(Y_t^{(N)})_{t\ge 0}\overset{d}{\Rightarrow}(X_t)_{t\ge 0}.$$

■

We also have this convergence to the Wright-Fisher diffusion for a space-time rescaled Wright-Fisher model. Following the description of the model in Section 4.4.2, we consider a population of size N with two allelic types, a and A. Let $X_t^{(N)}$ denote the number of allele a at time $t \geq 0$.

Theorem 4.26 ([16])**.** *Assume $X_0^{(N)} = pN$ for some $0 < p < 1$. Define the space-time rescaled process by*

$$Y_t^{(N)} \coloneqq \frac{1}{N} X_{\lfloor Nt \rfloor}^{(N)}.$$

Then we have

$$(Y_t^{(N)})_{t \geq 0} \overset{d}{\Rightarrow} (X_t)_{t \geq 0} \tag{4.17}$$

in the Skorokhod space $D(\mathbb{R}_+, \mathbb{R})$. X is the Wright-Fisher diffusion with $X_0 = p$.

Consider the allele a as a mutant variant introduced into the population. A central quantity of interest in population genetics is the duration until this mutation either reaches loss or total fixation. This duration, termed the *fixation time*, is formally defined for a population of size N as the first hitting time of an absorbing state:

$$T_N \coloneqq \inf\{t \geq 0 : X_t^{(N)} \in \{0, N\}\}.$$

Given that S is a finite state space where the absorbing states $\{0, N\}$ are accessible from every transient state, it follows that $\mathbb{P}(T_N < \infty) = 1$. Consequently, the process reaches an absorbing state in finite time almost surely.

Theorem 4.27. *Assume that $X_0^{(N)} = pN$ for some $0 < p < 1$. Define the events*

$$D \coloneqq \{X_{T_N}^{(N)} = 0\}, \qquad S \coloneqq \{X_{T_N}^{(N)} = N\}.$$

Then,

$$\mathbb{P}(S) = p, \qquad \mathbb{P}(D) = 1 - p.$$

Moreover, the expected fixation time T_N satisfies the asymptotic relation

$$\mathbb{E}[T_N] \sim -2N\big(p \ln(p) + (1-p)\ln(1-p)\big).$$

Proof. Let $(\mathcal{F}_t)_{t \in \mathbb{N}_0}$ be the canonical filtration of the process $(X_t^{(N)})_{t \in \mathbb{N}_0}$. This process is a martingale since

$$\begin{aligned}
\mathbb{E}\left[X_{t+1}^{(N)} \mid \mathcal{F}_t\right] &= \mathbb{E}\left[X_{t+1}^{(N)} \mid X_t^{(N)}\right] \quad \text{(Markov property)} \\
&= N\left(\frac{X_t^{(N)}}{N}\right) \\
&= X_t^{(N)}.
\end{aligned}$$

Moreover, since $(X_t^{(N)})_{t \in \mathbb{N}_0}$ is a uniformly bounded martingale by N, it is uniformly integrable, justifying the application of the Optional Stopping Theorem (Theorem A.22). Thus, we have $\mathbb{E}(X_0^{(N)}) = \mathbb{E}(X_{T_N}^{(N)})$, where $\mathbb{E}(X_0^{(N)}) = pN$.

Therefore,

$$pN = \mathbb{E}(X_{T_N}^{(N)}) = N\,\mathbb{P}(X_{T_N}^{(N)} = N) + 0 \cdot \mathbb{P}(X_{T_N}^{(N)} = 0) = N\,\mathbb{P}(X_{T_N}^{(N)} = N).$$

By dividing both sides by N, we get

$$\mathbb{P}(S) = p \quad \text{and} \quad \mathbb{P}(D) = 1 - p.$$

Now, we want to use the Wright-Fisher diffusion X with transition semigroup $(P_t)_{t\geq 0}$. Let T denote the time to absorption at either 0 or 1 for the diffusion process and let $f(p)$ be the expected absorption time starting from p, that is,

$$f(p) = \mathbb{E}_p(T).$$

Let $\mathcal{F}'_\varepsilon = \sigma(X_s : s \leq \varepsilon)$ be the canonical filtration of the diffusion. Then,

$$\begin{aligned}
f(x) &= \mathbb{E}_x[T] \\
&= \mathbb{E}_x\big[\mathbb{E}_x[T \mid \mathcal{F}'_\varepsilon]\big] \\
&= \mathbb{E}_x\Big[\mathbb{E}_x\big[(T-\varepsilon) + \varepsilon \mid \mathcal{F}'_\varepsilon\big]\Big] \\
&= \mathbb{E}_x\big[\varepsilon + \mathbb{E}_x[T - \varepsilon \mid \mathcal{F}'_\varepsilon]\big] \\
&= \mathbb{E}_x\big[\varepsilon + \mathbb{E}_{X_\varepsilon}[T]\big] && \text{(Markov property)} \\
&= \varepsilon + \mathbb{E}_x\big[f(X_\varepsilon)\big] \\
&= \varepsilon + P_\varepsilon f(x) \\
&= \varepsilon + f(x) + \varepsilon \mathcal{L} f(x) + o(\varepsilon).
\end{aligned}$$

The term $f(x)$ cancels out from both sides. Dividing the remaining equation by ε and taking the limit as $\varepsilon \to 0$, the $o(\varepsilon)/\varepsilon$ term vanishes, yielding

$$0 = 1 + \mathcal{L} f(x) \implies \mathcal{L} f(x) = -1. \implies \frac{1}{2} p(1-p) f''(p) = -1.$$

The boundary conditions are given by

$$f(0) = \mathbb{E}_0(T) = 0, \qquad f(1) = \mathbb{E}_1(T) = 0.$$

By solving the ordinary differential equation, we get

$$\mathbb{E}_p(T) = -2\big(p \ln(p) + (1-p)\ln(1-p)\big).$$

Moreover, since $\lim_{N\to\infty} 1/N\ \mathbb{E}[T_N] = \mathbb{E}(T)$ (see [16], Chapter 10), we can conclude $\mathbb{E}[T_N] \sim -2N\big(p \ln(p) + (1-p)\ln(1-p)\big)$. ■

A theoretical implication of the diffusion approximation is the duality relation between the block-counting process of Kingman's coalescent and the Wright-Fisher diffusion. To understand this connection, we first provide a formal definition of what it means for two stochastic processes to be in duality.

Definition 4.28 (Duality of processes)**.** Two processes X and Y, with respective laws $\mathbb{E}^{\rightarrow}$ and $\mathbb{E}^{\leftarrow}$, are said to be dual if there exists a function $\psi(x, y)$ such that, for all x, y and for all $t \geq 0$,

$$\mathbb{E}_x^{\rightarrow}[\psi(X_t, y)] = \mathbb{E}_y^{\leftarrow}[\psi(x, Y_t)]. \tag{4.18}$$

Remark 4.29. Duals are not necessarily unique.

Theorem 4.30. *Let $\mathbb{E}^{\rightarrow}$ and $\mathbb{E}^{\leftarrow}$ denote the laws of the Wright-Fisher diffusion $(X_t)_{t\geq 0}$ and of Kingman's coalescent $(\Pi_t^K)_{t\geq 0}$, respectively. Then, for all $0 < p < 1$ and all $n \geq 1$,*

$$\mathbb{E}_p^{\rightarrow}[(X_t)^n] = \mathbb{E}_n^{\leftarrow}\left[p^{\#\Pi_t^K}\right]. \tag{4.19}$$

Here, $\#\Pi_t^K$ denotes the number of non-empty blocks of the random partition Π_t^K.

Remark 4.31. In (4.19), the function ψ from Definition 4.28 is $\psi(x, y) = x^y$.

Proof. (Sketch) Consider the Moran model with a population size N and two possible allele types, A and a. At time $t = 0$, we flip a biased coin for each individual to determine their initial type. Specifically, with probability p the coin lands on heads and the individual carries allele a. We choose n individuals at random from the N individuals at time $Nt/2$. Denote by E the event, that all n chosen individuals are of type a at time $Nt/2$.

Recall from Theorem 4.18 that for $(\Pi_{Nt/2}^{N,n})_{t\geq 0}$, the speeded-up ancestral partition process of the Moran model, and $(\Pi_t^{(n)})_{t\geq 0}$, Kingman's n-coalescent, we have

$$\left(\Pi_{Nt/2}^{N,n}\right)_{t\geq 0} \overset{d}{=} \left(\Pi_t^{(n)}\right)_{t\geq 0}.$$

Therefore, we can deduce that after going backward $Nt/2$ units of time, we have approximately $\#\Pi_t^{(n)}$ distinct ancestral lineages. Assume that after going backward in time for t units, there are i ancestral lineages. Then,

$$\mathbb{P}(E \mid \#\Pi_t^{(n)} = i) = p^i.$$

This implies that

$$\begin{aligned}\mathbb{P}(E) &= \sum_{i=1}^{n} \mathbb{P}(E \mid \#\Pi_t^{(n)} = i)\, \mathbb{P}(\#\Pi_t^{(n)} = i) \\ &= \sum_{i=1}^{n} p^i\, \mathbb{P}(\#\Pi_t^{(n)} = i) \approx \mathbb{E}_n^{\leftarrow}\left(p^{\#\Pi_t^K}\right).\end{aligned}$$

Conversely, Theorem 4.24 states that as $N \to \infty$, the frequency of allele a at time $Nt/2$ is well-approximated by the Wright-Fisher diffusion X_t with initial frequency $X_0 = p$. The chance of randomly drawing a sample of n individuals that all carry allele a is therefore equivalent to the expected value of the n-th power of this random frequency. Thus, we obtain

$$\mathbb{P}(E) \approx \mathbb{E}_p^{\rightarrow}[(X_t)^n].$$

■

Consequently, the duality relation mathematically formalizes two equivalent temporal perspectives of a single probabilistic event. The forward-in-time expression on the left-hand side of (4.19) represents the probability that a random sample of n individuals drawn from the population at time t consists entirely of type a. Symmetrically, the backward-in-time expression on the right-hand side computes the probability that the surviving ancestral lineages of these n individuals at time 0 all carried allele a.

One of the primary advantages of this duality relation is that the block-counting process is often more analytically tractable than the continuous Wright-Fisher diffusion. This can be illustrated by the following calculation:

$$\mathbb{P}(X_\infty = 1 \mid X_0 = p) = \mathbb{E}[X_\infty \mid X_0 = p] = \mathbb{E}[p^{\#\Pi_\infty^K} \mid \#\Pi_0^K = 1] = p,$$

where we use $X_\infty \in \{0, 1\}$ and $\#\Pi_\infty^K = 1$.

4.5 Infinite alleles model

In this section, we consider the *infinite alleles model*. Imagine that together with the evolution forward in time, as described by the Wright-Fisher or Moran model, there exists a process by which mutations occur, inducing differences between the allele observed in a child and that of its parent. As the name implies, this model assumes there are so many possible alleles that each mutation always results in a new type never seen before in the population. To make sense of this assumption, consider a gene consisting of 500 nucleotides. Since there are 4 different nucleotides possible at each position, the number of possible DNA sequences is

$$4^{500} \approx 10^{301}.$$

For any of these possible sequences, there are $3 \cdot 500 = 1500$ sequences that can be reached by a single base change. Assuming equal probability for all replacements, the probability of returning to the original sequence by one specific reverse base change after a mutation is very small. Hence, it is reasonable to treat every mutation as producing a new allele that has never been seen before in the population [13].

To characterize the genetic diversity of a sample in this model, we need to define two fundamental objects:

- The *allelic partition*, denoted Π^n, is the partition of n individuals obtained by grouping those who carry the same allele (i.e., those who share the same mutational history).
- The *allele frequency spectrum* is the vector $(a_1, a_2, \ldots, a_n)$, where each a_i represents the number of distinct alleles carried by exactly i individuals in the sample.

Example 4.32. Imagine we are looking at the genealogy of a specific bird species, focusing on the gene for feather color (see Figure 4.7). We start at the very bottom of the tree with the common ancestor of all birds in our sample. This ancestor had dark blue feathers. If no changes had ever occurred, every bird passing through the generations would also be dark blue. However, as we follow the branches upward through time, a mutation randomly occurred in one specific bird. Under the infinite alleles assumption, this mutation creates a new allele that has not appeared before in the population, say, an allele corresponding to pink feathers. From that point onward, descendants along this branch would inherit the new allele, unless another mutation occurred later in the same lineage.

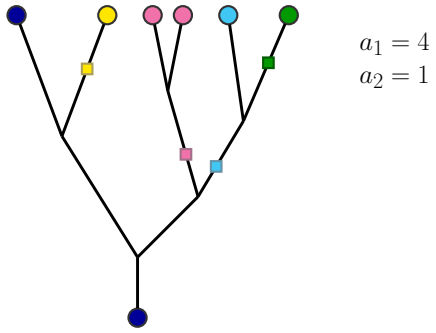


Figure 4.7: Allelic partition generated by mutations (symbolized by the squares) for feather color.

Assume now **(A1)**, **(A2)**, and **(A3)** are satisfied. Hence, the genealogy of a given sample can be modeled by Kingman's coalescent. We further assume that, conditional on the realized coalescent tree, mutations are distributed across the branches according to a Poisson random measure with constant intensity $\theta/2$ ($\theta > 0$) per unit length. We define Π^n as the allelic partition for a sample of n individuals. Given the consistency of these partitions for increasing n, we can define a random partition Π of the set of all natural numbers $\mathbb{N}$ such that $\Pi_{|[n]} = \Pi^n$ for every $n \geq 1$.

Theorem 4.33. *Let* Π *be the random allelic partition obtained from Kingman's coalescent and the infinite alleles model with mutation rate* $\frac{\theta}{2}$ $(\theta > 0)$*. Then,*

$$\Pi \sim \mathrm{PD}(0, \theta).$$

In particular, the probability

$$\mathbb{P}(A_1 = a_1, \ldots, A_n = a_n) = \frac{n!}{\theta(\theta+1)\cdots(\theta+n-1)} \prod_{j=1}^{n} \frac{(\theta/j)^{a_j}}{a_j!}. \tag{4.20}$$

Remark 4.34. Equation (4.20) gives the probability that a $\mathrm{PD}(0, \theta)$ partition of $[n]$ consists of a_1 blocks of size 1, a_2 blocks of size 2, $\ldots$, a_n blocks of size n, subject to the constraint $\sum_{j=1}^{n} j a_j = n$. This result reformulates Ewens' sampling formula (3.15), where the equivalence follows by letting the total number of blocks be $k = \sum_{j=1}^{n} a_j$ and noting that

$$\prod_{i=1}^{k} (n_i - 1)! = \prod_{j=1}^{n} [(j-1)!]^{a_j},$$

while the number of ways to partition the set $[n]$ into this specific configuration is

$$\frac{n!}{\prod_{j=1}^{n} a_j!\ (j!)^{a_j}}.$$

Proof of Theorem 4.33. We want to show that Π can be constructed as a Chinese restaurant process with parameter θ. Suppose the coalescent tree is drawn with the root at the bottom and the n leaves at the top. To every leaf of the tree, there is a unique path between it and the root. And on this path only the first mutation mark encountered when traveling from a leaf to the past matters. Any mutation or coalescence occurring further back do not change the allelic partition formed at the leaves. If no mutation mark is present, we can extend the coalescent tree.

The allelic partition can therefore be defined as follows: *Two leaves belong to the same allele group if and only if they merge with each other before either lineage hits a mutation mark.* From the coalescent tree, we can then extract a corresponding forest. This forest is a collection of subtrees containing all leaves and the nearest mutation marks (see Figure 4.8).

We define

$$0 < T_n < T_{n-1} < \cdots < T_2 < T_1$$

as the times at which the forest undergoes a branch reduction. This drop in the number of lineages occurs whenever we encounter a coalescence event or a mutation mark when tracing the history backwards in time. In both cases the number of branches is reduced by 1. Since we start with n branches (number of sampled individuals), we have exactly n such times.

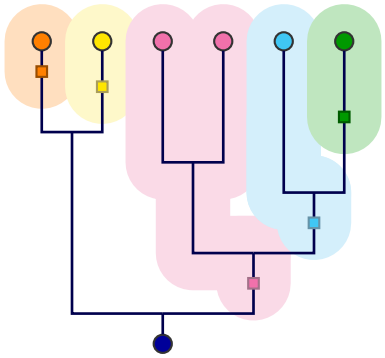


Figure 4.8: Forest of subtrees.

Now, we describe the evolution of the forest from the root upward by examining the sequence of event times $0 = T_1 < T_2 < \cdots < T_n$. We define T_1 as the time at the root. Moving forward in time, each T_m for $2 \leq m \leq n$ represents the emergence of a new lineage. Consequently, on the interval (T_m, T_{m+1}), the forest contains exactly m active lineages. We denote the allelic partition of these m lineages by Π_m^n.

Suppose that $m \geq 1$, and consider the distribution of Π_{m+1}^n given Π_m^n. Looking backward in time, the $(m+1)$-th lineage disappears from the forest for one of two reasons:

- A mutation event: The $(m+1)$-th lineage encounters a mutation mark. Under the infinite alleles assumption, this event represents the origin of a completely new and unique allele. Consequently, in the partition Π_{m+1}^n, this individual forms a new block of size 1.
- A coalescence event: The $(m+1)$-th lineage coalesces with one of the m lineages already present in the forest. This means that the $(m+1)$-th individual inherits the same allele as the lineage with which it merges. Therefore, it joins an existing block of Π_m^n.

We first consider the mutation event. Using the strong Markov property at time T_{m+1} of Kingman's n-coalescent, we obtain

$$\mathbb{P}(\text{new block} \mid \Pi_m^n) = \frac{\frac{(m+1)\theta}{2}}{\frac{m(m+1)}{2} + \frac{(m+1)\theta}{2}} = \frac{\theta}{m+\theta}, \tag{4.21}$$

where $m(m+1)/2$ is the total coalescent rate with $m+1$ lineages present and $(m+1)\theta/2$ the total mutation rate.

Suppose now the block sizes of Π_m^n are $n_1, \ldots, n_k$, where k is the number of distinct alleles discovered up to time T_m and $n_1 + \cdots + n_k = m$. In Kingman's n-coalescent, coalescence events are uniform among pairs. Therefore, given that a coalescence occurs, the probability that the new lineage joins a block with size n_j of the allelic partition Π_m^n is

$$\frac{n_j}{m}.$$

Hence, again using the strong Markov property

$$\mathbb{P}(\text{join block of size } n_j \mid \Pi_m^n) = \frac{\frac{m(m+1)}{2}}{\frac{m(m+1)}{2} + \frac{(m+1)\theta}{2}} \frac{n_j}{m} = \frac{n_j}{m+\theta}. \tag{4.22}$$

(4.21) and (4.22) are precisely the transition probabilities of the Chinese restaurant process with parameter $\theta > 0$. ■

We want to estimate the mutation rate $\theta/2$. Let K_n denote the number of blocks in the random allelic partition Π^n. Let π be a specific partition of n elements into exactly k blocks, with block sizes $n_1, \ldots, n_k$. Then, by conditioning on $K_n = k$ and using Ewens' sampling formula (3.15), we obtain

$$\begin{aligned}
\mathbb{P}(\Pi^n = \pi \mid K_n = k) &= \frac{\mathbb{P}(\Pi^n = \pi)}{\mathbb{P}(K_n = k)} \\
&= \frac{\frac{\theta^k}{\theta(\theta+1)\cdots(\theta+n-1)} \prod_{i=1}^k (n_i - 1)!}{\frac{\theta^k}{\theta(\theta+1)\cdots(\theta+n-1)} \sum_{\pi' : \#\pi' = k} \prod_{i=1}^k (n_i' - 1)!} \\
&= c_{n,k} \prod_{i=1}^k (n_i - 1)!.
\end{aligned}$$

By conditioning on K_n, we obtain an expression that no longer depends on θ. Therefore, K_n is a sufficient statistic for θ (see [33], Chapter 2.1, for further details). In simple terms, this means that knowing K_n is sufficient to estimate the parameter θ, and no additional information is gained from knowing which specific partition Π^n occurred.

Theorem 4.35. *Let Π be a $\mathrm{PD}(0, \theta)$ random partition, and let $\Pi_{|[n]}$ be its restriction to $[n]$, with K_n blocks. Then the following two conditions hold:*

(i)

$$\lim_{n\to\infty} \frac{K_n}{\ln(n)} = \theta \quad a.s. \tag{4.23}$$

(ii)

$$\frac{K_n - \mathbb{E}[K_n]}{\sqrt{\mathrm{Var}(K_n)}} \overset{d}{\Rightarrow} \mathcal{N}(0,1). \tag{4.24}$$

Proof. (i): By the definition of the Chinese restaurant process, each customer sits at a new table independently of the others. We can therefore define independent Bernoulli random variables

$$I_i \sim \mathrm{Ber}\left(\frac{\theta}{\theta + i - 1}\right),$$

where $I_i = 1$ if the i-th customer starts a new block. Consequently, the total number of blocks in the partition is given by the sum

$$K_n = \sum_{i=1}^{n} I_i.$$

Hence,

$$\mathbb{E}[K_n] = \mathbb{E}\left(\sum_{i=1}^{n} I_i\right) = \sum_{i=1}^{n} \mathbb{E}(I_i) = \sum_{i=1}^{n} \frac{\theta}{\theta + i - 1} = \theta \sum_{i=1}^{n} \frac{1}{\theta + i - 1}.$$

The sum can be seen as a Riemann sum that approximates an integral

$$\sum_{i=1}^{n} \frac{1}{\theta + i - 1} \approx \int_{\theta}^{n+\theta} \frac{1}{x}\, dx = \ln(n + \theta) - \ln(\theta) \sim \ln(n). \tag{4.25}$$

Therefore, $\mathbb{E}[K_n] \sim \theta \ln(n)$.

Moreover, since the I_i are independent and $I_1 = 1$,

$$\mathrm{Var}(K_n) = \sum_{i=2}^{n} \mathrm{Var}(I_i) = \sum_{i=2}^{n} \frac{\theta(i-1)}{(\theta + i - 1)^2}.$$

Then, by using $\lim\limits_{i\to\infty} \frac{(i-1)}{(\theta+i-1)} = 1$ and by (4.25), we have

$$\mathrm{Var}(K_n) \sim \sum_{i=2}^{n} \frac{\theta}{\theta + i - 1} \sim \theta \ln(n).$$

Choose a subsequence n_k such that

$$k^2 \leq \mathbb{E}[K_{n_k}] \leq (k+1)^2.$$

Then, using these bounds together with the approximation of $\mathbb{E}[K_n]$, we obtain

$$1 \leq \frac{\ln(n_{k+1})}{\ln(n_k)} \leq \frac{(k+2)^2}{k^2} \quad \overset{k\to\infty}{\longrightarrow} \quad 1. \tag{4.26}$$

Consider

$$A_k = \left\{ \left| \frac{K_{n_k}}{\ln(n_k)} - \theta \right| > \varepsilon \right\}.$$

Then,

$$\mathbb{P}(A_k) = \mathbb{P}\left(|K_{n_k} - \theta \ln(n_k)| > \varepsilon \ln(n_k) \right).$$

By Chebyshev's inequality (A.3) and by $\mathrm{Var}(K_n) \leq \mathbb{E}[K_n]$ for Bernoulli random variables, we have

$$\mathbb{P}(A_k) \leq \frac{\mathrm{Var}(K_{n_k})}{\varepsilon^2(\ln(n_k))^2} \leq \frac{\mathbb{E}[K_{n_k}]}{\varepsilon^2(\ln(n_k))^2} \sim \frac{\theta}{\varepsilon^2 \ln(n_k)} \leq \frac{\theta^2}{\varepsilon^2 k^2}.$$

Hence,

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) \leq \frac{\theta^2}{\varepsilon^2} \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty.$$

By the Borel-Cantelli Lemma,

$$\mathbb{P}(A_k \text{ infinitely often}) = 0.$$

Thus,

$$\lim_{k\to\infty} \frac{K_{n_k}}{\ln(n_k)} = \theta \quad \text{a.s.}$$

Moreover, K_n is monotone and therefore for $n_k \leq n \leq n_{k+1}$, we have $K_{n_k} \leq K_n \leq K_{n_{k+1}}$. This implies,

$$\frac{K_{n_k}}{\ln(n_{k+1})} \leq \frac{K_n}{\ln(n)} \leq \frac{K_{n_{k+1}}}{\ln(n_k)}.$$

By (4.26) and by the Sandwich Theorem, we can deduce

$$\lim_{n\to\infty} \frac{K_n}{\ln(n)} = \theta \quad \text{a.s.}$$

(ii): For this proof we need to apply the Lindeberg-Feller Theorem (Theorem A.11), which generalizes the Central Limit Theorem to triangular arrays of independent random variables. To apply this theorem the following two conditions must be satisfied:

(1) $\lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}[X_{n,i}^2] = \sigma^2 > 0$,

(2) For all $\varepsilon > 0$, $\lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}\Big(|X_{n,i}|^2 \, \mathbb{1}_{|X_{n,i}|>\varepsilon}\Big) = 0$.

Let $X_{n,i} = \frac{I_i - p_i}{V_n}$ for $1 \leq i \leq n$ and $p_i = \frac{\theta}{\theta+i-1}$, $V_n = \sqrt{\mathrm{Var}(K_n)}$.

Verification of (1): The $X_{n,i}$ are independent random variables by construction, and

$$\mathbb{E}[X_{n,i}] = \mathbb{E}\left[\frac{I_i - p_i}{V_n}\right] = \frac{1}{V_n} \, \mathbb{E}[I_i - p_i] = 0.$$

Therefore,

$$\sum_{i=1}^{n} \mathbb{E}[X_{n,i}^2] = \sum_{i=1}^{n} \mathrm{Var}(X_{n,i}) = \frac{1}{V_n^2} \sum_{i=1}^{n} \mathrm{Var}(I_i) = \frac{1}{V_n^2} \mathrm{Var}(K_n) = \frac{V_n^2}{V_n^2} = 1 =: \sigma^2.$$

Verification of (2): Since $I_i \in \{0,1\}$ and $0 < p_i < 1$, we have $|I_i - p_i| < 1$. Hence, for all $1 \leq i \leq n$,

$$|X_{n,i}| = \left|\frac{I_i - p_i}{V_n}\right| < \frac{1}{V_n}.$$

For $n \to \infty$, we have $V_n \to \infty$, since $\mathrm{Var}(K_n) \sim \theta \ln(n)$. Therefore, for every $\varepsilon > 0$ we can find N such that for all $n > N$

$$\frac{1}{V_n} < \varepsilon.$$

Thus, the event $\{|X_{n,i}| > \varepsilon\}$ for $n > N$ is empty, yielding

$$\mathbb{E}\left(|X_{n,i}|^2 \mathbb{1}_{\{|X_{n,i}|>\varepsilon\}}\right) = 0.$$

Finally, by applying the Lindeberg-Feller Theorem, we obtain

$$\sum_{i=1}^{n} X_{n,i} = \frac{K_n - \mathbb{E}[K_n]}{V_n} \overset{d}{\Rightarrow} \mathcal{N}(0,1).$$

■

4.6 Infinite sites model

In the infinite alleles model, we considered a single gene as a whole, and each mutation was assumed to produce an entirely new, previously unseen allele at that locus. However, with the increasing availability of DNA sequence data, it became more natural to investigate genetic variation at the level of individual nucleotides. This was provided by the *infinite sites model*, which assumes that each new mutation occurs at a different nucleotide site. Moreover, once a mutation occurs, it is transmitted unchanged to all the descendants of the corresponding individual and will be visible forever. Thus, the mutations accumulate instead of erasing each other.

To formally characterize this genetic variation, we need to introduce the following two objects:

- S_n is the number of *segregating sites*, representing the total count of loci where genetic variation is present among the n sampled individuals. In biology, these sites are known as Single Nucleotide Polymorphisms (SNPs).
- $M_j(n)$ is the number of sites at which exactly j individuals out of the n sampled individuals have a mutation. The vector $(M_1(n), \ldots, M_n(n))$ is called the *frequency spectrum of the sample*. We have $\sum_{j=1}^{n} M_j(n) = S_n$.

Example 4.36. Suppose we sample 22 humans and examine a specific chromosomal DNA sequence of 300 nucleotides. Variation is observed at a sequence of exactly 26 sites. Hence,

$$S_{22} = 26.$$

In the Table 4.1 below, the uppermost row represents the humans reference sequence of these 26 sites. The rows below represent the distinct sequences found across all 22 sampled individuals, where a dot indicates that the nucleotide matches the reference at that specific locus. As illustrated by the red-highlighted row, 9 individuals share sequence 7. The site frequency spectrum is constructed by evaluating the variant frequencies at each segregating site (i.e., column by column). As an example, the yellow highlights indicate two specific mutations that each occur in exactly four individuals across the sample, contributing a count of 2 to $M_4(22)$.

Ref	TCCGCTCTGTCCCCGCCCTGTTCTTA	
1	..T....C A.T....T.A....TA.C	3
2	........A.T....T..........	2
3	.....C....T....T..........	1
4	..........T....T....A...C.	1
5	.T. A ..T...T.....T..A....C.	2
6	.T. A ......T........A....C.	2
7	.T.........G....T..A....C.	9
8	.T......A.......T..A....C.	1
9	C...G.. C .C..TAA...C..A....	1

Table 4.1: DNA sequence alignment, where dots represent ancestral states and letters indicate derived mutations.

As in the previous model, we assume the genealogy follows Kingman's coalescent with mutations occurring as a Poisson random measure at constant rate $\theta/2$ (for $\theta > 0$) per unit length.

Theorem 4.37. *We have the following two formulas:*

(i)

$$\lim_{n\to\infty} \frac{S_n}{\ln(n)} = \theta \quad a.s. \tag{4.27}$$

(ii) For all $j \geq 1$,

$$\mathbb{E}[M_j(n)] = \frac{\theta}{j}. \tag{4.28}$$

Proof. (i): Let L_n denote the total branch length of a Kingman's coalescent tree for a sample of size n. Let T_k represent the holding time during which the coalescent consists of exactly k active lineages, for $2 \leq k \leq n$. As already seen, the holding

times are independent and $T_k \sim \text{Exp}\binom{k}{2}$. The total length L_n is the sum of the durations of all active lineages across all states, i.e.,

$$L_n = \sum_{k=2}^{n} k\ T_k\ . \tag{4.29}$$

By the linearity of expectation, we have

$$\mathbb{E}[L_n] = \sum_{k=2}^{n} k\ \mathbb{E}[T_k] = \sum_{k=2}^{n} k\ \frac{2}{k(k-1)} = 2\sum_{j=1}^{n-1}\frac{1}{j} = 2h_{n-1} \sim 2\ln(n) \tag{4.30}$$

where h_{n-1} is the $(n-1)$-th harmonic number.

Choosing a fixed $t \in (0, 1/4]$, we define $Y \coloneqq L_n - 2h_{n-1}$. For $x > 0$ and by Markov's inequality (A.2), it follows that

$$\mathbb{P}(Y > x) = \mathbb{P}(e^{tY} > e^{tx}) \leq e^{-tx}\ \mathbb{E}[e^{tY}]. \tag{4.31}$$

By using (4.29) and (4.30), the moment generating function of Y is

$$\begin{aligned}
\mathbb{E}[e^{tY}] &= \mathbb{E}\left[e^{t(L_n - 2h_{n-1})}\right] \\
&= \mathbb{E}\left[e^{t\sum_{k=2}^{n}\left(kT_k - \frac{2}{k-1}\right)}\right] \\
&= \mathbb{E}\left[\prod_{k=2}^{n} e^{t\left(kT_k - \frac{2}{k-1}\right)}\right] \\
&\overset{\perp\!\!\!\perp}{=} \prod_{k=2}^{n} \mathbb{E}\left[e^{t\left(kT_k - \frac{2}{k-1}\right)}\right] \\
&= \prod_{k=2}^{n} e^{-\frac{2t}{k-1}}\ \mathbb{E}\left(e^{tkT_k}\right) \\
&= \prod_{k=2}^{n} e^{-\frac{2t}{k-1}}\ \frac{1}{1-\frac{2t}{k-1}}\ . \qquad \left(\text{use: } kT_k \sim \text{Exp}\left(\frac{k-1}{2}\right)\right)
\end{aligned}$$

Taking logarithms and using logarithm properties,

$$\ln\left(\prod_{k=2}^{n}\frac{e^{-\frac{2t}{k-1}}}{1-\frac{2t}{k-1}}\right) = \sum_{k=2}^{n}\left(-\frac{2t}{k-1} - \ln\left(1-\frac{2t}{k-1}\right)\right).$$

Let $u = 2t/(k-1)$. Since $t \leq 1/4$, we have $u \leq 1/2$. Therefore, by the logarithmic inequality [30],

$$-u - \ln(1-u) \leq 2u^2,$$

we get

$$\sum_{k=2}^{n}\left(-\frac{2t}{k-1} - \ln\left(1-\frac{2t}{k-1}\right)\right) \leq \sum_{k=2}^{n} 2\left(\frac{2t}{k-1}\right)^2 \leq\ 8t^2\sum_{j=1}^{\infty}\frac{1}{j^2} =: C_t < \infty.$$

This logarithmic bound implies

$$\prod_{k=2}^{n} \frac{e^{-\frac{2t}{k-1}}}{1-\frac{2t}{k-1}} \leq e^{C_t}.$$

Putting all together, a bound for (4.31) can be given by

$$\mathbb{P}(L_n - 2h_{n-1} > x) \leq e^{-tx} e^{C_t}. \tag{4.32}$$

By doing the same for the lower tail and getting the same bound as (4.32), we can conclude that

$$\mathbb{P}(|L_n - 2h_{n-1}| > x) \leq 2e^{-tx} e^{C_t}. \tag{4.33}$$

Take now $n = 2^m$. As already seen before, we have $h_{2^m-1} \sim m\ln(2)$. Fix $\varepsilon > 0$. For m sufficiently large we have that

$$\left\{ \left| \frac{L_{2^m}}{m\ln(2)} - 2 \right| > \varepsilon \right\} \subseteq \left\{ |L_{2^m} - 2h_{2^m-1}| > \frac{\varepsilon}{2} m\ln(2) \right\}.$$

By using the bound (4.33), we can deduce that

$$\sum_{m=1}^{\infty} \mathbb{P}\left(\left| \frac{L_{2^m}}{m\ln(2)} - 2 \right| > \varepsilon \right) < \infty.$$

By Borel-Cantelli,

$$\lim_{m\to\infty} \frac{L_{2^m}}{m\ln(2)} = 2 \quad \text{a.s.} \tag{4.34}$$

Fix $\varepsilon' \in (0,1)$. For $X \sim \text{Poisson}(\mu)$, we can find the following bounds [30]:

$$\mathbb{P}(X \geq (1+\varepsilon')\mu) \leq \exp\left(-\mu\left[(1+\varepsilon')\ln(1+\varepsilon') - \varepsilon'\right]\right),$$
$$\mathbb{P}(X \leq (1-\varepsilon')\mu) \leq \exp\left(-\mu\left[\varepsilon' + (1-\varepsilon')\ln(1-\varepsilon')\right]\right).$$

Therefore, we can find $C(\varepsilon') > 0$ such that

$$\mathbb{P}\left(\left| \frac{X}{\mu} - 1 \right| > \varepsilon' \right) \leq e^{-\mu C(\varepsilon')}.$$

Note that, given the coalescent tree $\mathcal{G}$, and therefore the total branch length L_n, the number of segregating sites S_n is distributed as a Poisson random variable with parameter $\mu_n \coloneqq \theta L_n/2$. Therefore, conditioning on the genealogy, we have

$$\mathbb{P}\left(\left| \frac{S_n}{\mu_n} - 1 \right| > \varepsilon' \mid \mathcal{G} \right) \leq 2e^{-C(\varepsilon')\mu_n}.$$

We again restrict our attention to the dyadic indices $n = 2^m$. By (4.34), we know that $\mu_{2^m} \sim \theta m \ln(2)$. Thus, for sufficiently large m, the conditional probabilities are bounded by the terms of a convergent geometric series. By the conditional Borel-Cantelli Lemma, we have

$$\lim_{n\to\infty} \frac{S_{2^m}}{\mu_{2^m}} = 1 \quad \text{a.s.} \tag{4.35}$$

By using (4.34) and (4.35), we can deduce that

$$\lim_{n\to\infty} \frac{S_{2^m}}{m \ln(2)} = 1 \quad \text{a.s.} \tag{4.36}$$

To finally extend the result from dyadic indices to arbitrary $n \in \mathbb{N}$, note that $(S_n)_{n\in\mathbb{N}}$ is non-decreasing. For $2^m \leq n \leq 2^{m+1}$, this yields

$$S_{2^m} \leq S_n \leq S_{2^{m+1}}, \qquad m\ln(2) \leq \ln(n) \leq (m+1)\ln(2).$$

We can write

$$\frac{m}{m+1}\,\frac{S_{2^m}}{m\ln(2)} \leq \frac{S_n}{\ln(n)} \leq \frac{m+1}{m}\,\frac{S_{2^{m+1}}}{(m+1)\ln(2)}. \tag{4.37}$$

As m tends to infinity, both sides of the inequality (4.37) approach θ almost surely. The result for arbitrary n follows by the Sandwich Theorem.

(*ii*): For the proof we refer the reader either to [5] or to [36]. ■

Outlook

This thesis provides an introduction to Kingman's coalescent. A defining feature of Kingman's model is its restriction to binary coalescent events, meaning that only two ancestral lineages can coalesce at any given time. A natural mathematical next step is to relax this restriction, leading to the study of multiple-merger coalescents. For instance, the Λ-coalescent allows several ancestral lineages to merge into a single lineage in one coalescent event ([5], Chapter 3), while the more general Ξ-coalescent permits multiple groups of lineages to merge at the same time ([28], [34]). Exploring these generalized processes builds directly upon the exchangeability and partition theory developed in this thesis and provides a deeper understanding of complex genealogical structures.

Appendices

Appendix A

Theoretical background

In the following, we provide several foundational results utilized throughout this thesis, followed by a selection of relevant biological definitions. [1]

A.1 Distributions

Definition A.1 (Poisson Distribution, [33])**.** A random variable X is said to follow the Poisson distribution with parameter $\lambda > 0$, denoted by $X \sim \text{Poisson}(\lambda)$, if

$$f(x;\lambda) = e^{-\lambda}\frac{\lambda^x}{x!} \quad \text{for} \quad x \in \mathbb{N}_0.$$

The mean, variance and moment generating function (MGF) of $X \sim \text{Poisson}(\lambda)$ are given by

$$\mathbb{E}[X] = \lambda, \qquad \text{Var}[X] = \lambda, \qquad M(t) = \exp(\lambda(e^t - 1)).$$

Lemma A.2 ([33])**.** *Let $X_i \sim \text{Poisson}(\lambda_i)$ for $1 \leq i \leq n$. Then*

$$Y = \sum_{i=1}^{n} X_i \sim \text{Poisson}\left(\sum_{i=1}^{n} \lambda_i\right).$$

Proof. We compute the MGF of Y. Since the X_i are independent,

$$M_Y(t) = \mathbb{E}[e^{tY}] = \mathbb{E}\left[e^{t\sum_{i=1}^{n} X_i}\right] \overset{\perp\!\!\!\perp}{=} \prod_{i=1}^{n} \mathbb{E}\left[e^{tX_i}\right] = M_{X_1}(t) \cdot M_{X_2}(t) \cdots M_{X_n}(t).$$

For a Poisson(λ) random variable X_i, the MGF is

$$M_{X_i}(t) = e^{\lambda_i(e^t-1)}.$$

Therefore,

$$M_Y(t) = e^{\sum_{i=1}^{n} \lambda_i(e^t-1)}.$$

This is exactly the moment generating function of a Poisson $(\sum_{i=1}^{n} \lambda_i)$ random variable. ∎

[1] Unless otherwise cited, the mathematical results in this appendix are based on the Probability I and II lecture courses by Prof. Dr. Jean Bertoin.

Lemma A.3 ([33]). *Let $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ be independent. Then*

$$X \mid (X + Y = k) \sim \text{Binomial}\left(k, \frac{\lambda}{\lambda + \mu}\right).$$

Definition A.4 (Exponential distribution, [33]). A random variable X is said to follow the exponential distribution with parameter $\lambda > 0$, denoted $X \sim \text{Exp}(\lambda)$, if

$$f_X(x; \lambda) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

The mean, variance and moment generating function of $X \sim \text{Exp}(\lambda)$ are given by

$$\mathbb{E}[X] = \lambda^{-1}, \quad \text{Var}[X] = \lambda^{-2}, \quad M(t) = \frac{\lambda}{\lambda - t}, \quad t < \lambda.$$

Theorem A.5 ([32]). *Let I be a countable set and let X_k, $k \in I$, be independent random variables with $X_k \sim \text{Exp}(q_k)$ and $0 < q \coloneqq \sum_{k \in I} q_k < \infty$. Set $X = \inf_k X_k$. Then $X \sim \text{Exp}(q)$.*

Proof. We have

$$\mathbb{P}(X > x) = \mathbb{P}\left(\inf_{k \in I} X_k > x\right) = \mathbb{P}(X_k > x, \forall k \in I) \overset{\perp\!\!\!\perp}{=} \prod_{k \in I} \mathbb{P}(X_k > x).$$

Using the fact that $X_k \sim \text{Exp}(q_k)$, we have $\mathbb{P}(X_k > x) = e^{-q_k x}$. Thus,

$$\prod_{k \in I} e^{-q_k x} = \exp\left(-x \sum_{k \in I} q_k\right) = \exp(-xq).$$

■

Theorem A.6. *Let $X_n \sim \text{Geom}(p_n)$ for $p_n = \lambda/n$. Then $X_n/n \overset{d}{\Rightarrow} Z$, where $Z \sim \text{Exp}(\lambda)$.*

Proof. The moment generating function for a Geometric random variable with parameter λ/n is given by

$$M_{X_n/n}(t) = M_{X_n}\left(\frac{t}{n}\right) = \frac{\frac{\lambda}{n}}{1 - e^{t/n}(1 - \frac{\lambda}{n})} = \frac{\lambda}{n - ne^{t/n}(1 - \frac{\lambda}{n})} = \frac{\lambda}{n(1 - e^{t/n}) + \lambda e^{t/n}}.$$

As $n \to \infty$, we find

$$\lim_{n \to \infty} M_{X_n/n}(t) = \frac{\lambda}{\lambda - t},$$

which is exactly the moment generating function of an exponential random variable with parameter λ. ■

Definition A.7 (Multinomial distribution, [31]). A random vector $(X_1, \ldots, X_s)$ is said to have a multinomial distribution with parameters n and $p_1, \ldots, p_s$ ($s \geq 2$, $n \geq 1$, $p_i \geq 0 \; \forall \, 1 \leq i \leq s$, $p_1 + \ldots + p_s = 1$) if for all $i_1, \ldots, i_s \in \mathbb{N}_0$ mit $i_1 + \cdots + i_s = n$ we have

$$\mathbb{P}(X_1 = i_1, \ldots, X_s = i_s) = \frac{n!}{i_1!\, i_2! \cdots i_s!}\, p_1^{i_1} p_2^{i_2} \cdots p_s^{i_s} \tag{A.1}$$

We denote the distribution of such a random vector by

$$(X_1, \ldots, X_s) \sim \text{Mult}(n; p_1, \ldots, p_s).$$

Lemma A.8 (Invariance of the Uniform Distribution). *Let $f : (0,1) \to (0,1)$ be a measurable, Lebesgue-measure-preserving map, and let $(U_i)_{i \in \mathbb{N}}$ be an i.i.d. sequence of random variables with $U_i \sim \text{Uniform}(0,1)$. Then the sequence $(f(U_i))_{i \in \mathbb{N}}$ is again i.i.d. with $f(U_i) \sim \text{Uniform}(0,1)$ for each $i \in \mathbb{N}$.*

Theorem A.9 (Inverse Transform Sampling, [7]). *Let X be a real-valued random variable with cumulative distribution function F, and let q be the generalized inverse of F, defined by*

$$q(x) = \inf\{y : F(y) > x\}, \quad x \in (0,1).$$

Then, $Y = q(U)$ has the same distribution as X for $U \sim \text{Uniform}(0,1)$.

Theorem A.10 (Central Limit Theorem, [33]). *Let $Y_1, \ldots, Y_n$ be i.i.d. random variables such that $\mathbb{E}[Y_i] = \mu < \infty$ and $\text{Var}[Y_i] = \sigma^2 < \infty$. Let $\overline{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i$. Then*

$$\sqrt{n}\,(\overline{Y}_n - \mu) \overset{d}{\Rightarrow} \mathcal{N}(0, \sigma^2).$$

A generalization of the Central Limit Theorem for triangular arrays of independent random variables is given by the following result:

Theorem A.11 (Lindeberg-Feller Theorem, [14]). *For each n, let $X_{n,m}$, $1 \leq m \leq n$, be independent random variables with $\mathbb{E}[X_{n,m}] = 0$. Suppose*

(1) $\lim\limits_{n \to \infty} \sum_{m=1}^n \mathbb{E}[X_{n,m}^2] = \sigma^2 > 0$,

(2) For all $\varepsilon > 0$, $\lim\limits_{n \to \infty} \sum_{m=1}^n \mathbb{E}\Big(|X_{n,m}|^2 \, \mathbb{1}_{|X_{n,m}| > \varepsilon}\Big) = 0$.

For $n \to \infty$,

$$S_n = X_{n,1} + \cdots + X_{n,n} \overset{d}{\Rightarrow} \mathcal{N}(0, \sigma^2).$$

A.2 Results from probability theory

Definition A.12 ([25])**.** Let $(A_n)_{n\in\mathbb{N}}$ and $(B_n)_{n\in\mathbb{N}}$ be sequences of events in $\mathcal{F}$. Then:

(i) **Continuity from below:** If $A_n \subseteq A_{n+1}$ for every $n \in \mathbb{N}$, then

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

(ii) **Continuity from above:** If $B_{n+1} \subseteq B_n$ for every $n \in \mathbb{N}$, then

$$\mathbb{P}\left(\bigcap_{n\in\mathbb{N}} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n).$$

Definition A.13 (Conditional probability, [10])**.** The conditional probability of A given B is the number

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

defined when $\mathbb{P}(B) > 0$ for A, $B \in \mathcal{F}$.

Theorem A.14 (Bayes' rule of total causes, [10])**.** *For $B_1, B_2, \ldots \in \mathcal{F}$ such that*

$$\sum_{i=1}^{\infty} B_i = \Omega$$

and for all $A \in \mathcal{F}$ we have,

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \mid B_i)\ \mathbb{P}(B_i).$$

Theorem A.15 (Monotone Convergence Theorem)**.** *Let $(X_n)_{n\geq 0}$ be a sequence of non-negative real-valued random variables such that $X_n \leq X_{n+1}$ for all $n \in \mathbb{N}$. If*

$$\lim_{n\to\infty} X_n = X_\infty,$$

then

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X_\infty].$$

Theorem A.16 (Dominated Convergence Theorem)**.** *If $(X_n)_{n\geq 0}$ is a sequence of real-valued random variables and*

$$|X_n| \leq Z, \qquad \mathbb{E}[Z] < \infty, \qquad \lim_{n\to\infty} X_n = X \ \text{a.s.},$$

then

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

Theorem A.17 (Strong Law of Large Numbers, [25]). *Let $(X_n)_{n\geq 0}$ be a sequence of independent and identically distributed real-valued random variables such that $X_1 \in L^1(\mathbb{P})$. Then*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mathbb{E}[X_1] \quad a.s.$$

Theorem A.18 (Markov's inequality). *Let $X \in L^1(\mathbb{P})$ and $a > 0$. Then*

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[|X|]}{a}. \tag{A.2}$$

Theorem A.19 (Chebychev's inequality). *If $X \in L^2(\mathbb{P})$ and $a > 0$, then we have*

$$\mathbb{P}(|X - \mathbb{E}[X]\ | > a) \leq \frac{\mathrm{Var}(X)}{a^2}. \tag{A.3}$$

Lemma A.20 (Borel-Cantelli Lemma, [25]). *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of events.*
(i) If $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$, then

$$\mathbb{P}(\limsup_{n\to\infty} A_n = 0)$$

or equivalently

$$\sum_{n\in\mathbb{N}} \mathbb{1}_{A_n} < \infty \quad a.s.$$

(ii) If $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$ and A_n are independent events for all $n \in \mathbb{N}$, then

$$\mathbb{P}(\limsup_{n\to\infty} A_n = 1)$$

or equivalently

$$\sum_{n\in\mathbb{N}} \mathbb{1}_{A_n} = \infty \quad a.s.$$

Definition A.21 (Discrete-time martingale). Let $(X_n)_{n\geq 0}$ be a random process such that $X_n \in L^1(\mathbb{P})$ for all $n \geq 0$. Then $(X_n)_{n\geq 0}$ is called a martingale if, for any $n \geq 0$, we have

$$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = X_n \quad \text{a.s.}$$

We use the following version of Doob's Optional Stopping Theorem for discrete-time martingales:

Theorem A.22 (Optional Stopping Theorem). *Let $(X_n)_{n\geq 0}$ be a uniformly integrable martingale, i.e.,*

$$\lim_{r\to\infty} \sup_{n\geq 0} \mathbb{E}\big[|X_n|\mathbf{1}_{\{|X_n|>r\}}\big] = 0.$$

Let T be a stopping time. Then

$$\mathbb{E}[X_T] = \mathbb{E}[X_0],$$

where $X_T := X_\infty$ on the event $\{T = \infty\}$, and X_∞ exists almost surely by uniform integrability.

A.3 Change of meassure

Definition A.23 (Change of measure, [11])**.** Let $Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ with $Z \geq 0$ and $\mathbb{E}_{\mathbb{P}}[Z] = 1$. Then, Z defines a new probability measure $\mathbb{Q}$ on $\mathcal{F}$ via

$$\mathbb{Q}(A) \coloneqq \int_A Z d\mathbb{P} = \mathbb{E}_{\mathbb{P}}[Z \mathbb{1}_A].$$

We call Z a density of $\mathbb{Q}$ w.r.t. $\mathbb{P}$ (or Radon-Nicodym derivative).

Definition A.24 (Absolutely continuous, [11])**.** Let $\mathbb{Q}$ and $\mathbb{P}$ be two probability measures on a measurable space $(\Omega, \mathcal{F})$. We say that $\mathbb{Q}$ is absolutely continuous w.r.t. $\mathbb{P}$ on $\mathcal{F}$, and write $\mathbb{Q} \ll \mathbb{P}$ if for all $A \in \mathcal{F}$, we have

$$\mathbb{P}(A) = 0 \;\Rightarrow\; \mathbb{Q}(A) = 0.$$

If $\mathbb{Q} \ll \mathbb{P}$ and $\mathbb{P} \ll \mathbb{Q}$, we say that the probability measures $\mathbb{Q}$ and $\mathbb{P}$ are equivalent and write $\mathbb{Q} \sim \mathbb{P}$.

Theorem A.25 ([11])**.** *(a) $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$ if and only if there exists a density of $\mathbb{Q}$ with respect to $\mathbb{P}$.*

(b) $\mathbb{Q}$ is equivalent to $\mathbb{P}$ if and only if there exists a density of $\mathbb{Q}$ with respect to $\mathbb{P}$ that is strictly positive.

The notation used for the density is as follows:

$$\mathbb{Q} = Z \cdot \mathbb{P}, \qquad d\mathbb{Q} = Z \cdot d\mathbb{P}, \qquad Z = \frac{d\mathbb{Q}}{d\mathbb{P}}.$$

Proposition A.26 ([11])**.** *Let Z be the Radon-Nicodym derivative of $\mathbb{Q}$ w.r.t. $\mathbb{P}$. Then,*

$$\mathbb{E}_{\mathbb{Q}}[Y] = \mathbb{E}_{\mathbb{P}}[ZY]$$

for all $Y \in L^1(\Omega, \mathcal{F}, \mathbb{Q})$.

A.4 Analytical results

Definition A.27 (Gamma function, [27])**.** The Gamma function Γ is defined for all $t > 0$ by

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x}\, dx.$$

It satisfies the following fundamental properties:

- the functional equation
$$\Gamma(t+1) = t\,\Gamma(t),$$
- and for every $n \in \mathbb{N}$,
$$\Gamma(n) = (n-1)!.$$

Definition A.28 (Beta function, [27])**.** The Beta function B is defined for $\alpha, \beta > 0$ as follows:

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx.$$

Remark A.29 ([27])**.** The Beta function B and the Gamma function Γ are connected by

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Proof.

$$\begin{aligned}\Gamma(z_1)\Gamma(z_2) &= \int_0^\infty e^{-u}u^{z_1-1}\, du \cdot \int_0^\infty e^{-v}v^{z_2-1}\, dv \\ &= \int_0^\infty \int_0^\infty e^{-u-v}u^{z_1-1}v^{z_2-1}\, du\, dv.\end{aligned}$$

We change the variables: $u = st$ and $v = s(1-t)$. Since $u + v = s$ and $\frac{u}{u+v} = t$, we have that the limits of integration for s are from 0 to ∞, and for t from 0 to 1. Therefore, by changing variables we get

$$\begin{aligned}\Gamma(z_1)\Gamma(z_2) &= \int_0^\infty \int_0^1 e^{-s}(st)^{z_1-1}(s(1-t))^{z_2-1} s\, dt\, ds \\ &= \int_0^\infty e^{-s}s^{z_1+z_2-1}\, ds \cdot \int_0^1 t^{z_1-1}(1-t)^{z_2-1}\, dt \\ &= \Gamma(z_1+z_2) \cdot B(z_1, z_2).\end{aligned}$$

■

A.5 Kolmogorov's Extension Theorem

Suppose S is a Polish space. Let $S^{\mathbb{N}}$ denote the product space $S \times S \times \dots$, which is viewed as the set of sequences $(x_1, x_2, \dots)$ where $x_i \in S$. Let $\mathcal{B}(S^n)$ denote the Borel σ-algebra on S^n.

Definition A.30 (Consistency, [4])**.** Let μ_n be probability measures on S^n for $n \geq 1$. The sequence $(\mu_n)_{n\geq 1}$ is said to be consistent if

$$\mu_{n+1}(A \times S) = \mu_n(A)$$

for $A \in \mathcal{B}(S^n)$.

Theorem A.31 (Kolmogorov's Extension Theorem, [4])**.** *Suppose that for each n, we have a probability measure μ_n on S^n. If the measures μ_n are consistent, then there exists a unique probability measure μ on $(S^{\mathbb{N}}, \mathcal{B}(S^{\mathbb{N}}))$ such that*

$$\mu(A \times S \times S \times \dots) = \mu_n(A)$$

for all $A \in \mathcal{B}(S^n)$.

Appendix B

Basic biological terminology

We introduce some essential terminology from biology [3]:

- **DNA:** Genetic material carrying hereditary information.
- **Gene:** A piece of the DNA, that encodes for a protein.
- **Nucleotides:** The building blocks of DNA, consisting of four types: Adenine (A), Cytosine (C), Guanine (G), and Thymine (T). In DNA base pairing, A pairs with T and C pairs with G.
- **Allele/Type:** One of two or more variants of a gene.
- **Mutation:** A change in the DNA sequence that may lead to genetic variation and drives evolution.
- **Natural selection:** The differential survival and reproduction (fitness) of individuals due to heritable differences in phenotype.
- **Genetic drift:** Random fluctuations in allele frequencies caused by random sampling.
- **Genetic variation:** The diversity of alleles within a population.
- **Diploid:** In diploid populations every individual carries two copies of each gene. Example: Humans
- **Haploid:** In haploid populations every individual carries one copy of each gene. Example: Bacteria

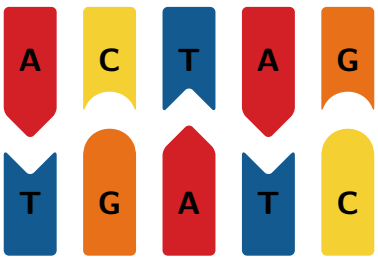


Figure B.1: Schematic representation of DNA base pairing.

Bibliography

[1] D. J. Aldous. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour XIII - 1983*. Springer, 1985.

[2] V. Bansaye and S. Méléard. *Stochastic Models for Structured Populations*. Mathematical Biosciences Institute Lecture Series. Springer, 2015.

[3] K. Basler. Molekulare und Klassische Genetik, 2021. Fall Semester.

[4] R. F. Bass. *Stochastic Processes*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2011.

[5] N. Berestycki. Recent Progress in Coalescent Theory. *Ensaios Matemáticos*, 2009. URL `https://doi.org/10.48550/arXiv.0909.3985`.

[6] J. Bertoin. *Random Fragmentation and Coagulation Processes*. Cambridge University Press, 2006.

[7] J. Bertoin. Probabilités: Cours de Licence de Mathématiques Appliquées, LM 390, n.d.

[8] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. John Wiley & Sons, New York, 2nd edition, 1999.

[9] B. Błaszczyszyn. Lecture Notes on Random Geometric Models - Random Graphs, Point Processes and Stochastic Geometry, 2017. URL `https://inria.hal.science/cel-01654766v3`. Lecture notes, HAl Open Science.

[10] P. Brémaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation and Queues*, volume 31 of *Texts in Applied Mathematics*. Springer, 2nd edition, 2020.

[11] D. Coculescu. FIM202 Mathematical Finance II, 2024. Fall Semester.

[12] T. Duquesne. Processus markoviens 1: Chaînes de markov, processus de markov à temps continu et espace d'état discret, nuages poissonniens, 2023-2024. Lecture notes.

[13] R. Durrett. *Probability Models for DNA Sequence Evolution*. Probability and Its Applications. Springer, 2nd edition, 2008.

[14] R. Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 5th edition, 2019.

[15] E. B. Dynkin. *Markov Processes*, volume 1 of *Die Grundlehren der mathematischen Wissenschaften*. Springer, 1965. Translated with the authorization and assistance of the author by J. Fabius et al.

[16] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 1986.

[17] R. Grassl and O. Levin. Bell Numbers, n.d. URL `https://discrete.openmathbooks.org/more/mdm/sec_adv-bell.html`.

[18] G. R. Grimmet. *Percolation*. Springer, 2nd edition, 1999.

[19] O. Kallenberg. *Random Measures, Theory and Applications*, volume 77 of *Probability Theory and Stochastic Modelling*. Springer, 2017.

[20] J. F. C. Kingman. The coalescent. *Stochastic Processes and their Applications*, 13:235–248, 1982. URL `https://people.eecs.berkeley.edu/~jordan/sail/readings/archive/kingman-coalescent1982.pdf`.

[21] J. F. C. Kingman. On the Genealogy of Large Populations. *Journal of Applied Probability*, 19:27–43, 1982. URL `https://www.jstor.org/stable/3213548`. Essays in Statistical Science.

[22] J.F.C. Kingman. *Poisson Processes*. Oxford University Press Inc., 1993.

[23] A. Klenke. *Wahrscheinlichkeitstheorie*. Springer Spektrum, 4th edition, 2020.

[24] G. Last and M. Penrose. *Lectures on the Poisson Process*. Institute of Mathematical Statistics textbooks. Cambridge University Press, 2018.

[25] J.-F. Le Gall. *Measure Theory, Probability and Stochastic Processes*. Springer, 2022.

[26] H. M. Mahmoud. *Pólya Urn Models*. CRC Press, 2009.

[27] T. Michaels. *Analysis 1: Eine praxisorientierte Einführung für Mathematiker und Physiker mit über 900 gerechneten Beispielen*. Editrices, 2nd edition, 2015.

[28] M. Möhle and S. Sagitov. A Classification of Coalescent Processes for Haploid Exchangeable Population Models. *The Annals of Probability*, 29(4):1547–1562, 2001. URL `https://www.jstor.org/stable/2691969`.

[29] D. Navarro and A. Perfors. The Chinese restaurant process, n.d. URL `https://compcogsci-3016.djnavarro.net/technote_chineserestaurantprocesses.pdf`. Lecture notes.

[30] A. Nikeghbali. Advanced Mathematics for Data Science, 2024. Spring Semester.

[31] H. Norbert. *Stochastik für Einsteiger: Eine Einführung in die faszinierende Welt des Zufalls*. Springer Spektrum, 14th edition, 2023.

[32] J. R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2009.

[33] V. M. Panaretos. *Statistics for Mathematicians: A Rigorous First Course*. Birkhäuser, 2016.

[34] J. Schweinsberg. Coalescents with Simultaneous Multiple Collisions. *Electronic Journal of Probability*, 5:1–50, 2000. URL `http://www.math.washington.edu/~ejpecp/EjpVol5/paper12.abs.html`.

[35] K. Siegrist. Time reversal in discrete-time chains, 2022. URL `https://stats.libretexts.org/Bookshelves/Probability_Theory/Probability_Mathematical_Statistics_and_Stochastic_Processes_(Siegrist)/16:_Markov_Processes/16.07:_Time_Reversal_in_Discrete-Time_Chains`.

[36] J. Wakeley. *Coalescent theory: An introduction*. W. H. Freeman and Company, 2009.

This thesis was prepared with the assistance of tools such as ChatGPT (OpenAI) and Google Gemini, used for linguistic refinements and LaTeX-related tasks.

List of symbols

$\mathbb{1}_B$	Indicator function of a set B
δ_x	Dirac mass at x
$\delta_{x,y}$	Kronecker Delta
B^c	Complement of the set B in a sample space Ω
$\#B$	Cardinality of a set B
$\|I\|$	Length of the interval I
$\lfloor x \rfloor$	$\max\{n \in \mathbb{Z} \mid n \leq x\}$
$\mathbb{N}$	Set of positive integers $\{1, 2, 3, \dots\}$
$\mathbb{N}_0$	$\mathbb{N} \cup \{0\}$
$\mathbb{Z}$	Set of integers $\{\dots, -2, -1, 0, 1, 2, \dots\}$
$\mathbb{R}$	Real numbers
$\mathbb{R}_+$	Non-negative real numbers
$[n]$	Set of the first n integers $\{1, 2, \dots, n\}$
$[\infty]$	Different notation for $\mathbb{N}$
$\mathcal{P}_n$	Space of partitions of $[n]$
$\mathcal{P}_\infty$	Space of partitions of $\mathbb{N}$
$\mathcal{S}_0$	Space of mass-partitions
$D(E, F)$	Space of càdlàg functions $f : E \to F$, called Skorokhod space
$C_b(E)$	Set of bounded functions $f : E \to \mathbb{R}$
$C_b^2(E)$	Set of bounded functions $f : E \to \mathbb{R}$ with continuous and bounded derivatives up to second order
$C_0^2([0, 1])$	Set of functions $f \in C^2([0, 1])$ such that $f(0) = f(1) = 0$
π	Partition usually in $\mathcal{P}_n$ or $\mathcal{P}_\infty$
π_i	i-th block of partition π
$\pi_{\|B}$	Partition π restricted to a set B
$\#\pi$	Number of non-empty blocks of a partition π
$\|\pi_i\|$	Asymptotic frequency of block π_i
$\Pi^{(n)}$	Kingman's n-coalescent
Π^K	Kingman's coalescent
$\Pi^{N,n}$	Ancestral partition process with population size N and sample size n