TIME TO COALESCEENCE FOR A CLASS OF NONUNIFORM ALLOCATION PROCESSES

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of the Ohio State University

By

John Kingen McSweeney, B.S., M.S.

* * * * *

The Ohio State University
2009

Dissertation Committee:
Professor Boris G. Pittel, Advisor
Professor Gerald Edgar
Professor Janet Best

Approved by

Advisor
Graduate Program in Mathematics
ABSTRACT

We study a so-called coalescent process that can be described as follows: start with a set of $n$ boxes and $b_0$ balls. Let $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ be any probability vector. Throw each ball into box $j$ with probability $p_j$, independently for each ball. Any balls that land in the same box are fused into a single ball, and the process is repeated with this (possibly smaller) number of balls. Continue this process until there is only one ball left; the time at which this happens is called the coalescence time, denoted $T$. This problem can also be phrased in the context of population genetics, where it is referred to as the Generalized Wright-Fisher Model. In that formulation, the balls represent ancestral lineages, and $T$ is the number of generations back in time one has to go to find a common ancestor for $b_0$ individuals from the current generation. We shall mainly study the expected coalescence time $E[T]$. For $b_0 = n$, and $\mathbf{p}$ nonuniform, little is known about the expected time spent when the number of balls is relatively large. We show that for vectors $\mathbf{p}$ satisfying a mild uniformity condition, this quantity is negligible compared to the expected time spent when the number of balls is “small”, which is asymptotically $2(\sum_j p_j^2)^{-1}$. We further show that this condition is sharp, in that if it is not met, there are vectors $\mathbf{p}$ which give rise to processes which do not have this qualitative behavior, and thus where the expected coalescence time far exceeds $2(\sum_j p_j^2)^{-1}$.
To my Parents, Kerry and Susanne
ACKNOWLEDGMENTS

I would first like to thank my advisor Boris Pittel. His steady yet enthusiastic guidance, as well as his unwillingness to accept anything but my best efforts, have made the current work possible. I am thankful for all the time he has been able to give me; our meetings are invariably a pleasure because his passion for the material is so contagious.

I would also like to thank my fellow students, especially Adam Hammett, Craig Lennon, and Jia Yeum who, over the course of numerous seminars, not only followed attentively but offered many constructive suggestions.

I have benefited from the tutelage of several faculty and staff here at Ohio State, in particular my dissertation committee members, Professors Edgar and Best, who have graciously taken time from their busy schedules to carefully review this work and offer helpful comments and suggestions.

I am also indebted to the participants of the 2008 Summer School in Probability at Cornell University who helped me view my work from a biological point of view, which allowed me to gain a deeper understanding of it and paved the way for some natural questions for future research.

My family, while of little help in the gritty details of my mathematical excursions,
has been a great help in shaping my overall intellectual and academic outlook; specifically my parents Kerry and Susanne, as well as my sister Kendra, who has been very helpful in getting me through the day-to-day challenges which graduate school brings.
VITA

1980   Born - Kingston, Ontario, Canada

2002   B.S. in Applied Mathematics (hons.), McGill University

2004   M.S. in Mathematics, The Ohio State University

2002-2007   VIGRE Graduate Fellow, The Ohio State University

2007-Present   Graduate Teaching Associate, The Ohio State University
PUBLICATIONS

Research Papers


FIELDS OF STUDY

Major Field: Mathematics

Specialization: Combinatorial Probability
# TABLE OF CONTENTS

Abstract . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . ii
Dedication . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . iii
Acknowledgments . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . iv
Vita . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . vi
List of Figures . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . x

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1</td>
</tr>
<tr>
<td>1.1</td>
<td>Setup . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1</td>
</tr>
<tr>
<td>1.2</td>
<td>The Most Recent Common Ancestor . . . . . . . . . . . . . . . . . . . . . . . . 2</td>
</tr>
<tr>
<td>1.3</td>
<td>Phylogenetic Tree . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3</td>
</tr>
<tr>
<td>1.4</td>
<td>Relation to Branching Processes . . . . . . . . . . . . . . . . . . . . . . . . 4</td>
</tr>
<tr>
<td>1.5</td>
<td>Multinomial Case . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5</td>
</tr>
<tr>
<td>1.6</td>
<td>Balls-into-boxes Formulation . . . . . . . . . . . . . . . . . . . . . . . . . . 6</td>
</tr>
<tr>
<td>1.7</td>
<td>Applications to Computer Science . . . . . . . . . . . . . . . . . . . . . . . . 7</td>
</tr>
<tr>
<td>1.8</td>
<td>General Markov Case . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8</td>
</tr>
<tr>
<td>1.9</td>
<td>Outline of the Dissertation . . . . . . . . . . . . . . . . . . . . . . . . . . . 9</td>
</tr>
<tr>
<td>2</td>
<td>Previous Work, Statement of Theorems . . . . . . . . . . . . . . . . . . . . . 10</td>
</tr>
<tr>
<td>2.1</td>
<td>Kingman’s coalescent . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10</td>
</tr>
<tr>
<td>2.1.1</td>
<td>Discrete-time process . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10</td>
</tr>
<tr>
<td>2.1.2</td>
<td>Continuous-time process . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12</td>
</tr>
<tr>
<td>2.1.3</td>
<td>Limiting Behavior . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13</td>
</tr>
<tr>
<td>2.2</td>
<td>Previous Results for the Nonuniform Case . . . . . . . . . . . . . . . . . . . 15</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Möhle . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Adler et al. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16</td>
</tr>
</tbody>
</table>

viii
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>4</td>
</tr>
</tbody>
</table>

Tracing the MRCA in a Phylogenetic Tree
CHAPTER 1
INTRODUCTION

1.1 Setup

The problem central to this thesis is most easily described through the language of population genetics. Take a finite population which reproduces according to some random mechanism. We will assume that the population is haploid; that is, each individual has exactly one parent. This assumption is valid for certain organisms such as bacteria and mitochondrial DNA. In diploid populations (where each individual has two parents) such as humans, we can still use this framework by, for example, restricting our attention to the females in the population and ignoring the males. We also assume that the generations are discrete and non-overlapping: each generation is indexed by \( t, t = 0, 1, 2, \ldots \), and all individuals of a given generation produce their offspring (if any) simultaneously, and die immediately thereafter (in terms of the genealogy, we need not assume that they die right after reproduction, only that they do not reproduce ever again). Suppose that such a process has been going on for an arbitrarily long amount of time, and let the current generation be labelled with \( t = 0 \). For the sake of convenience, we will index the generations backwards in time; that is,
the parents of the current generation are alive at time $t = 1$; the grandparents, $t = 2$; and so on.

1.2 The Most Recent Common Ancestor

Define $n(t)$ to be the number of individuals alive in the $t$-th generation backwards in time (this could in general be random). For an individual $i$ alive at generation $t$ ($1 \leq i \leq n(t)$), let $\nu_i(t)$ denote the (random) number of offspring it produces. It must therefore be the case that $\nu_1(t) + \nu_2(t) + \cdots + \nu_{n(t)}(t) = n(t-1)$ for any $t$. The random vectors $\mathbf{N}(t) := (\nu_1(t), \nu_2(t), \ldots, \nu_{n(t)}(t))$ will always be assumed independent and identically distributed over all $t$; that is, the reproductive mechanism does not change from one generation to the next. However, for a given $t$, the random variables $\nu_1(t), \ldots, \nu_{n(t)}(t)$ need not be independent — within a given generation, there may be factors such as competition for limited resources which makes one individual’s offspring size depend on another’s. In fact, we will mostly be considering the case where $n(t)$ is fixed value $n$, in which case the $\nu_i(t)$ cannot be independent except for certain trivial distributions. This constant population assumption might be reasonable, for example, in the case of a bacterial population with a sufficiently strong propensity to reproduce, but which are constrained to live in a Petri dish of fixed size.

Unless otherwise noted, once we condition on the $\nu_i(t)$, the assignment of children to generation $t$ is performed uniformly at random among all allowable assignments (that is, among all assignments of offspring that assign $\nu_i(t)$ children to individual $i$ for each $i$). This can be thought of as saying that as children, the individuals are
indistinguishable (but not necessarily as parents). We will call this the memorylessness property and we may discuss some simple cases where this is not assumed to hold. For a given \( b_0 \leq n(0) \), the coalescence time \( T_{b_0} \) is defined to be the smallest number of generations backwards in time we must go to find an individual who is an ancestor to \( b_0 \) individuals chosen from the current population (the memoryless property ensures that the distribution of \( T_{b_0} \) does not depend on which \( b_0 \) individuals are chosen). This individual is the Most Recent Common Ancestor (MRCA) for these \( b_0 \) individuals. In the case where \( b_0 = n(0) \), we seek the time back to the MRCA of the entire current population. Once the MRCA is reached, we say that coalescence of these \( b_0 \) ancestral lines has occurred.

1.3 Phylogenetic Tree

If the individuals and their kinship are represented through a standard phylogenetic tree, the MRCA is found by tracing the lineages of \( b_0 \) individuals back up the tree until all their ancestral lines merge. When we talk of a ‘number of lineages’ at a given time, it is understood that we mean ‘number of lineages for which there is a descendant alive at time \( t = 0 \)’. Figure 1.1 is a tree where the lineages to the MRCA are traced back for the entire population of \( n = 6 \) individuals. In this case, the coalescence time is \( T_6 = 5 \). All individuals with descendants alive at time \( t = 0 \) are marked with an ‘x’.
1.4 Relation to Branching Processes

Over one hundred years ago, Francis Galton and H. W. Watson [23] considered a related inverse problem, motivated by computing the probability of extinction of surnames (there is no danger of extinction of these authors’ names, as they live on through the so-called Galton-Watson Branching Process). The inspiration for their research was to explain “[t]he decay of the families of men who occupied conspicuous positions in past times”, and to determine whether it was the case that “a rise in physical comfort and intellectual capacity is necessarily accompanied by diminution in ‘fertility’, ” or whether such extinctions were merely a mathematical inevitability. Their setup is isomorphic to ours if, in a given population, we consider the last names passed down from father to son (for the purpose of survival of last names, any
daughters born are irrelevant). Their analysis however proceeded “forwards” in time: they considered a given individual and computed the (forwards) time (if any) after which he will have no descendants; we take a given generation and look backwards to determine the time at which only one individual has any descendants. Indeed, coalescent processes can be regarded generally as inverses of branching processes; see Möhle [14] for a much more detailed account of this relation.

1.5 Multinomial Case

The case of most interest to us is when \( n(t) = n \) is constant, and the offspring sizes are multinomially distributed according to some probability vector \( \mathbf{p} = (p_1, \ldots, p_n) \), independently across generations. That is, identically and independently for all \( t \),

\[
P(N(t) = (m_1, m_2, \ldots, m_n)) = \binom{n}{m_1 \ldots \cdot m_n} p_1^{m_1} \ldots p_n^{m_n},
\]

for \( m_1 + \cdots + m_n = n \), and 0 otherwise. This is referred to in the biological context as the Generalized Wright-Fisher Model; the case \( \mathbf{p} = \mathbf{u} := (1/n, \ldots, 1/n) \) is called the neutral Wright-Fisher Model.

Following Wakeley [22], here is a biologically-motivated interpretation of this model. Consider \( p_j \) to be a (scaled) measure of the “reproductive fitness” of individual \( j \). In a given generation, each individual \( j \) then produces \( M \cdot p_j \) offspring, where \( M \gg n \) is meant to be a very large number. However, we require that the population size stay constant at \( n \); therefore, immediately after the birth of these \( M \) individuals, we select \( n \) of them (without replacement) uniformly at random to survive until maturity, obtaining \( \nu_j \) surviving offspring for individual \( j \). For \( M \) large enough, we may ignore
the discrepancy between sampling with or without replacement, and therefore the distribution of \((\nu_1(t), \ldots, \nu_n(t))\) is Multinomial\((p_1, \ldots, p_n)\). It should also mentioned that this mechanism is *selectively neutral*, in that once the \(n\) individuals which survive are chosen, the fitness levels for the next generation are assigned uniformly at random (i.e. independently of the parentage). From now on when we refer to ‘offspring’, we will only mean any individuals that survive to maturity, as for the sake of the reproductive process, all others can be safely ignored.

Note that for any given \(t\), each \(\nu_j(t)\) is marginally a Binomial\((n, p_j)\) random variable, but that \(\nu_1(t), \ldots, \nu_n(t)\) are certainly not independent, as they must sum to \(n\). This lack of independence makes the analysis of this process unwieldy from this ‘forwards’ point of view. However, there is a very convenient ‘backwards’ interpretation, which is the following: at each generation, each offspring chooses its parent according to the distribution \(p\), and all of these choices are *independent*. This interpretation of course is biologically nonsensical, but this independence makes this formulation much easier to work with. So now we’ve got our independence — what are we going to do with it? One of our main tools will be to build a generating function to express \(\pi_{kb}\), the probability of going from \(k\) ancestral lines to \(b\) lines in one generation, where this independence plays a crucial role.

### 1.6 Balls-into-boxes Formulation

This reverse interpretation lends itself to an alternate description of the process, which is perhaps more familiar to combinatorial probabilists – this is the description
mentioned in the abstract. We start with $b_0$ balls (individuals from the current generation), and throw them into $n$ boxes independently of each other, each ball landing in box $j$ with probability $p_j$. (A ball going from box $i$ to box $j$ corresponds to individual $j$ giving birth to individual $i$.) Any balls that land into the same box are fused into one (merging of ancestral lines), and we repeat this allocation according to the same rules with this new, possibly smaller, number of balls. The process terminates when all balls have fused into one ball (all $b_0$ ancestral lines have merged), which is the coalescence time $T_{b_0}$, which of course is random. We will investigate mainly how certain properties of $p$ influence the expected coalescence time $E[T_{b_0}]$.

1.7 Applications to Computer Science

Another interpretation of this problem is in terms of the compositions of random functions. Choose random functions $\{f_s : [n] \to [n]\}_{s \in \mathbb{N}}$ independently, in such a way that for all $i \in [n]$, and for all $s \in \mathbb{N}$, $f_s(i) = j$ with probability $p_j$, independently for all $i$ and $s$. The total coalescence time $T_n$ is then the smallest value of $t$ for which $f_t \circ \cdots \circ f_2 \circ f_1$, restricted to $[b_0]$, is a constant function. It is this formulation that has been used in connection with computer science: this problem is potentially useful in bounding the running time of so-called ‘Coupling from the Past’ (CFTP) algorithms introduced by Propp and Wilson [18, 19], which we describe here.

The general (discrete) random sampling problem involves generating a sample that is distributed according to some given probability vector $q = (q_1, \ldots, q_n)$. In typical “Monte Carlo” random-sampling algorithms, one might construct a (ergodic) Markov
Chain with $n$ states which has $q$ as a stationary distribution, and then run the chain for a predetermined “large” amount of time and output the current state of the chain as our sample. However, the sample returned will not in general be distributed exactly according to the desired distribution. Thus much work has been done (using, for example, results about conductance and mixing times of ergodic Markov Chains) in trying to determine the relationship between the accuracy of the sample and the amount of time the chain must be run.

The CFTP algorithms, on the other hand, have the striking property that samples generated are distributed exactly according to the desired distribution $q$. The algorithms runs by performing multiple simultaneous flows on a Markov Chain that has $q$ as a stationary distribution. Such algorithms terminate once these flows all coalesce, and therefore bounds on the running time of such algorithms can be achieved through bounds on the coalescence time of the underlying chain. We should note that the current setup is of no immediate practical interest for this purpose, as the stationary distribution of our chain is $p$, but to run the chain we need to be able to sample from $p$ in the first place! In the next section we provide the framework for a possible future line of research that avoids this pitfall, which we discuss in more detail in Chapter 6.

### 1.8 General Markov Case

To be applicable for this purpose, we would need to consider the case of a chain evolving according to stochastic matrix $P = (p_{ij})_{1 \leq i, j \leq n}$ where $p_{ij}$ denotes the probability of a ball from box $i$ landing in box $j$; we will refer to this as the General Markov
Case. (Our multinomial model described above is the special case where each row of $P$ is the same vector $p$.) This general case does not satisfy the memorylessness property described above in Section 1.2. We would then run simultaneous flows on the chain and determine how long it takes for them to completely coalesce. Biologically, this represents the case where the reproductive fitness of a child is a function of its parent's; this leads to the evolutionarily crucial phenomenon of natural selection. Realistically, we would deal with a transition matrix $P$ which was quite sparse to allow for efficient running of the Markov process.

1.9 Outline of the Dissertation

The rest of the current work will be structured as follows. In Chapter 2, we will review some of the existing literature and precisely state our main theorem and some auxiliary theorems. In Chapter 3, we will present some key technical lemmas that we will need to prove our main theorem, which we will do in Chapter 4. In Chapter 5, we will show that the conclusions of our main theorem need not hold if $p$ does not satisfy certain uniformity requirements. Finally, in Chapter 6 we discuss possible future lines of research and some open questions.
CHAPTER 2
PREVIOUS WORK, STATEMENT OF THEOREMS

2.1 Kingman’s coalescent

There is a wealth of literature on this problem, going back to Kingman’s work in the early 1980s [7, 8, 9]. His main contribution was to show that for fixed $b_0$ (i.e. not tending to $\infty$ with $n$), for constant population size $n$, and for exchangeable offspring sizes $N(t) = (\nu_1(t), \ldots, \nu_n(t))$ with time-independent distributions satisfying certain moment conditions, the process converged in the limit to a continuous-time process now called Kingman’s coalescent. What follows is a discussion of his work as it relates to our problem.

2.1.1 Discrete-time process

We begin by restating the MRCA problem in terms of a partition-valued discrete stochastic process.

The following results hold for processes where $n$ is constant, and the $\nu_j$ are exchangeable (and hence identically distributed) random variables which satisfy certain uniformity conditions; the most important of which is that $Var[\nu_1]$ tends to a finite nonzero limit $\sigma^2$ as $n \to \infty$. 

10
Take a fixed set \( B_0 \) of \( b_0 \) individuals (\( b_0 \leq n \)) from the current generation, and without loss of generality label them \( \{1, 2, \ldots, b_0\} \). For each \( t, t = 0, 1, \ldots \) define a partition \( R_t \) of \( B_0 \) by the following rule: \( i \) and \( j \) \( (i, j \in B_0) \) belong to the same block of \( R_t \) iff individuals \( i \) and \( j \) have a common ancestor \( t \) generations backwards in time. Notice that \( R_t \) is necessarily a refinement of \( R_{t+1} \), for any \( t \). Thus

\[
R_0 = \Delta_{B_0} := \{ \{i\} : i \in B_0 \},
\]

and then the coalescence time \( T_{b_0} \) is therefore the smallest time \( t \) for which \( R_t \) consists of a single block, i.e.

\[
T_{b_0} = \min_{t \geq 0} \{ R_t = \{B_0\} \}.
\]

Define the transition probabilities

\[
p_{\xi\eta} = P(R_{t+1} = \eta | R_t = \xi), \tag{2.1}
\]

which must be 0 unless the partition \( \xi \) is a refinement of \( \eta \). Define a partial order \( \prec \) on the set of partitions of \( B_0 \) by: \( \xi \prec \eta \) iff \( \eta \) is obtained from \( \xi \) by merging exactly two of its blocks. (In particular, if \( \xi \prec \eta \), then \( \xi \) is a refinement of \( \eta \), but not conversely.) Also, define

\[
B(t) := |R_t| \tag{2.2}
\]

to be the number of blocks of \( R_t \); that is, the number of individuals at time \( t \) which are ancestors of at least one of the \( b_0 \) individuals of the sample at generation 0. In the balls-into boxes formulation, \( B(t) \) is the number of balls left at time \( t \).

The key feature of the limiting process is that simultaneous multiple mergers of ancestral lines are rare; that is, in the limit, as \( n \to \infty \) but \( b_0 \) remains fixed, \( p_{\xi\eta} \) is “negligibly small” unless \( \xi = \eta \) or \( \xi \prec \eta \).
To quantify this notion of “rare simultaneous mergers”, letting \( n \to \infty \), Kingman showed that
\[
p_{\xi \eta} = \delta_{\xi \eta} + q_{\xi \eta} n^{-1} \sigma^2 + O(n^{-2})
\]
where \( \delta \) is the Kronecker delta and
\[
q_{\xi \eta} = \begin{cases} 
-(k^2) & \text{if } \xi = \eta \text{ and } k = |\xi| \\
1 & \text{if } \xi \prec \eta \\
0 & \text{otherwise.}
\end{cases}
\]

2.1.2 Continuous-time process

We now define Kingman’s \( b \)-coalescent as a continuous-time partition-valued stochastic process \( \{Q_t\}_{t \geq 0} \) with the following properties:

\( Q_t \) is a partition of \([b]\) for all \( t \geq 0 \)
\[
Q_0 = \Delta_{[b]},
\]
\[
\lim_{h \downarrow 0} \frac{1}{h} P(Q_{t+h} = \eta|Q_t = \xi) = r_{\xi \eta},
\]
where the transition rates \( r_{\xi \eta} \) are defined by
\[
r_{\xi \eta} = \begin{cases} 
1 & \text{if } \xi \prec \eta \\
0 & \text{otherwise.}
\end{cases}
\]

Note that the probability of a merger of blocks only depends on the number of blocks present, and not on their sizes. Thus the process \( \{Q_t\} \) only allows for mergers of at most two blocks at any one time. If we define \( B_t = |Q_t| \), then \( B_t \) is an integer-valued continuous-time stochastic process on \([b]\) with initial value \( B_0 = b \), and with
probability 1 takes on every value in $[b]$ (because by definition of $r_{\xi_0}$, all reductions in the number of blocks of $Q_t$ are by at most 1). Thus $B_t$ is what is known as a pure death process. Now if we define the transition moments $\tau(k)$ as

$$\tau(k) = \inf_t \{B_t = k\},$$

then the process $\{\tau(k)\}_{k=b,b-1,\ldots,2,1}$ can be thought of as a Poisson process, where the inter-arrival waiting times are independent and exponentially distributed, but with different parameters: the waiting times $T(k) := \tau(k-1) - \tau(k)$ are exponentially distributed with parameter $\left(\frac{k}{2}\right)$. Therefore, the coalescence time $\tau(1)$ of the process $\{Q_t\}_{t\geq0}, Q_0 = [b]$ is distributed as

$$\sum_{k=2}^{b} Y_k,$$

where the $Y_k$ are independent and Exp$(\frac{k}{2})$ distributed.

Note. There is now a wealth of literature on so-called $\Lambda$- and $\Xi$- coalescents, which are generalizations of Kingman’s coalescent which allow for simultaneous and/or multiple mergers. We will not attempt a full description here, but we refer the reader to the seminal work of Pitman [16], Sagitov [20], and Schweinsberg [21].

### 2.1.3 Limiting Behavior

Based on the definitions of these processes $\{R_t\}_{t=0,1,\ldots}$ and $\{Q_t\}_{t\geq0}$ and an inspection of their transition rates (2.3) and (2.6), it should be expected that the the continuous-time process $Q_t$ is represents the limiting process (in some well-defined sense) for the discrete-time process $R_t$, under scaling by a factor of $n\sigma^{-2}$. This is indeed the case, although we will not go into the details of this notion of convergence of stochastic
processes. For our purposes of analyzing the coalescence time, a consequence is that, as \( n \to \infty \) and \( b_0 \) remains fixed,

\[
T_{b_0}/(n\sigma^{-2}) \Rightarrow \sum_{k=2}^{b_0} Y_k,
\]

where \( \Rightarrow \) denotes convergence in distribution, and where \( Y_k \) are as in (2.7). This result was later reproved in a different manner by Fill [4] and Goh et al. [6], who gave a detailed description of the limiting density function. It follows immediately from (2.8) that

\[
E[T_{b_0}]/(\sigma^{-2}n) \to \sum_{k=2}^{b_0} \frac{2}{k(k-1)} = \sum_{k=2}^{b_0} 2 \left( \frac{1}{k-1} - \frac{1}{k} \right) = 2(1 - 1/b_0).
\]

(2.9)

So far, these results only apply to the case where \( b_0 \) does not grow with \( n \). To see what (2.9) says about \( E[T_n] \), the following simple coupling result will be helpful. The statement is perhaps obvious, but it is crucial enough to warrant a proof.

**Proposition 2.1.1.** Let \( p \) be an arbitrary probability vector. Let \( m_1 \leq m_2 \leq n \). Let \( T_{m_1} \) and \( T_{m_2} \) be the coalescence time of processes \( \{B_{m_1}(t)\}, \{B_{m_2}(t)\} \) respectively, with \( B_{m_1}(0) = m_1 \) and \( B_{m_2}(0) = m_2 \) (on any probability space). Then \( T_{m_2} \) stochastically dominates \( T_{m_1} \), and so in particular \( E[T_{m_2}] \geq E[T_{m_1}] \).

**Proof.** Start with \( m_2 \) balls, of which \( m_1 \) are colored red. Perform the usual allocation-coalescence process according to \( p \), where a fused ball is considered red if it results from the fusion of at least one red ball. Then \( X \), the time at which there is only one red ball left is distributed as \( T_{m_1} \), and \( Y \), the time at which there is only one ball left at all, is distributed as \( T_{m_2} \). If there is only one ball left there is necessarily only one red ball left, so \( X \leq Y \) and the result follows. \( \square \)
Using this proposition on (2.9), we see that for \( B(0) = n \), for any fixed \( b_0 \leq n \),
\[
E[T_n] \geq E[T_{b_0}] \geq 2n\sigma^{-2}(1 - 1/b_0)(1 - o(1)),
\]
and so letting \( b_0 \) be arbitrarily large,
\[
E[T_n] \geq 2n\sigma^{-2}(1 - o(1)) \quad \text{as} \quad n \to \infty.
\]
(2.10)

Kingman proved as well (by different means involving clever manipulations of Stirling numbers) that for \( B(0) = n \), for the specific uniform multinomial case (where \( p = u = (1/n, \ldots, 1/n) \), which implies \( \sigma^2 = 1 \)) we have
\[
E[T_n(u)] \leq 2n - 2,
\]
(2.11)
and so combining this with (2.10) gives
\[
E[T_n(u)] = 2n(1 + o(1))
\]
(2.12)
for the uniform model. (We use the notation \( T_m(p) \) to denote the coalescence time of a sample of size \( m \) for the multinomial model evolving according to \( p \), when we wish to emphasize the dependence on \( p \)). This result was actually reproved 20 years later in a very different manner by Dalal and Schmutz [2].

2.2 Previous Results for the Nonuniform Case

2.2.1 Möhle

Several authors, notably Martin Möhle, have extended this “convergence to the coalescent” work a great deal, with some results for nonuniform cases, with most of these for \( B(0) \) fixed. First, for any probability vector \( p \), define
\[
c_2 = c_2(p) := \sum_{j=1}^{n} p_j^2 \quad \text{and} \quad c_3 = c_3(p) := \sum_{j=1}^{n} p_j^3,
\]
In [12], M"ohle proved convergence to the Kingman coalescent for a wider class of models, if time is scaled by the appropriate factor. Applied to the nonuniform multinomial case, a consequence of his result is that

$$E[T_{b_0}(p)] \rightarrow 2c_2^{-1}(1 - 1/b_0)$$

under the rather restrictive condition

$$\max_j p_j \leq K/n \quad \text{for some absolute constant } K. \quad (2.13)$$

To consider another result of M"ohle, let us define $\pi_{kb} := P(B(t + 1) = b | B(t) = k)$, recalling the definition (2.2) of $B(t)$.

In [13, Theorem 4.1], he shows that for $B(0) = n$, if we can show that

$$\sum_{b=1}^{k} \pi_{kb}/b \geq 1/k + c_2/2, \quad (2.14)$$

then the result

$$E[T_n(p)] = 2c_2^{-1}(1 + o(1))$$

would follow. However, neither he nor we have been able to verify (2.14) for the nonuniform multinomial case at hand.

### 2.2.2 Adler et al.

The paper with results most closely related to our own is by Adler et al. [1].
Their Main Result

Theorem 3 in [1] provides a result for the multinomial case where $p$ is not uniform, and $b_0$ is on the order of $n$. In our notation, it can be written as: for $B(0) = n$, we have

$$E[T_n(p)] = 2c_2^{-1}(1 + o(1)) \quad \text{as} \quad n \to \infty,$$

under the condition $c_3/c_2 < 3/n$.

Reassuringly, this agrees with Kingman’s result (2.11), as $c_2 = 1/n$ and $c_3/c_2 = 1/n < 3/n$ for the uniform case.

Schur Concavity

There is another result in [1] which bears mentioning. First, define a partial order (the dominance order) $\geq_M$ on the set of probability $n$-vectors by

$$p \geq_M q \quad \text{iff} \quad \sum_{j=1}^i p_{[j]} \geq \sum_{j=1}^i q_{[j]} \quad \forall i \in [n]$$

where $(p[,1],\ldots,p[n])$ and and $(q[,1],\ldots,q[n])$ are the permutations of $p$ and $q$ respectively with components written in nonincreasing order. Note that

$$e := (1,0,\ldots,0) \geq_M p \geq_M (1/n,\ldots,1/n) =: u$$

for any probability vector $p$.

Theorem 7 in [1] shows that $g(p) := P(T_n(p) > t)$ is a Schur concave function of $p$; that is,

$$P(T_n(p) > t) \leq P(T_n(q) > t) \quad (2.15)$$

17
whenever \( p \geq_M q \). In other words, \( T_n(p) \) is stochastically dominated by \( T_n(q) \) whenever \( p \geq_M q \). In particular, \( p \geq_M q \Rightarrow E[T_n(p)] \leq E[T_n(q)] \).

Since \( u \leq_M p \) for any \( p \), this result says that the \( u \)-process is slowest, which is not surprising — in some sense, the \( u \)-process produces the least amount of “clustering”.

Since \( e \geq_M p \) for any \( p \), this theorem is also consistent with the obvious fact that the \((1, 0, \ldots, 0)\)-process is the fastest (of course, \( T_n(e) = 1 \) with probability 1 in this case).

This result is however of limited use, as the partial order \( \leq_M \) is rather sparse. For example, if \( p \) and \( q \) are such that \( \sum_j p_j^2 = \sum_j q_j^2 \), then \( p \) and \( q \) cannot be comparable unless \( p = q \) (since the function \( c_2(p) = \sum_j p_j^2 \) is strictly Schur-concave). Also, while it may allow us to (stochastically) compare the coalescence times for a few pairs of processes, it says nothing directly about how these times may be distributed.

**Conditions on \( p \)**

We shall briefly discuss why we might expect the quantities \( c_3/c_2 \) and \( c_2 \) to show up in these results. \( c_3 \) is the probability that three balls collide at one step of the allocation. The presence of an upper bound on \( c_3/c_2 \) ensures that triple collisions are sufficiently rare compared to double collisions, and thus that there should be some time threshold, after which is a phase where all collisions (if any) are only one-on-one, and thus no states are skipped during that phase. (These conditions are analogous to those that guarantee (2.3) in Kingman’s setup.) Since we expect this “late” phase to contribute overwhelmingly to the total coalescence time, \( c_2 \) should be the central quantity controlling \( E[T_n] \). Indeed, in most of the models considered by Möhle in
c_2 (or an analogue) is used as the natural scaling factor. The fact that \( E[T_n] \) is a decreasing function of \( c_2 \) is to be expected — given (2.15), we would expect that the closer \( p \) is to being the uniform \( u \) (where \( c_2 \) achieves its minimum \( 1/n \)), the larger \( E[T_n] \) will be. We may thus think of \( c_2(p) \) as being a crude one-dimensional measure of the “nonuniformity” of \( p \).

### 2.3 Our Main Theorem

**Note.** From now on, unless stated otherwise, \( T \) will denote \( T_u(p) \), i.e., the coalescence time for the entire population for the multinomial process with probability vector \( p \). We will prove the same asymptotic result as in [1], under conditions on \( p \) which are much less restrictive than \( c_3/c_2 < 3/n \):

**Theorem 2.3.1.** Let \( B(0) = n \). Suppose that for some \( \varepsilon \in (0, 1/4) \), the following conditions on \( p \) are met:

\[
c_2 \ll \ln^{-2} n \quad \text{and} \quad c_3 \leq c_2^{3/2} \ln^{-1/2-\varepsilon} n. \tag{2.16}
\]

Then

\[
E[T] = 2c_2^{-1}(1 + o(1)) \quad \text{for} \quad n \to \infty. \tag{2.17}
\]

This first-moment result is the key to proving the following extension of (2.8):

**Theorem 2.3.2.** Under the conditions (2.16),

\[
\frac{T}{c_2^{-1}} \Rightarrow \sum_{k \geq 2} \frac{2}{k(k-1)} X_k, \tag{2.18}
\]

where the \( X_k \) are independent and exponentially distributed, \( P(X_k > x) = e^{-x} \).
Notes.

1. If not for the logarithmic factors, the conditions (2.16) would not have disallowed any \( p \), since it can be easily verified that for any \( p \), \( c_2 \leq 1 \) and \( c_3 \leq c_2^{3/2} \).

2. Our condition \( c_3/c_2 \leq c_2^{1/2} \ln^{-1/2-\varepsilon} n \) is far less restrictive than \( c_3/c_2 < 3/n \) from [1], as \( c_2^{1/2} \ln^{-2-\varepsilon} n \geq n^{-1/2} \ln^{-2-\varepsilon} n \gg 3/n \). It is also much less restrictive than Möhle’s condition (2.13), as our conditions allow some components of \( p \) to be \( \gg 1/n \).

2.4 A Process with a Long Early Phase

We will also show that it is not always the case that the early phase contribute a negligible amount to \( E[T] \). Interestingly, \( \ln^{-2} n \) appearing as the upper bound for \( c_2 \) in (2.16) is a genuine threshold for the property “\( E[T] \) is of order \( c_2^{-1} \) exactly”. We will provide the proof of the following theorem in Chapter 5.

Note. We say that an event \( A \) holds “with high probability” if \( P(A) \to 1 \) as \( n \to \infty \), and will often abbreviate it \( \text{whp} \).

Theorem 2.4.1. Let \( B(0) = n \). For \( c_2 \gg \ln^{-2} n \), there exists a probability vector \( p \) with \( \sum_j p_j^2 = c_2 \) such that with high probability, \( T \gg c_2^{-1} \), and thus \( E[T] \gg c_2^{-1} \).

Adler et al. [1, Theorem 5] had proved that \( E[T] \) could be \( \gg c_2^{-1} \), under conditions however which required \( c_2 \) to be bounded away from 0. As we allow the possibility that \( c_2 \to 0 \), our Theorem 2.4.1 can be viewed as a sharpening of their result.
2.5 Outline of the Proof of Theorem 2.3.1

We will provide a full proof of Theorem 2.3.1 in Chapter 4, but here is a brief description of our approach.

There is a fundamental qualitative difference between two phases of this process: the "early phase" (large values of \( B(t) \)) and the "late phase" (small values of \( B(t) \)). (The threshold between "early" and "late" will be made more precise later). We may expect that the late phase takes up the majority of the total coalescence time, as mergers are rare during this phase, and thus inter-merger times are long; our analysis will confirm that this is usually the case. This late phase is much simpler to analyze: there are long time intervals without collisions, and when collisions do happen they are likely to be only binary, which allows scaling to a combinatorially simple continuous-time process in the style of Kingman. For the early phase, there will be many multiple mergers at every round, and this does not permit a similar scaling. Moreover, the collisions during this phase could be of a wide variety of types, so that it is not at all clear what parameter of \( p \) is relevant (as opposed to the late phase where \( c_2 \) is clearly central).

For the early phase approach, observe that the sequence \( \{B(t)\}_{t \geq 0} \) is a Markov chain: the number of balls remaining at time \( t + 1 \), \( B(t + 1) \), depends only on \( B(t) \), the number of balls present at time \( t \). (Note that this would not be the case if we were in the General Markov Case, described in Section 1.8, where the destination of a ball depended on its current location.)
We want to get an idea of the expected behavior of the stochastic sequence \{B(t)\}_{t \geq 0}.
To that end, we have the following simple result concerning its one-step behavior.

**Proposition 2.5.1.** The sequence \{B(t)\}_{t \geq 0} has the following property:

\[ E[B(t + 1) | B(t) = k] = \sum_{j=1}^{n} [1 - (1 - p_j)^k], \quad B(0) = n. \]

**Proof.** Conditioned on the event \{B(t) = k\}, for 1 \leq j \leq n, define \(A_j\) to be the event that box \(j\) receives at least one ball at time \(t + 1\). Then \(P(A_j^c) = (1 - p_j)^k\), and so using linearity of expectations,

\[ E[B(t + 1) | B(t) = k] = \sum_{j} E[I_{A_j}] = \sum_{j} P(A_j) = \sum_{j} [1 - (1 - p_j)^k]. \]

\(\square\)

It would seem natural to try to prove that the conditional distribution of \(B(t + 1)\) is concentrated around \(E[B(t + 1) | B(t) = k]\), as long as \(k\) is large enough. Curiously, we will be able to show instead that for smallish \(t\), \textbf{whp} \(B(t + 1)\) is relatively close to \(\Phi_p(B(t))\), where

\[ \Phi_p(k) := \sum_{j=1}^{n} (1 - e^{-p_j})^k. \quad (2.19) \]

Note that \(\Phi_p(B(t))\) is close to \(E[B(t + 1) | B(t)]\) when most of \(p_jB(t)\) are small. We will show that with high probability, the sequence “closely” follows this expected recurrence for all values of \(B(t)\) greater than some threshold \(k^*\) to be determined (this will define the early phase). We will then use this recurrence to bound the expected time spent in states \(\geq k^*\). We should note that not only will this allow us to bound the expected time spent in the early phase, but it will provide sharp concentration
bounds for $B(t)$ for small times $t$. Therefore, if we are interested not in the time
to the MRCA of the whole population, but rather, say, the likely number of distinct
ancestral lineages in the very recent past, then this approach will allow us to estimate
that as well.
We will use this chapter to get out of the way some of the more technical parts of the proof of the main theorem, so as to make the argument in the next chapter easier to follow.

3.1 Variational Problems

In order to make use of the recurrence

\[ E[B(t + 1)|B(t)] \approx \Phi_p(B(t)), \]  

explained at the end of the previous chapter, we will need estimates for the unwieldy \( \sum_j e^{-p_j k} \), given the definition of \( \Phi_p(k) \) in (2.19). We would like to obtain estimates which retain dependence on \( p \) but without having to keep track of all \( n \) of its components. One obvious way would be to use Taylor expansions, such as

\[
\sum_j e^{-p_j k} \geq \sum_j \left( 1 - p_j k + \frac{p_j^2 k^2}{2} - \frac{p_j^3 k^3}{6} \right)
\]

\[
= n - k + \frac{k^2}{2} \sum_j p_j^2 - \frac{k^3}{6} \sum_j p_j^3
\]

\[
= n - k + \frac{k^2}{2} c_2 - \frac{k^3}{6} c_3,
\]
which now depends on $\mathbf{p}$ only through two of its statistics, namely $c_2$ and $c_3$. This approach yielded some modest results, however it is not ideal in that the validity of the Taylor approximation relies on each of the $p_jk$ being small enough, whereas we would like to be able to deal with large $k$ (for the early phase) and potentially large $p_j$ (for dealing with vectors that are far from uniform). We shall therefore use the following more intricate variational approach to bounding $\Phi_p(B(t))$.

Let us define $D_n$ to be the set of probability $n$-vectors. For any $\mathbf{q} \in D_n$, define

$$F_{\mathbf{q}}(k) := \sum_{j=1}^{n} e^{-kq_j}, \quad (3.2)$$

so that $\Phi_{\mathbf{q}}(k) = n - F_{\mathbf{q}}(k)$, and set

$$D(c_2) := \left\{ \mathbf{q} \in D_n \mid \sum_j q_j^2 = c_2(\mathbf{p}) \right\},$$

$$D(c_2, c_3) := \left\{ \mathbf{q} \in D_n \mid \sum_j q_j^2 = c_2(\mathbf{p}), \sum_j q_j^3 = c_3(\mathbf{p}) \right\}.$$ 

That is, $D(c_2)$ (resp. $D(c_2, c_3)$) is the set of probability vectors that share the same sum of squares (resp., squares and cubes) as $\mathbf{p}$. The functional $F_{\mathbf{q}}(k)$ is continuous and the sets $D_n$, $D(c_2)$ and $D(c_2, c_3)$ are compact, so the infima of $F$ (as a function of $\mathbf{q}$, for fixed $k$) over these sets are attained. Note that $D(c_2)$ and $D(c_2, c_3)$ are never empty because each contains $\mathbf{p}$ at the very least. (If $\mathbf{p} = \mathbf{u}$, then these sets actually contain nothing else.)

Note that the order in which the components of any probability vector are listed in no way affects anything about the process; we will therefore use whatever ordering is most convenient.
Proposition 3.1.1. For any \( k \in [n] \),

\[
\min_{q \in D_n} F_q(k) = F_u(k),
\]

where \( u \) is the uniform vector \( u = (1/n, \ldots, 1/n) \).

This result should not be surprising, given our intuition that the uniform process decays ‘slowest’, and the result (2.15); this crude estimate will be most useful when \( p \) is close to uniform.

Proposition 3.1.2. For any \( k \in [n] \),

\[
\min_{q \in D(c_2)} F_q(k) = F_{\theta(c_2)}(k),
\]

where \( \theta(c_2) = (\theta_1, \theta_2, \ldots, \theta_n) \) (when its components are listed in nonincreasing order) has the property that \( \theta_1 \geq \theta_2 = \cdots = \theta_n \). That is, \( \theta \) has at most two distinct entries, and the larger one has support size equal to 1. We will refer to such vectors as being of “top heavy” type.

Proposition 3.1.3. For any \( k \in [n] \),

\[
\min_{q \in D(c_2, c_3)} F_q(k) = F_{r(c_2, c_3)}(k),
\]

where \( r(c_2, c_3) := (r_1, r_2, \ldots, r_n) \) (when ordered in nonincreasing order) has the following property: for some \( \nu \in \{1, 2, \ldots, n\} \),

\[
r_1 = \cdots = r_\nu \geq r_{\nu+1} \geq r_{\nu+2} = \cdots = r_n.
\]

(3.3)

That is, \( r(c_2, c_3) \) has at most 3 distinct entries, and the middle one (if any) has support size equal to 1.
Clearly, \( D_n \supset D(c_2) \supset D(c_2, c_3) \), and so \( F_p(k) \geq F_r(k) \geq F_\theta(k) \geq F_u(k) \). These propositions therefore provide sharper and sharper estimates, so at various junctures we will use whichever one is easiest to work with, while still being sharp enough.

**Proof of Proposition 3.1.1.** Using the fact that \( e^{-x} \) is concave up,

\[
F_p(k) = \sum_j e^{-p_jk} \\
= n \sum_j \frac{1}{n} e^{-p_jk} \\
\geq n \exp \left( - \sum_j \frac{1}{n} p_jk \right) \\
= ne^{-k/n} = F_u(k).
\]

\( \square \)

**Proofs of Propositions 3.1.2 and 3.1.3.** To prove Proposition 3.1.2, there are two steps:

1. Show that a minimizer \( \theta \) of \( F_q(k) \) on \( D(c_2) \) cannot have a configuration \( \theta_{j_1} > \theta_{j_2} > \theta_{j_3} \) for \( j_1, j_2, j_3 \in [n] \); that is, that it cannot have three distinct entries.

2. Show that a minimizer \( \theta \) cannot have a \( \theta_{j_1} = \theta_{j_2} > \theta_{j_3} \) configuration, which will imply that the larger entry is unique.

We will only prove Proposition 3.1.3 as it is more difficult; the interested reader would not find it difficult to adapt the argument to prove Proposition 3.1.2.

We may recast Proposition 3.1.3 as follows: letting \( z_j := q_jk \), \( z := (z_1, \ldots, z_n) \), we want to minimize \( G(z) := \sum_j e^{-z_j} \) under the constraints

\[
\sum_j z_j = k, \quad \sum_j z_j^2 = k^2 c_2, \quad \text{and} \quad \sum_j z_j^3 = k^3 c_3. \tag{3.4}
\]
Our task is to show that the minimizer $x = (x_1, \ldots, x_n)$ of $G(z)$, with components listed in nonincreasing order, must have the form $x_1 = \cdots = x_\nu \geq x_{\nu+1} \geq x_{\nu+2} = \cdots = x_n$ for some $\nu \in [n]$.

**Case I.** We first show that a minimizer of $G(z)$ over $D(c_2, c_3)$ cannot have four distinct entries. Suppose for the sake of contradiction that we have a minimizing vector $x = (x_1, \ldots, x_n)$ for which there exist $j_1, j_2, j_3, j_4$ such that $x_{j_1} > x_{j_2} > x_{j_3} > x_{j_4} \geq 0$ (without loss of generality, let $(j_1, j_2, j_3, j_4) = (1, 2, 3, 4)$). Let $y_j = x_j + \varepsilon_j$ for $j = 1, \ldots, 4$, and $y_j = x_j$ for $j = 5, \ldots, n$; we will show that for a suitable choice of $\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)$, $y = (y_1, \ldots, y_n)$ satisfies the conditions (3.4), but $G(y) < G(x)$, and thus such an $x$ cannot be a minimizer.

First note that we require $\varepsilon_4 \geq 0$ because of the possibility that $x_4 = 0$, but $\varepsilon_1, \varepsilon_2, \varepsilon_3$ can be of either sign. For $y$ to satisfy the conditions (3.4), we require

$$\sum_{j=1}^{4} \varepsilon_j = 0, \quad (3.5)$$

$$2 \sum_{j=1}^{4} x_j \varepsilon_j + \sum_{j=1}^{4} \varepsilon_j^2 = 0, \quad (3.6)$$

$$3 \sum_{j=1}^{4} x_j^2 \varepsilon_j + 3 \sum_{j=1}^{4} x_j \varepsilon_j^2 + \sum_{j=1}^{4} \varepsilon_j^3 = 0. \quad (3.7)$$

Now we want $G(x) - G(y) > 0$, where

$$G(x) - G(y) = \sum_{j=1}^{4} e^{-x_j} \left(1 - e^{-\varepsilon_j}\right).$$

By linearizing the $e^{-\varepsilon_j}$ factors, it will be sufficient (by letting the $\varepsilon_j$ be as small as we wish) to show that

$$e^{-x_1} \varepsilon_1 + e^{-x_2} \varepsilon_2 + e^{-x_3} \varepsilon_3 + e^{-x_4} \varepsilon_4 > 0. \quad (3.8)$$
We now obtain expressions for the $\varepsilon_j$. For given $x_j$, the system (3.5)-(3.7) is a system of 3 nonlinear equations in 4 unknowns $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$; treating $\varepsilon_4$ as a parameter, we hope to be able to solve it uniquely for $\epsilon_1, \varepsilon_2, \varepsilon_3$ near $(0, 0, 0)^T$. Let $\varepsilon := (\varepsilon_1, \varepsilon_2, \varepsilon_3)^T$, and write (3.5)-(3.7) as the vector equation

$$f(\varepsilon) = b(\varepsilon_4), \quad b(\varepsilon_4) := (-\varepsilon_4, -2x_4^2 \varepsilon_4 - \varepsilon_4^2, -3x_4^2 \varepsilon_4 - 3x_4 \varepsilon_4^2 - \varepsilon_4^3)^T.$$ 

The derivative (Jacobian) matrix of $f$ at 0 is

$$L := \begin{pmatrix} 1 & 1 & 1 \\ 2x_1 & 2x_2 & 2x_3 \\ 3x_1^2 & 3x_2^2 & 3x_3^2 \end{pmatrix}.$$ 

Its determinant is equal to $6\Delta(x_1, x_2, x_3)$, where $\Delta(x_1, x_2, x_3)$ is the Vandermonde determinant for $x_1, x_2, x_3$,

$$\Delta(x_1, x_2, x_3) = (x_2 - x_1)(x_3 - x_1)(x_3 - x_2),$$

which is non-zero (negative), as the $x_i$ are distinct (decreasing). Therefore, by the Inverse Vector Function Theorem (IVFT), for $|\varepsilon_4|$ sufficiently small there exists a differentiable solution $\varepsilon = \varepsilon(\varepsilon_4), \varepsilon(0) = 0$, such that

$$\varepsilon = \gamma \varepsilon_4 + O(\varepsilon_4^2), \quad \gamma := L^{-1}b'(0) = L^{-1}(-1, -2x_4, -3x_4^2)^T.$$ 

Explicitly, by Cramer’s rule,

$$\gamma_1 = -\frac{\Delta(x_1, x_2, x_3)}{\Delta(x_1, x_2, x_3)}, \quad \gamma_2 = -\frac{\Delta(x_1, x_4, x_3)}{\Delta(x_1, x_2, x_3)}, \quad \gamma_3 = -\frac{\Delta(x_1, x_4, x_4)}{\Delta(x_1, x_2, x_3)}.$$

With these formulas, (3.8) is equivalent to showing (by letting $\varepsilon_4 > 0$ be as small as needed)

$$e^{-x_4} \Delta(x_1, x_2, x_3) < e^{-x_1} \Delta(x_4, x_2, x_3) + e^{-x_2} \Delta(x_1, x_4, x_3) + e^{-x_3} \Delta(x_1, x_2, x_4),$$

29
which in turn is equivalent to showing that

\[
D := \begin{vmatrix}
  e^{-x_1} & e^{-x_2} & e^{-x_3} & e^{-x_4} \\
  1 & 1 & 1 & 1 \\
  x_1 & x_2 & x_3 & x_4 \\
  x_1^2 & x_2^2 & x_3^2 & x_4^2 \\
\end{vmatrix} > 0.
\]

By performing a sequence of column operations and simple cofactor expansions typical of Vandermonde-type determinants, we get

\[
D = - \begin{vmatrix}
  e^{-x_2} - e^{-x_1} & e^{-x_3} - e^{-x_1} & e^{-x_4} - e^{-x_1} \\
  x_2 - x_1 & x_3 - x_1 & x_4 - x_1 \\
  x_2^2 - x_1^2 & x_3^2 - x_1^2 & x_4^2 - x_1^2 \\
\end{vmatrix} = (x_2 - x_1)(x_3 - x_1)(x_4 - x_1) \begin{vmatrix}
  e^{-x_3-x_1} & e^{-x_2-x_1} & e^{-x_4-x_1} \\
  x_3 - x_2 & x_4 - x_2 & x_2 - x_1 \\
\end{vmatrix}.
\]

Now define

\[
\lambda(x) := \frac{e^x - 1}{x}
\]

and

\[
C = C(x_1, x_2, x_3, x_4) := (x_2 - x_1)(x_3 - x_1)(x_4 - x_1)(x_3 - x_2)(x_4 - x_2)e^{-x_1} < 0.
\]

We then get, by factoring out terms judiciously,

\[
D = C \cdot \begin{vmatrix}
  \frac{\lambda(x_1 - x_2) - \lambda(x_1 - x_3)}{x_3 - x_2} & \frac{\lambda(x_1 - x_2) - \lambda(x_1 - x_4)}{x_4 - x_2} \\
  1 & 1 \\
\end{vmatrix} = C \cdot \left[ \frac{\lambda(x_1 - x_2) - \lambda(x_1 - x_3)}{(x_1 - x_2) - (x_1 - x_3)} - \frac{\lambda(x_1 - x_2) - \lambda(x_1 - x_4)}{(x_1 - x_2) - (x_1 - x_4)} \right].
\]

(3.9)
Now $\lambda(x)$ is concave up for $x > 0$, as can be seen by the fact that the Maclaurin expansion of $\lambda''(x)$ has only positive coefficients:

$$\lambda''(x) = \sum_{k=0}^{\infty} \frac{x^k}{(k + 3)k!}.$$  

Therefore, since

$$0 < x_1 - x_2 < x_1 - x_3 < x_1 - x_4,$$

the quantity in square brackets in (3.9) must be strictly negative, by interpreting its terms to be slopes of secant lines to the graph of $\lambda(x)$. Using this and the fact that $C < 0$, we get the desired conclusion, i.e. $D > 0$.

**Case II.** Now we show that a vector $\mathbf{x}$ with a configuration $x_1 > x_2 = x_3 > x_4 \geq 0$ cannot be a minimizer of $G$ either. Define $x := x_2 = x_3$. Now that $f'(0)$ is singular, determination of small feasible $\varepsilon_1, \ldots, \varepsilon_4$ such that $G(\mathbf{y}) < G(\mathbf{x})$ is more of a challenge. The fact that the linear terms in (3.5)-(3.7) now depend on only $\varepsilon_1, \varepsilon_2 + \varepsilon_3, \varepsilon_4$ hints that $|\varepsilon_1|, |\varepsilon_2 + \varepsilon_3|, |\varepsilon_4|$ should be equally small, and that $|\varepsilon_2|$ and $|\varepsilon_3|$, while small, should be much larger.

Believing in this scenario, we set

$$\varepsilon_1 = \delta_1 \varepsilon^2, \quad \varepsilon_2 = \varepsilon + \delta_2 \varepsilon^2, \quad \varepsilon_3 = -\varepsilon, \quad \varepsilon_4 = \delta_4 \varepsilon^2,$$

and seek the feasible $\delta_i(\varepsilon)$ for small $\varepsilon$. To begin with, we again require $\delta_4 \geq 0$. The conditions (3.5)-(3.7) become

$$\delta_1 + \delta_2 + \delta_4 = 0, \quad (3.10)$$

$$x_1 \delta_1 + x \delta_2 + x_4 \delta_4 = -1 + \varepsilon b_2(\varepsilon, \delta), \quad (3.11)$$

$$x_1^2 \delta_1 + x^2 \delta_2 + x_4^2 \delta_4 = -2x + \varepsilon b_3(\varepsilon, \delta), \quad (3.12)$$
where the \( b_i(\varepsilon, \delta) \) are polynomials. Notice that \( \Delta := \Delta(x_1, x, x_4) \), the determinant of the matrix in (3.10)-(3.12), is nonzero. So for \(|\varepsilon|\) small enough, there exists a differentiable solution \( \delta(\varepsilon) \), such that \( \delta(0) \) is the solution of (3.10)-(3.12) with 0, -1, and -2\(x\) respectively on the right hand side. By Cramer’s rule,

\[
\begin{align*}
\delta_1(0) &= \frac{1}{\Delta} \begin{vmatrix} 0 & 1 & 1 & 1 \\ -1 & x & x_4 & x_1^2 \end{vmatrix}, \\
\delta_2(0) &= \frac{1}{\Delta} \begin{vmatrix} 1 & 0 & 1 & 1 \\ x_1 & -1 & x_4 & x_1^2 \end{vmatrix}, \\
\delta_4(0) &= \frac{1}{\Delta} \begin{vmatrix} 1 & 1 & 0 & 1 \\ x_1 & x & -1 & x_4 \end{vmatrix},
\end{align*}
\]

which gives

\[
\begin{align*}
\delta_1(0) &= \frac{(x - x_4)^2}{\Delta}, \\
\delta_2(0) &= \frac{(x_1 - x_4)(x_1 + x_4 - 2x)}{\Delta}, \\
\delta_4(0) &= -\frac{(x - x_1)^2}{\Delta}.
\end{align*}
\]

Reassuringly, \( \delta_4(0) \) is positive (because \( \Delta \) is negative). Again we want \( G(x) - G(x + \varepsilon) > 0 \); we have

\[
G(x) - G(x + \varepsilon) = \sum_{j=1}^{4} (1 - e^{-\varepsilon_j})e^{-x_j}
\]

\[
e^{-x_1}(1 - e^{-\varepsilon_1}) + e^{-x}(1 - e^{-\varepsilon_2} + 1 - e^{-\varepsilon_3}) + e^{-x_4}(1 - e^{-\varepsilon_4})
\]

\[
= e^{-x_1}\left(\delta_1 \varepsilon^2 + O(\varepsilon^4)\right)
\]

\[
+ e^{-x}(\varepsilon + \delta_2 \varepsilon^2 - \frac{1}{2}(\varepsilon + \delta_2 \varepsilon^2)^2 + O(\varepsilon^3)) - \varepsilon - \frac{1}{2}\varepsilon^2 + O(\varepsilon^3)
\]

\[
+ e^{-x_4}(\delta_4 \varepsilon^2 + O(\varepsilon^4))
\]

\[
= \varepsilon^2(\delta_1 e^{-x_1} + (\delta_2 - 1)e^{-x} + \delta_4 e^{-x_4} + O(\varepsilon)).
\]

(3.13)

Thus, by taking \( \varepsilon \) sufficiently small, (3.13) will be > 0 if

\[
\delta_1(0)e^{-x_1} + (\delta_2(0) - 1)e^{-x} + e^{-x_4}\delta_4(0) > 0.
\]

(3.14)

In light of the formulas for the \( \delta_i(0) \) and the fact that \( \Delta < 0 \), (3.14) is equivalent to \( T(x) > 0 \), where

\[
T(x) := -(x - x_4)^2e^{-x_1} + (\Delta - (x_1 - x_4)(x_1 + x_4 - 2x))e^{-x} + (x - x_1)^2e^{-x_4}.
\]

(3.15)
Now multiplying $T(x)$ by $e^x$ and using the inequalities

$$e^{-(x_1-x)} < 1 - (x_1 - x) + (x_1 - x)^2/2, \quad e^{x-x_4} > 1 + (x - x_4) + (x - x_4)^2/2,$$

we get

$$e^x T(x) = (x - x_4)^2 (1 - e^{x-x_1}) + (x - x_1)^2 (e^{x-x_4} - 1) + \Delta$$

$$> (x - x_4)^2 \left( (x_1 - x) - \frac{(x_1 - x)^2}{2} \right) + (x - x_1)^2 \left( (x - x_4) + \frac{(x - x_4)^2}{2} \right) + \Delta$$

$$= (x - x_4)^2 (x_1 - x) + (x - x_1)^2 (x - x_4) + (x - x_1)(x_4 - x_1)(x_4 - x)$$

$$= 0. \quad (!)$$

Therefore (3.15) holds, and thus as before, $x$ cannot be a minimizer, and this concludes Case II.

This only leaves the possibility that the minimizer $x$ of $G$ is of the form

$$x_1 = x_2 = \cdots = x_\nu \geq x_{\nu+1} \geq x_{\nu+2} = \cdots = x_n,$$

and thus that the minimizer $r$ of $F_q(k)$ over $D(c_2, c_3)$ is of the form (3.3), for any $k$.

This concludes the proof of Proposition 3.1.3. \qed

### 3.2 Exponential Tail Bounds

Recall that conditioned on the value of $B(t)$, we expected $B(t+1)$ to be “likely” to be “near” its approximate expected value $\Phi(B(t))$. In this section we quantify this intuition.

The bounds we will develop are similar in spirit to Chernoff bounds, as they rely on a generating-function approach for bounding large deviations. Along the same lines as
Pittel [17], we will first express the unwieldy transition probabilities $\pi_{kb}$ by means of generating functions. We will then use these to establish two-sided exponential tail bounds for the distribution of $B(t + 1)$ conditioned on $B(t)$, which we will use in Theorems 2.3.1 and 2.4.1.

Recall that $\{B(t)\}_{t \geq 0}$ is a Markov chain, and so we may write the (time-independent) transition probabilities as

$$\pi_{kb} := P(B(t + 1) = b | B(t) = k),$$

and note that $\pi_{kb} = 0$ for $k < b$. We also introduce the tail probabilities

$$\pi^-_{kb} = P(B(t + 1) < b | B(t) = k) \quad \text{and} \quad \pi^+_{kb} = P(B(t + 1) > b | B(t) = k).$$

**Theorem 3.2.1.**

$$\pi^-_{kb} \leq 3\sqrt{k} \exp \left[ -\frac{(\Phi_p(k) - b)^2}{2k} \right], \quad b \leq \Phi_p(k) \tag{3.17}$$

and

$$\pi^+_{kb} \leq 3\sqrt{k} \exp \left[ -\frac{(b - \Phi_p(k))^2}{2k} \right], \quad b \geq \Phi_p(k). \tag{3.18}$$

Qualitatively, these theorems say that the tail probabilities are exponentially small in the deviation from their (approximate) expected value.

**Proof of Theorem 3.2.1.** The heart of the proof is an expression for $\pi_{kb}$ by means of generating functions.

### 3.2.1 Generating-Function expression

For the uniform case $p = u$, one can express $\pi_{kb}$ by means of Stirling numbers of the second kind, a fact which Kingman [9] and Fill [4] deftly exploited. However,
things are not as simple for arbitrary \( p \), which necessitates the following extension.

As usual, the expression \([y^m]f(y)\) denotes the coefficient of \( y^m \) in the power series expansion of \( f(y) \).

**Lemma 3.2.2.**

\[
\pi_{kb} = \frac{k!}{n^k} [x^k z^b] \left( \prod_{j=1}^n (1 + z(e^{np_j x} - 1)) \right), \quad 1 \leq b \leq k \leq n. \quad (3.19)
\]

**Proof.**

\[
\pi_{kb} = P(B(t + 1) = b|B(t) = k) = \sum_{U \subseteq [n], \ |U| = b} P(k \text{ balls go into exactly the boxes indexed by } U)
\]

\[
= \sum_U \sum_{\varepsilon_1 + \cdots + \varepsilon_b = k} \sum_{\varepsilon_j > 0 \ \forall j \in [b]} P(\text{each box } j \text{ from } U \text{ gets } \varepsilon_j \text{ balls})
\]

\[
= \sum_U \sum_{\varepsilon} \binom{k}{\varepsilon_1 \varepsilon_2 \cdots \varepsilon_b} \prod_{j \in U} p_j^{\varepsilon_j}
\]

\[
= k! \sum_U \sum_{\varepsilon} \prod_{j \in U} p_j^{\varepsilon_j} \varepsilon_j!.
\]

Now we build a *bivariate* generating function for the probabilities \( \pi_{kb} \), starting with the \( k \) index. Incorporating for future convenience an \( n^k \) factor, we have

\[
\sum_k \pi_{kb} \frac{n^k}{k!} x^k = \sum_U \sum_{\varepsilon_1 + \cdots + \varepsilon_b = k} \sum_{\varepsilon_j > 0} \prod_{j \in U} \frac{(nx)^{\varepsilon_j} p_j^{\varepsilon_j}}{\varepsilon_j!}.
\]

We merge the second and third sums, yielding for the right-hand side

\[
\sum_U \sum_{\varepsilon_j > 0} \prod_{j \in U} \frac{(nx)^{\varepsilon_j} p_j^{\varepsilon_j}}{\varepsilon_j!}.
\]
Reversing the order of summation and multiplication, we get

\[
\sum_k \pi_{kb} \frac{n^k}{k!} x^k = \sum_U \prod_{j \in U} \sum_{\varepsilon = 1}^{\infty} \frac{(n\varepsilon p_j)^{\varepsilon}}{\varepsilon!} = \sum_U \prod_{j \in U} (\exp(n\varepsilon p_j x) - 1).
\]

Multiplying by \( z^b \) and summing for \( b \geq 0 \), we obtain

\[
\sum_b \sum_k \pi_{kb} \frac{n^k}{k!} x^k z^b = \sum_b z^b \sum_{|U| = b, j \in U} (\exp(n\varepsilon p_j x) - 1)
\]

\[
= \sum_{U \subseteq [n]} z^b \prod_{j \in U} (\exp(n\varepsilon p_j x) - 1)
\]

\[
= \sum_{U \subseteq [n]} \prod_{j \in U} z(\exp(n\varepsilon p_j x) - 1)
\]

\[
= \prod_{j = 1}^n \left[ 1 + z(\exp(n\varepsilon p_j x) - 1) \right].
\]

Therefore we have

\[
\pi_{kb} \frac{n^k}{k!} = [x^k z^b] \prod_{j = 1}^n \left[ 1 + z(\exp(n\varepsilon p_j x) - 1) \right]
\]

and from here the lemma follows. \( \square \)

### 3.2.2 Probability Generating Functions

To make use of the above generating-function expressions for the transition probabilities, we will now look at the probability generating functions. By Lemma 3.2.2,

\[
g_k(z) := E[z^{B(t+1)}|B(t) = k] = \sum_{b=1}^k \pi_{kb} z^b = \frac{k!}{n^k} [x^k] \prod_{j=1}^n (1 + z(e^{n\varepsilon p_j x} - 1)).
\]
Now, for \( b \leq k \), we have

\[
\begin{align*}
\sum_{i=1}^{b-1} \pi_{ki} z^i & \leq g_k(z), \quad 0 < z \leq 1, \\
\sum_{i=1}^{b} \pi_{ki} z^i & \leq g_k(z), \quad 0 < z \leq 1.
\end{align*}
\]

This gives

\[
\begin{align*}
\pi_{kb}^- & \leq \frac{g_k(z)}{z^b} = \frac{k!}{z^b n^k} [z^k] \prod_{j=1}^{n} (1 + z(e^{np_j r} - 1)), \quad \text{for } 0 < z \leq 1, \quad (3.20) \\
\pi_{kb}^+ & \leq \frac{g_k(z)}{z^b} = \frac{k!}{z^b n^k} [z^k] \prod_{j=1}^{n} (1 + z(e^{np_j r} - 1)), \quad \text{for } z \geq 1. \quad (3.21)
\end{align*}
\]

Since the coefficients of the products in (3.20)-(3.21) are nonnegative, we may use the inequality \([x^k] f(x) \leq f(x)/x^k\ (\forall x > 0)\) to obtain (after replacing \( x \) by \( r \))

\[
\begin{align*}
\pi_{kb}^- & \leq \frac{k!}{z^b (nr)^k} \prod_{j=1}^{n} (1 + z(e^{np_j r} - 1)), \quad \forall r > 0, \quad 0 < z \leq 1, \quad (3.22) \\
\pi_{kb}^+ & \leq \frac{k!}{z^b (nr)^k} \prod_{j=1}^{n} (1 + z(e^{np_j r} - 1)), \quad \forall r > 0, \quad z \geq 1. \quad (3.23)
\end{align*}
\]

We use Stirling's formula \( k! \leq 3\sqrt{k}(k/e)^k \) to transform the product-type formulas (3.22) and (3.23) into

\[
\begin{align*}
\pi_{kb}^- & \leq 3\sqrt{k} \exp(H(z, r, b)), \quad z \leq 1, \quad (3.24) \\
\pi_{kb}^+ & \leq 3\sqrt{k} \exp(H(z, r, b)), \quad z \geq 1, \quad (3.25)
\end{align*}
\]

where

\[
H(z, r, b) := k \ln \left( \frac{k}{r n e} \right) - b \ln(z) + \sum_j \ln(1 + z(e^{np_j r} - 1)). \quad (3.26)
\]
3.2.3 Choosing $z$ and $r$

Our task is to get the most out of these bounds (3.24)-(3.25) by choosing values for $z$ and $r$ judiciously.

For a given $b$, we want to use a stationary point of $H(z, r, b)$, i.e. a solution to

$$H_z = -\frac{b}{z} + \sum_j \frac{e^{np_jr} - 1}{1 + z(e^{np_jr} - 1)} = 0, \quad (3.27)$$

$$H_r = -\frac{k}{r} + z \sum_j \frac{np_j e^{np_jr}}{1 + z(e^{np_jr} - 1)} = 0. \quad (3.28)$$

This complicated system has a simple solution $(z_*, r_*) = (1, k/n)$ for $b = b_* := n - \sum_j e^{-p_jk} (= \Phi_p(k))$; this algebraic fact justifies the appearance of $\Phi_p$ in the statement of the theorems, and agrees with our intuitive interpretation of $\Phi_p(B(t))$ as the approximate conditional expected value of $B(t + 1)$.

Moreover, from (3.26) it is immediate that $H(z_*, r_*, b_*) = 0$. This is a first sign that the inequalities (3.24)-(3.25) may indeed lead to meaningful explicit bounds for $\pi_{kb}^\pm$.

Of course, we need to know that (3.27)-(3.28) has a solution $(z, r)$ for $b \neq b_*$ as well, such that $z < 1$ for $b < b_*$, and $z > 1$ for $b > b_*$. This leads us to consider the following determinant

$$\chi := \det \begin{pmatrix} H_{zz} & H_{zr} \\ H_{rz} & H_{rr} \end{pmatrix} = H_{zz}H_{rr} - H_{zr}^2.$$

**Lemma 3.2.3.** $\chi > 0$ for all $(z, r, b)$ on the curve $C$ defined by (3.27) and (3.28). Thus for a given $b$, by the Implicit Vector Function Theorem, (3.27) and (3.28) define infinitely differentiable functions $z(b), r(b)$ uniquely, such that $(z(b_*), r(b_*)) = (1, k/n)$. 38
Proof. By the definition (3.26) of $H$, we have

$$H_{zz} = \frac{b}{z^2} - \sum_j \frac{(e^{np_j r} - 1)^2}{(1 + z(e^{np_j r} - 1))^2}, \quad (3.29)$$

$$H_{rr} = \frac{k}{r^2} + z(1 - z)n^2 \sum_j \frac{p_j^2 e^{-np_j r}}{((1 - z)e^{-np_j r} + z)^2}, \quad (3.30)$$

and

$$H_{rz} = \frac{n}{z^2} \sum_j \frac{p_j e^{np_j r}}{(z^{-1} + e^{np_j r} - 1)^2}. \quad (3.31)$$

We can recast (3.27) and (3.28) as

$$b = \sum_j \frac{z(e^{np_j r} - 1)}{1 + z(e^{np_j r} - 1)}, \quad k = nrz \sum_j \frac{p_j e^{np_j r}}{1 + z(e^{np_j r} - 1)}. \quad (3.32)$$

Using these in (3.29)-(3.31), we get that on $C$,

$$H_{zz} = \frac{1}{z^2} \sum_j \frac{z(e^{np_j r} - 1)}{1 + z(e^{np_j r} - 1)} - \sum_j \frac{(e^{np_j r} - 1)^2}{(1 + z(e^{np_j r} - 1))^2}, \quad (3.33)$$

$$H_{rr} = \frac{nz}{r} \sum_j \frac{p_j e^{np_j r}}{1 + z(e^{np_j r} - 1)} + z(1 - z)n^2 \sum_j \frac{p_j^2 e^{np_j r}}{(1 + z(e^{np_j r} - 1))^2},$$

and

$$H_{rz} = n \sum_j \frac{p_j e^{np_j r}}{(1 + z(e^{np_j r} - 1))^2}. \quad (3.34)$$

Now (3.33) simplifies to

$$H_{zz} = \frac{1}{z} \sum_j \frac{(e^{np_j r} - 1)}{(1 + z(e^{np_j r} - 1))^2}. \quad (3.35)$$

For $H_{rr}$ we can put under a common denominator and get

$$H_{rr} = \sum_j \frac{nzp_j e^{np_j r}(1 + z(e^{np_j r} - 1)) + rz(1 - z)n^2 p_j^2 e^{np_j r}}{r(1 + z(e^{np_j r} - 1))^2}.$$
Using the inequality $e^x - 1 \geq x$ gives

$$H_{rr} \geq \sum_j npj e^{npjr} (1 + znpjr + rz(1 - z)) r^2 p_j e^{npjr}$$

which then leads to some very convenient cancelling (in particular, of the $z^2$ term) to get

$$H_{rr} \geq \sum_j npj e^{npjr} (1 + rnpj)$$

(3.36)

Multiply (3.35) and (3.36) together:

$$H_{zz} H_{rr} \geq \sum_j npj e^{npjr} (1 + rnpj)$$

(3.37)

and use the Cauchy-Schwarz inequality to get

$$H_{zz} H_{rr} \geq \sum_j npj e^{npjr} (1 + rnpj)$$

(3.38)

We need to show that this is strictly greater than $H_{rz}^2$, which can be expressed, using (3.34), as

$$H_{rz}^2 = n^2 \left( \sum_j p_j e^{npjr} \right)^2$$

(3.39)

Taking square roots of the expressions (3.37) and (3.38), the condition $\chi > 0$ is equivalent to

$$\sum_j \left[ (e^{npjr} - 1)p_j (1 + npjr)e^{npjr} \right]^{1/2} - \sqrt{rnpj e^{npjr}} > 0.$$
This is equivalent to showing that \( \forall j \),

\[
f(np_j r) \geq 0, \quad (3.40)
\]

where we define \( f(x) = (e^x - 1)x(1 + x)e^x - x^2 e^{2x} \). Now

\[
f(x) = xe^{2x} + x^2 e^{2x} - xe^x - x^2 e^x - x^2 e^{2x}
\]

\[
= xe^{2x} - xe^x - x^2 e^x
\]

\[
= xe^x(e^x - 1 - x) > 0 \quad \text{for} \quad x > 0.
\]

Therefore the inequalities in (3.40) hold (and at least one of them is strict), and so the lemma follows. \( \square \)

Moreover, \( z(b) \) is strictly increasing (see (3.48) below), so that indeed \( z(b) < 1 \) for \( b < b_\ast \), and \( z(b) > 1 \) for \( b > b_\ast \). So, introducing

\[
h(b) = H(z(b), r(b), b), \quad (3.41)
\]

we have

\[
\pi_{kb}^- \leq 3\sqrt{k} \exp(h(b)), \quad b < b_\ast, \quad (3.42)
\]

\[
\pi_{kb}^+ \leq 3\sqrt{k} \exp(h(b)), \quad b > b_\ast; \quad (3.43)
\]

here \( h(b_\ast) = H(z_\ast, r_\ast, b_\ast) = 0 \).
3.2.4 Taylor Expansion of $h$ about its Stationary Point

To get efficient bounds from (3.42)-(3.43), let us approximate $h(b)$ by its Taylor polynomial about $b_*$. First, using (3.27)-(3.28),

$$h'(b) = \frac{d}{db} H(z(b), r(b), b) = H_z(z(b), r(b), b) z'(b) + H_r(z(b), r(b), b) r'(b) + H_b(z(b), r(b), b) = H_b(z(b), r(b), b) = -\ln(z(b)).$$  \hfill (3.44)

It follows that $h(b)$ is unimodal (concave down, in fact), attaining its zero maximum at $b = b_*$. Consequently

$$h(b) = h(b_*) + h'(b_*)(b - b_*) + \frac{h''(\tilde{b})}{2} (b - b_*)^2 = \frac{h''(\tilde{b})}{2} (b - b_*)^2,$$  \hfill (3.45)

$\tilde{b}$ being between $b$ and $b_*$. To get a bound on $h''$, we now show

**Lemma 3.2.4.** With $h$ as defined in (3.41), we have, uniformly for $\tilde{b} \leq k$,

$$h''(\tilde{b}) \leq -\frac{1}{k}.$$  

**Proof.** First note that

$$h''(\tilde{b}) = \frac{d}{db} \left( h'(\tilde{b}) \right) = \frac{d}{db} \left( -\ln(z(\tilde{b})) \right) \quad \text{by (3.44)}$$

$$= -z'(\tilde{b})/z(\tilde{b}).$$  \hfill (3.46)

Notice that

$$H_{rb} = 0 \quad \text{and} \quad H_{zb} = -1/z.$$  \hfill (3.47)
To find $z'(\tilde{b})$, differentiate $H_z(z(\tilde{b}), r(\tilde{b}), \tilde{b}) = 0$ and $H_r(z(\tilde{b}), r(\tilde{b}), \tilde{b}) = 0$ with respect to $\tilde{b}$, using (3.47). We can solve for $z'(\tilde{b})$ in this system to get

$$z'(\tilde{b}) = \frac{H_{rr}}{z(\tilde{b})(H_{zz}H_{rr} - H_{rz}^2)} = \frac{H_{rr}}{z(\tilde{b})\chi}, \quad (3.48)$$

which is strictly positive by (3.36) and Lemma 3.2.3. Now note that, using the expression (3.35) for $H_{zz}$, and (3.32) to express $\tilde{b}$,

$$\frac{H_{zz}}{\tilde{b}} = \frac{1}{z(\tilde{b})^2b} \sum_j \frac{z(\tilde{b})(e^{np_jr(\tilde{b})} - 1)}{[1 + z(\tilde{b})(e^{np_jr(\tilde{b})} - 1)]^2} \leq \frac{1}{z(\tilde{b})^2b} \sum_j \frac{z(\tilde{b})(e^{np_jr(\tilde{b})} - 1)}{1 + z(\tilde{b})(e^{np_jr(\tilde{b})} - 1)} = \frac{1}{z(\tilde{b})^2}. \quad (3.49)$$

Therefore, using (3.46), (3.48), and (3.50),

$$h''(\tilde{b}) = \frac{z'(\tilde{b})}{z(\tilde{b})} = -\frac{H_{rr}}{z(\tilde{b})^2\chi} \leq -\frac{H_{rr}}{b\chi} \leq -\frac{1}{\tilde{b}}, \quad (3.51)$$

where the last inequality holds because $0 < \chi < H_{rr}H_{zz}$. The fact that $b \leq k$, $\Phi_p(k) \leq k$, and $\tilde{b}$ is between $b$ and $\Phi_p(k)$ means that $\tilde{b} \leq k$. Therefore, (3.51) yields

$$h''(\tilde{b}) \leq -\frac{1}{k}.$$ 

Lemma 3.2.4, combined with (3.42), (3.43), and (3.45) imply (3.17) and (3.18), which completes the proof of Theorem 3.2.1.
CHAPTER 4

PROOF OF THE MAIN THEOREM

We are now in position to prove our main Theorem 2.3.1. We will break it up into several steps. By far the most involved is to show that under our conditions (2.16), the expected time spent in certain “early” states is negligible. Then, in a much more straightforward manner, we show that the time spent in the late phase is bounded above by $2c_2^{-1}(1 + o(1))$. We conclude the proof of the main theorem by providing a quick proof of the lower bound $E[T] \geq 2c_2^{-1}(1 - o(1))$, which will allow us to prove the distribution result, which is Theorem 2.3.2.

4.1 Early-Phase Upper Bound

The fact that this early phase provably contributes a negligible amount to the total coalescence time has a practical biological application. Suppose one wishes to determine the time $T$ back to the MRCA from the whole of a very large population of size $n$. If the reproductive mechanism fits the current model, then one can get away with taking a much smaller sample $M$ of size $m(n)$, where $m(n) \to \infty$ however slowly. The time $T_M$ back to the MRCA of $M$ will then only differ from $T$ (in expectation) by a $o(1)$ factor.
4.1.1 Identifying and iterating a likely recurrence inequality

Let

\[ k_* := c_2^{-1} \ln^{-\varepsilon} n, \]  

with \( \varepsilon \) coming from (2.16). (\( k_* \) is meant to be an integer, as is another parameter \( k_1 \) defined later, but for simplicity we omit the “integer part” notation.) This \( k_* \) will serve as a threshold separating the “early” states (\( B(t) > k_* \)) from the “late” states (\( B(t) \leq k_* \)). So, in light of the informal discussion in the introduction, “\( k_* = o(c_2^{-1}) \)” should be more or less expected, because we will allow all states \( \leq k_* \) to be visited at least once. The need for an additional factor, \( \ln^{-\varepsilon} n \), will become clear later, in Section 4.2.

Given our expected one-step behavior given by (3.1), our immediate task is to identify a function \( \Psi_p(k) \), such that, intuitively at least, the random sequence \( \{B(t)\} \) \( \text{whp} \) satisfies a recurrence inequality

\[ B(t + 1) \leq \Psi_p(B(t)), \quad \text{if} \quad B(t) \geq k_. \]  

Then, for \( k \geq k_* \), \( \Psi_p(k) \) needs to be large enough so that, conditionally on \( \{B(t) = k\} \), the event \( \{B(t + 1) \leq \Psi_p(B(t))\} \) is very likely. To this end, because of (3.1), we must have \( \Psi_p(k) > \Phi_p(k) \). Also, to be of any use, \( \Psi_p(k) \) must fall below \( k \): it does us no good to know that \( \text{whp} \ B(t + 1) \leq B(t) \) — this is always the case. Last, but not least, our choice of \( \Psi_p \) must allow us to solve the recurrence (4.2). The function

\[ \Psi_p(k) = \frac{1}{2} (k + \Phi_p(k)) = \frac{1}{2} (k + n - F_p(k)), \]

45
with $F_p(k)$ as defined in (3.2), certainly meets the first two requirements, and will also be amenable to the third.

Define the intermediate coalescence times $\tau(k)$ by

$$\tau(k) := \min_t \{B(t) \leq k\},$$

and define a “good” event

$$\Delta := \{\forall t : B(t) \geq k_* \implies B(t + 1) \leq \Psi_p(B(t))\}.$$ 

In Section 4.1.2, we will show that $P(\Delta) \to 1$ as $n \to \infty$. In this section, assuming that the event $\Delta$ holds, we solve the recurrence (4.2) and estimate sharply $\tau(k_*)$, the first moment $t$ when $B(t) \leq k_*$. 

**Lemma 4.1.1.** On the event $\Delta$,

$$\tau(k_*) \leq 5c_2^{-1/2} \ln n = o(c_2^{-1}).$$

**Note:** For the rest of this section, we consider $B(t) \geq k_*$, and assume that $\Delta$ holds.

**Proof.** The proof is divided into two cases, depending on how close $p$ is to uniform.

**Case I.** First consider $c_2 \geq 2n^{-1}$, meaning $p$ is ‘far enough’ from being uniform. First of all, by Proposition 3.1.2, we have

$$\Psi_p(k) \leq \Psi_{\theta(c_2)}(k) \quad \forall k \in [n],$$

$\theta(c_2)$ being the topheavy distribution $(\theta_1, \theta_2, \ldots, \theta_2)$ with parameter $c_2$. Therefore on the event $\Delta$,

$$B(t + 1) \leq \Psi_{\theta(c_2)}(B(t)), \quad \forall t \leq \tau(k_*). \quad (4.4)$$
Let us bound $\Psi_{\theta(c_2)}(B(t))$ from above. Using the equations

$$\theta_1 + (n - 1)\theta_2 = 1 \quad \text{and} \quad \theta_1^2 + (n - 1)\theta_2^2 = c_2,$$

we obtain

$$\theta_1 = 1 + \sqrt{(n - 1)(c_2 n - 1)} \quad \text{and} \quad \theta_2 = \frac{1 - \theta_1}{n - 1}.$$  \hfill (4.5)

Since $c_2 \geq 2n^{-1}$, we get

$$\theta_1 \geq \frac{1}{2}c_2^{1/2} \quad \Rightarrow \quad \theta_2 \leq \frac{1}{n - 1} - \frac{(1/2)c_2^{1/2}}{n - 1}.$$  \hfill (4.6)

By (4.6) and

$$\Psi_{\theta}(B(t)) = \frac{1}{2}(B(t) + n - e^{-\theta_1 B(t)} - (n - 1)e^{-\theta_2 B(t)}),$$

we have (using the inequality $e^{-x} \geq 1 - x$)

$$B(t + 1) \leq \frac{1}{2}(B(t) + n - 0 - (n - 1)(1 - \theta_2 B(t)))$$
$$\leq \frac{1}{2} \left(B(t) + (1 - c_2^{1/2}/2)B(t) + 1\right)$$
$$\leq \left(1 - \frac{c_2^{1/2}}{4}\right)B(t) + \frac{1}{2},$$  \hfill (4.7)

a linear recurrence inequality. (Implicit in this derivation is an intuition that, for the distribution $\theta$ in question, a large enough proportion of collisions happen in box 1, and that we may disregard collisions in boxes 2, . . . , n without inducing too large an error. In effect, the inequality $e^{-\theta_2 B(t)} \geq 1 - \theta_2 B(t)$ ignores any collisions that may happen in boxes 2, . . . , n.) We can solve the recurrence (4.8) (using the fact that $B(0) = n$) and get

$$B(t) \leq n \left(1 - \frac{c_2^{1/2}}{4}\right)^t + 2c_2^{-1/2} \quad \text{for} \quad B(t) \geq k_*. \hfill (4.9)$$
To get a bound on $\tau(k_*)$, notice that

$$k_* < B(\tau(k_*) - 1) \leq n \left( 1 - \frac{c_2^{1/2}}{4} \right)^{\tau(k_*)^{-1}} + 2c_2^{-1/2}$$

and let $\tau := \tau(k_*) - 1$. Now $c_2^{-1/2} = o(k_*)$ if $c_2 = o(\ln^{-2\varepsilon} n)$, which is certainly implied by (2.16). So we can crudely use the bound $2c_2^{-1/2} \leq (1/2)k_*$ in (4.10) to get

$$\left( 1/2 \right) k_* \leq n(1 - c_2^{1/2}/4)\tau \leq n \exp \left(-c_2^{1/2}\tau/4\right).$$

Taking logarithms and solving for $\tau$, we obtain

$$\tau \leq 4 \left( \ln n + \ln c_2 + \varepsilon \ln \ln n + \ln 2 \right) c_2^{-1/2} \leq 5c_2^{-1/2} \ln n.$$

Hence $\tau(k_*) = o(c_2^{-1})$, since $c_2 = o(\ln^{-2} n)$, which is the first condition in (2.16).

**Case II.** Now consider $c_2 \leq 2n^{-1}$. This time $\theta(c_2)$ is too close to being uniform, and the inequality (4.7) is too crude. A bit of reflection shows that we should not expect that $B(t)$ decay exponentially here. We show instead:

**Lemma 4.1.2.** For some absolute constant $A$,

$$B(t) \leq \frac{An}{t + 1}, \quad t \leq \tau(k_*).$$

**Proof.** We proceed by induction. The case $t = 0$ holds if $A \geq 1$. Suppose (4.12) holds for some $t$. Observe that

$$\Psi_p(k) = \sum_{j=1}^n \psi(p_jk), \quad \psi(x) := \frac{1}{2} (x + 1 - e^{-x}).$$
and that $\psi(x)$ is increasing and concave down. Then by the inductive assumption,

$$B(t + 1) \leq \sum_{j=1}^{n} \psi(p_j B(t))$$

$$\leq \sum_{j=1}^{n} \psi\left(p_j \frac{A n}{t + 1}\right)$$

$$\leq n \psi\left(\frac{1}{n} \sum_{j=1}^{n} p_j \frac{A n}{t + 1}\right)$$

$$= n \psi\left(\frac{A}{t + 1}\right).$$

So we need to find $A \geq 1$ such that

$$\psi\left(\frac{A}{t + 1}\right) \leq A/(t + 2), \quad (4.13)$$

or, defining $x := A/(t + 1),

$$\psi(x) \leq x \frac{t + 1}{t + 2} \iff 1 - e^{-x} \leq x \left(1 - \frac{2}{t + 2}\right). \quad (4.14)$$

We therefore define $x(t)$ implicitly by

$$1 - e^{-x(t)} = x(t) \left(1 - \frac{2}{t + 2}\right); \quad (4.15)$$

by considering the graphs of the functions of $x(t)$ on the left- and right-hand sides of $\,(4.15),\,$ it is clear that $x(t)$ is well-defined and decreasing (to 0) in $t$. Therefore $\,(4.14)\,$ is satisfied iff $x \geq x(t)$. Using the Taylor expansion of $e^{-x(t)}$, we have

$$\frac{x(t)}{2} - \frac{x(t)^2}{6} \leq \frac{2}{t + 2} \leq \frac{x(t)}{2}$$

which therefore implies (since $x(t) \to 0$ as $t \to \infty$) that

$$x(t) \sim \frac{4}{t + 2}, \quad t \to \infty,$$
and so \( A_* := \limsup_{t \to \infty} (t + 1)x(t) \) is finite. Thus to satisfy (4.13) and thereby to complete the inductive proof, we can pick \( A = \max\{1, A_*\} \).

Therefore, on the event \( \Delta \), we have

\[
k_* < B(\tau(k_*) - 1) \leq \frac{An}{\tau(k_*)}
\]

which we can invert to get

\[
\tau(k_*) \leq \frac{An}{k_*} = A' \ln^\varepsilon n,
\]

which is certainly \( o(c_2^{-1/2} \ln n) \). Combining this with case I, \( c_2 \geq 2/n \), we have \( \tau(k_*) = O(c_2^{-1/2} \ln n) \) on the event \( \Delta \). This completes the proof of Lemma 4.1.1.

We now need to show that \( P(\Delta) \) converges to 1, and that it does so sufficiently fast.

### 4.1.2 Using the exponential tail bounds

We will need (3.18) from Theorem 3.2.1. It gives, for \( b = \Psi_p(k) \) (recall the definition (4.3) of \( \Psi_p(k) \)),

\[
\pi_{k,\Psi_p(k)}^+ \leq 3\sqrt{k} \exp \left[ -\frac{(\Psi_p(k) - \Phi_p(k))^2}{2k} \right]
\]

\[
= 3\sqrt{k} \exp \left[ -\frac{(k - \Phi_p(k))^2}{8k} \right].
\]

Introducing

\[
H_p(k) := \frac{1}{k}(\Psi_p(k) - \Phi_p(k))^2;
\]

we rewrite (4.16) as

\[
\pi_{k,\Psi_p(k)}^+ \leq 3\sqrt{k} \exp \left( -\frac{1}{2}H_p(k) \right).
\]
The next Lemma states, roughly, that the larger $k$ is, the more likely it is that the next state $B(t+1)$ is close to the prediction based on information $B(t) = k$.

**Lemma 4.1.3.** For all $k$ and $p$, $H_p(k)$ is increasing in $k$.

*Proof.* Define

$$N_p(k) := \Psi_p(k) - \Phi_p(k) = \frac{1}{2} \left( k - n + \sum_{j=1}^{n} e^{-p_j k} \right)$$

so that $H_p(k) = k^{-1} N_p(k)^2$. Then

$$H'_p(k) = \frac{N_p(k)}{k^2} \left[ 2k N'_p(k) - N_p(k) \right],$$

where

$$2k N'_p(k) - N_p(k) = \frac{1}{2} \sum_{j=1}^{n} (p_j k + 1 - (2p_j k + 1)e^{-p_j k}) > 0,$$

because

$$f(x) := x + 1 - (2x + 1)e^{-x} > x + 1 - \frac{2x + 1}{x + 1} = \frac{x^2}{x + 1} > 0 \text{ for } x > 0.$$

\[\square\]

**Lemma 4.1.4.** For all $k \geq k_*$,

$$H_p(k) \geq A_* \ln^{1+\varepsilon} n,$$

where $A_*$ is some absolute constant. Thus the probability in (4.18) is superpolynomially small.
Proof. As before, we consider separately the cases where \( p \) is close to uniform or not.

**Case I.** Consider first \( c_2 \geq 2n^{-1} \). From Proposition 3.1.3 and Lemma 4.1.3, it follows that, for all \( k \geq k_* \),

\[
H_p(k) \geq H_p(k_*) \geq H_r(k_*);
\]

here

\[
H_r(k_*) = \frac{(\Psi_r(k_*) - \Phi_r(k_*))^2}{k_*}
\]

\[
= \frac{1}{4k_*} \left[ \sum_{j=1}^{n} (e^{-r_j k_*} - 1 + r_j k_*) \right]^2
\]

\[
\geq \frac{1}{4k_*} \left[ \nu (e^{-r_1 k_*} - 1 + r_1 k_*) \right]^2. \tag{4.19}
\]

To bound (4.19) from below we need to have sharp bounds for \( r_1 \) and \( \nu \). Recalling the definition of \( r \) in Lemma 3.1.3, and letting \( \mu := n - \nu - 1 \), we have

\[
\nu r_1 + r_2 + \mu r_3 = 1, \quad \nu r_1^2 + r_2^2 + \mu r_3^2 = c_2, \quad \nu r_1^3 + r_2^3 + \mu r_3^3 = c_3. \tag{4.20}
\]

Obviously \( r_3 \leq n^{-1} \). To bound \( c_3 \), notice that by the Cauchy-Schwarz inequality,

\[
c_3 = \left( \sum_j p_j^3 \right) \left( \sum_j p_j \right)
\]

\[
\geq \left( \sum_j p_j^{3/2} p_j^{1/2} \right)^2
\]

\[
= c_2^2. \tag{4.21}
\]

Since we are assuming that \( c_2 \geq 2n^{-1} \), this gives

\[
c_3 \geq c_2^2 \geq \frac{4}{n^2}.
\]
Hence
\[
\mu r^2 \leq \frac{\mu}{n^2} \leq \frac{1}{n} \leq \frac{c_2}{2}, \quad \mu r^3 \leq \frac{\mu}{n^3} \leq \frac{1}{n^2} \leq \frac{c_3}{4}.
\]

(4.22)

Combining (4.20), (4.22) and \( r_2 \leq r_1 \) we get
\[
\frac{c_2}{4} \leq \nu r^2_1 \leq c_2, \quad \frac{3c_3}{8} \leq \nu r^3_1 \leq c_3.
\]

These double inequalities imply directly that
\[
\frac{3c_3}{8} \leq r_1 \leq \frac{4c_3}{c_2}, \quad \frac{1}{64} \frac{c_2^3}{c_3^2} \leq \nu \leq \frac{64c_2^3}{9c_3^3}.
\]

(4.23)

Armed with (4.23) we return to (4.19). Recalling that \( k_* = c_2^{-1} \ln^{-\varepsilon} n \), we need to consider separately the subsequences \( \{n_i\} \) such that \( r_1 k_* = O(1) \) for \( n \in \{n_i\} \), and the subsequences \( \{n_i\} \) along which \( r_1 k_* \to \infty \). In the first case
\[
\inf_{n_i} \frac{e^{-r_1 k_*} - 1 + r_1 k_*}{r_1^2 k_*^2} \geq \delta > 0.
\]

So, using (4.23), we obtain from (4.19):
\[
H_r(k_*) \geq \frac{\delta^2}{4} \nu^2 r_1^4 k_*^3 \geq \frac{\delta^2}{256} c_2^{-1} \ln^{-3\varepsilon} n.
\]

In the second case, for \( n_i \) large enough,
\[
e^{-r_1 k_*} - 1 + r_1 k_* \geq \frac{1}{2} r_1 k_*.
\]

So, using (4.23) again,
\[
H_r(k_*) \geq \frac{1}{16} \nu^2 r_1^2 k_* \geq \frac{9}{2^{22} c_3^3} \ln^{-\varepsilon} n.
\]

Therefore there exists a constant \( A > 0 \) such that, for \( n \) large enough,
\[
H_r(k_*) \geq A \min \left\{ c_2^{-1} \ln^{-3\varepsilon} n, \frac{c_3^3}{c_3^3} \ln^{-\varepsilon} n \right\}.
\]
Since
\[ c_3 \leq c_2^{3/2} \ln^{- (1/2 + \varepsilon)} n, \quad c_2 \leq \ln^{-2} n, \quad (4.24) \]
the last inequality leads to
\[ H_r(k_*) \geq A \ln^{1+\varepsilon} n, \]
as long as \( \varepsilon < 1/4 \), which of course we may assume without loss of generality.

**Case II.** Consider now the case \( c_2 \leq 2n^{-1} \). This time, by Proposition 3.1.1 and Lemma 4.1.3, for all \( k \geq k_* = c_2^{-1} \ln^{-\varepsilon} n \),
\[ H_p(k) \geq H_p(k_*) \geq H_u(k_*), \]
where
\[ H_u(k_*) = \frac{1}{4k_*} (k_* - n + ne^{-k_*/n})^2 \geq \frac{k_*^3}{36n^2} \geq \frac{1}{288} n \ln^{-3\varepsilon} n. \]
For the first inequality we used
\[ e^{-x} - 1 + x \geq \frac{x^2}{2} - \frac{x^3}{6} \geq \frac{x^2}{3}, \quad \text{for} \ x \in (0,1). \]
Thus \( H_p(k) \gg \ln^{1+\varepsilon} n \) for this case, and this concludes the proof of Lemma 4.1.4.

**4.1.3 Validation of the deterministic approximation**

Now that we have established the (superpolynomially) small bound for \( \pi_{k,\Psi_p(k)}^{+} \), we can finally show that the event \( \Delta \) is extremely likely indeed.

**Lemma 4.1.5.** For some constant \( A > 0 \),
\[ P(\Delta) \geq 1 - n \exp \left( -A \ln^{1+\varepsilon} n \right) \geq 1 - n^{-K}, \quad \forall K > 0, \ n \geq n(K). \]
Remark. Borrowing a term from Knuth et al. [10], the event $\Delta$ holds quite surely ($qs$).

Proof. Introduce the events

$$C_t = \left\{ B(s + 1) \leq \Psi_p(B(s)) \forall s < t, \ B(t + 1) > \Psi_p(B(t)), B(t) \geq k_s \right\};$$

that is, $C_t$ is the event that the recursive inequality $B(s + 1) \leq \Psi_p(B(s))$ is violated at a state $B(t) \geq k_s$, and $t$ is the first such moment. Clearly

$$\Delta^c = \bigcup_{t \geq 0} C_t.$$ We expect each $C_t$ to be quite rare, but we still need an upper bound on the possible number of these ‘bad’ events. To that end, let us show that

$$C_t = \emptyset, \quad \forall t \geq t_* := 4c_2^{-1/2} \ln n.$$ Suppose that on the contrary $C_{t_1} \neq \emptyset$ for some $t_1 > t_*$. Then, by the definition of $C_{t_1}$, we have

$$B(s + 1) \leq \Psi_p(B(s)), \quad \forall s \leq t_*,$$

and certainly $B(t_*) \geq k_s$. However, using this recurrence inequality exactly as in the derivation of (4.9), we must have

$$B(t_*) \leq n \left( 1 - c_2^{1/2}/4 \right)^{t_*} + 2c_2^{-1/2}$$

$$\leq n \exp \left[ - \left( 4c_2^{-1/2} \right) (\ln n) \left( c_2^{1/2}/4 \right) \right] + 2c_2^{-1/2}$$

$$= 1 + 2c_2^{-1/2} < k_s$$
since $c_2 \ll \ln^{-2\varepsilon} n$. Contradiction! Thus $\Delta^c$ is a union of at most $t_*$ events $C_t$. Now, by (4.16) and Lemma 4.1.3, we have

$$P(C_t) \leq 3n^{1/2} \exp(-H_p(k_*)/2), \quad t \geq 0,$$

since $B(t) \geq k_*$ on $C_t$. Therefore

$$P(\Delta^c) = P\left(\bigcup_{t \leq t_*} C_t\right) \leq \sum_{t \leq t_*} P(C_t) \leq 3t_* n^{1/2} \exp(-H_p(k_*)/2) = 12c_2^{-1/2} n^{1/2} \ln n \exp(-H_p(k_*)/2) \leq n \exp(-0.5A_* \ln^{1+\varepsilon} n) \quad \text{by Lemma 4.1.4.}$$

This completes a program we put forth at the end of Chapter 2. Combining Lemma 4.1.1 and this last Lemma 4.1.5, we have proved the following.

**Lemma 4.1.6.** Let $k_* = c_2^{-1} \ln^{-\varepsilon} n$ for some $\varepsilon \in (0, 1/4)$ however small, and let $\tau(k_*)$ denote the random moment when $B(t)$ falls to or below $k_*$ for the first time. Then, for some constant $a > 0$,

$$P\{\tau(k_*) \leq ac_2^{-1/2} \ln n\} \geq 1 - n^{-K}, \quad \forall K > 0, \ n \geq n(K).$$

In short, $qs \tau(k_*) = o(c_2^{-1})$. 

56
4.1.4 Bounding the expectation of $\tau(k_*)$

Even though $\tau(k_*) = o(c_2^{-1}) qs$, proving that $E[\tau(k_*)] = o(c_2^{-1})$ as well is not straightforward, since we do not have a polynomial (worst-case) bound for $\tau(k_*)$. As a first step, introducing the event indicators $I_\Delta, I_{\Delta'}$, we split $E[\tau(k_*)]$ using $1 = I_\Delta + I_{\Delta'}$ and bound the second summand via the Cauchy-Schwarz inequality:

$$E[\tau(k_*)] = E[\tau(k_*) I_\Delta] + E[\tau(k_*) I_{\Delta'}]$$

$$\leq 5(\ln n)c_2^{-1/2} + \sqrt{E[\tau^2(k_*)] \cdot E[I_{\Delta'}^2]}$$

$$\leq 5(\ln n)c_2^{-1/2} + \sqrt{E[\tau^2(k_*)] \cdot P(\Delta')}.$$  \hspace{1cm} (4.25)

It remains to show that $E[\tau^2(k_*)]$ is at most polynomially large. To do so, introduce $T(k)$, the random time the process $\{B(t)\}$ spends at state $k$, i.e.

$$T(k) = |\{t \geq 0 : B(t) = k\}|,$$  \hspace{1cm} (4.26)

and note that

$$\tau(k) = \sum_{i=k+1}^n T(i).$$

Then

$$\tau^2(k_*) = \left( \sum_{k=k_*+1}^n T(k) \right)^2 \leq (n - k_*) \sum_{k=k_*+1}^n T(k)^2,$$

again by the Cauchy-Schwarz inequality. Therefore

$$E[\tau^2(k_*)] \leq n \sum_{k=k_*+1}^n E[T(k)^2] \leq n \sum_{k=k_*+1}^n E[T(k)^2 | T(k) > 0].$$

Recalling the notation $\pi_{kk} = P(B(t+1) = k | B(t) = k)$, we observe that

$$P\{T(k) = j | T(k) > 0\} = \pi_{kk}^{j-1}(1 - \pi_{kk}), \quad j > 0,$$  \hspace{1cm} (4.27)

57
i.e. conditioned on \( \{ T(k) > 0 \} \), \( T(k) \) is geometrically distributed, with success probability \( 1 - \pi_{kk} \). In particular,

\[
E[T(k) \mid T(k) > 0] = \frac{1}{1 - \pi_{kk}}, \quad \text{Var}[T(k) \mid T(k) > 0] = \frac{\pi_{kk}}{(1 - \pi_{kk})^2}.
\]

It is obvious intuitively, and can be easily proved, that \( \pi_{kk} \) decreases with \( k \): the larger the number of balls—the larger the probability of collision. Consequently both conditional moments of \( T(k) \) decrease with \( k \). So

\[
E[\tau^2(k^*)] \leq n \sum_{k=k_*+1}^{n} \left( \frac{\pi_{kk}}{(1 - \pi_{kk})^2} + \frac{1}{(1 - \pi_{kk})^2} \right)
\]

\[
\leq n \sum_{k=k_*+1}^{n} \frac{2}{(1 - \pi_{kk})^2}
\]

\[
\leq \frac{2n^2}{(1 - \pi_{k, k_*})^2}.
\]

Therefore, it remains to show only that \( \frac{1}{1 - \pi_{k, k_*}} \) is at most polynomially large in \( n \).

Using the simplest Inclusion-Exclusion lower bound for the probability of the union of events, we write

\[1 - \pi_{kk} = P(\text{there is a collision during a } k\text{-allocation}) \geq \sum_{\{a,b\}\subset[k]} P(\text{balls } a \text{ and } b \text{ collide}) - \sum_{\{c,d\}\neq\{e,f\}\subset[k]} P(\text{c, d collide and e, f collide}).\]

Now there are two ways for two pairs of balls \( \{c, d\}, \{e, f\} \) to collide: either these sets overlap or they don’t. This explains the presence of the last two terms in what follows.
The $a_i$ will denote some absolute constants, whose exact values are not important to us.

$$1 - \pi_{kk} \geq \left(\frac{k}{2}\right) c_2 - \left(a_1 k^3 c_3 + a_2 k^4 c_2^2\right)$$

$$= \left(\frac{k}{2}\right) c_2 \left(1 - a_3 k \frac{c_3}{c_2} - a_4 k^2 c_2\right). \quad (4.28)$$

Now introduce $k_1 := c_2^{-1/2} \ln^{-\varepsilon/4} n$; clearly $k_1 < k_*$, and we also have

$$\lim k_1 \frac{c_3}{c_2} = 0, \quad \lim k_1^2 c_2 = 0.$$

So, by (4.28), uniformly for all $k \leq k_1$,

$$\frac{1}{1 - \pi_{kk}} \leq \frac{1}{\left(\frac{k}{2}\right) c_2} (1 + o(1)). \quad (4.29)$$

Consequently

$$\frac{1}{1 - \pi_{k_* k_*}} \leq \frac{1}{1 - \pi_{k_1 k_1}} = O(k_1^{-2} c_2^{-1}) = O(\ln^{\varepsilon/2} n), \quad (4.30)$$

and so not only is $(1 - \pi_{kk})^{-1}$ polynomially small, but actually logarithmically so.

Thus by (4.25),

$$E[\tau(k_*)] = O(c_2^{-1/2} \ln n) = o(c_2^{-1}). \quad (4.31)$$

### 4.2 Late-Phase Upper Bound

To complete the proof of Theorem 2.3.1 it remains to bound the expected duration of the process after the number of balls has dropped below $k_*$. 
We define a middle phase as \([\tau(k_*), \tau(k_1)]\), the time interval during which the number of balls is below \(k_*\) and above \(k_1\). Using the \(T(k)\) defined in (4.26), we have

\[
E[\tau(k_1) - \tau(k_*)] = \sum_{k=k_1+1}^{k_*} E[T(k)]
\]

\[
= \sum_{k=k_1+1}^{k_*} E[T(k)|T(k) > 0] \cdot P(T(k) > 0)
\]

\[
\leq \sum_{k=k_1+1}^{k_*} \frac{1}{1 - \pi_{kk}}.
\]

Then, by decreasing monotonicity of \((1 - \pi_{kk})^{-1}\) and (4.30),

\[
E[\tau(k_1) - \tau(k_*)] \leq \frac{k_* - k_1}{1 - \pi_{k_1k_1}}
\]

\[
= O(c_2^{-1} \ln^{-\varepsilon/2} n)
\]

\[
= O(c_2^{-1} \ln^{-\varepsilon/2} n)
\]

\[
(4.32)
\]

\[
= o(c_2^{-1}).
\]

(4.33)

(The last computation explains at long last why we needed the \(\ln^{-\varepsilon} n\) factor in the definition (4.1) of \(k_*\).

Naturally, we define a late phase as \([\tau(k_1), \tau(1)]\). By (4.29),

\[
E[\tau(1) - \tau(k_1)] \leq 1 + \sum_{k=2}^{k_1} \frac{1 + o(1)}{c_2(k)} = 2c_2^{-1}(1 + o(1)),
\]

(4.34)

where we used the following to make the sum telescope:

\[
\frac{1}{k(k-1)} = \frac{1}{k-1} - \frac{1}{k}, \quad k \geq 2.
\]

As anticipated, this late phase contribute the majority of the expected time.
That does it! Adding the estimates (4.31), (4.33), and (4.34), we obtain

$$E[\tau(1)] \leq 2c_2^{-1}(1 + o(1)).$$

The proof Theorem 2.3.1 is complete.

### 4.3 Lower Bound

Here we prove the lower bound needed to complete our main Theorem 2.3.1. This part is much simpler than the upper bound, and in fact could be deduced from some existing results in the literature (e.g. [1, Theorem 2]); however we will provide our own short proof in our familiar notation. The key feature that we will exploit is that

*if we start with few enough balls, then whp every (remaining) state of the process is visited.*

The conditions under which this lower bound holds will actually be less restrictive than those which guarantee the upper bound. We state this part as a separate proposition:

**Proposition 4.3.1.** If $p$ is any probability vector that satisfies

$$c_3(p)/c_2(p) \to 0,$$  \hfill (4.35)

then we have

$$E[T_n(p)] \geq 2c_2^{-1}(1 - o(1)).$$

**Remark.** On account of the relations (using (4.21))

$$c_2^2 \leq c_3 \leq c_2^{3/2} \implies c_2 \leq \frac{c_3}{c_2} \leq c_2^{1/2},$$

we could equivalently state condition (4.35) as “$c_2(p) \to 0$”.

61
Proof. For any $m \leq n$, define $\{B_m(t)\}_{t \geq 0}$ to be a coalescence process evolving according to $p$ starting with $m$ balls, and let $T_m$ be its coalescence time. To obtain a lower bound on $E[T] = E[T_n]$, it suffices to obtain one for $E[T_m]$ for any $m \leq n$, by Proposition 2.1.1. It is therefore up to us to choose $m$ which is easiest to work with, while still providing the correct asymptotic result. Note that we are effectively “ignoring” the earliest states, because they are cumbersome to work with and (usually) contribute a negligible amount to the coalescence time.

Let $m = m(p)$, to be determined. Define $A_k$ to be the event that state $k$ is visited; that is,

$$A_k := \{T_m(k) > 0\},$$

with $T_m(k)$ defined as in (4.26). We will show that for small enough $m$, every state $\leq m$ in the process starting at $m$ is visited. Define this “good” event $S$ by

$$S := \{T_m(k) > 0 \text{ for all } k, 1 \leq k \leq m\} = \bigcap_{k=1}^{m} A_k.$$

Clearly, $P(A_k) \geq P(S)$ for any $k \leq m$, and so if we can show that $P(S) = 1 - o(1)$, then it will follow that $P(A_k) = 1 - o(1)$, uniformly for all $k \leq m$. Consider the following alternate description of $S$, which is more amenable to probabilistic estimates:

$$S^c = \bigcup_{k=2}^{m} (A_k \cap A_{k-1}^c).$$

That is, if a state is skipped, then there must be at least one instance where a state $k$ is visited but the next state $k-1$ is not (this is guaranteed by the fact that the state
$m$ is necessarily visited at $t = 0$; this is why we need to use this coupling approach).

We may therefore write

$$
P(S^c) \leq \sum_{k=2}^{m} P \left( A_k \cap A_{k-1}^c \right)$$

$$= \sum_{k=2}^{m} P \left( A_{k-1}^c | A_k \right) P(A_k)$$

$$\leq \sum_{k=2}^{m} P \left( A_{k-1}^c | A_k \right). \quad (4.36)$$

To calculate $P(A_{k-1}^c | A_k)$, notice that this is the probability that given state $k$ is visited, the next distinct state visited is $k - 2$ or lower. We therefore first reach state $k$ at time $\tau(k)$, we stay there $j$ times for some $j \geq 0$ (each time with probability $\pi_{kk}$), and then go to a state $\leq k - 2$, (with probability $1 - \pi_{kk} - \pi_{k,k-1}$). Thus

$$P(A_{k-1}^c | A_k) = \sum_{j=0}^{\infty} \pi_{kk}^j (1 - \pi_{kk} - \pi_{k,k-1})$$

$$= \frac{1 - \pi_{k,k-1}}{1 - \pi_{kk}}. \quad (4.37)$$

Now $1 - \pi_{kk} - \pi_{k,k-1}$ is the probability that the “drop” upon a $k$-allocation is greater than or equal to 2; there are two ways for this to happen: either 3 balls collide at the same box, or there are two pairs of balls that collide in different boxes (or both).

For $A \subset [k]$, let $R_A$ denote the event that the balls indexed by $A$ all land in the same
box on a $k$-allocation. Then

$$1 - \pi_{kk} - \pi_{k,k-1} = P \left( \bigcup_{a,b,c \in [k]} R_{\{a\neq b\neq c\}} \cup \bigcup_{a \neq b \neq c, d \in [k]} R_{\{a,b\} \cap R_{\{c,d\}}} \right)$$

$$\leq \sum_{a,b,c \in [k]} P(R_{\{a,b,c\}}) + \sum_{a \neq b \neq c, d \in [k]} P(R_{\{a,b\} \cap R_{\{c,d\}}})$$

$$= O(k^3) \sum_j p_j^3 + O(k^4) \left( \sum_j p_j^2 \right)^2$$

$$= O \left( k^3 \sum_j p_j^3 + k^4 \sum_j p_j^3 \right) \quad \text{by (4.21)}$$

$$= O \left( k^4 c_3 \right). \quad (4.38)$$

Now from our inclusion-exclusion bounds we have

$$1 - \pi_{kk} = P(\text{collision in a } k\text{-allocation})$$

$$\leq \sum_{\{a,b\}} P(\text{balls } a, b \text{ collide})$$

$$= \binom{k}{2} \sum_j p_j^2,$$

so that

$$\frac{1}{1 - \pi_{kk}} \geq \frac{1}{\binom{k}{2} c_2}. \quad (4.39)$$

Combining (4.37), (4.38), and (4.39) into (4.36), we get

$$P(S^c) \leq \sum_{k=2}^m O \left( \frac{k^4 c_3}{k^2 c_2} \right) \leq O(m^3 c_3/c_2).$$

Recall our assumption that $c_3/c_2 \to 0$. Therefore, as long as we choose $m = m_* = o \left( (c_2/c_3)^{1/3} \right)$ (take $m_* := (c_2/c_3)^{1/4}$, for instance), we have

$$P(S^c) = o(1),$$

64
and so $P(A_k^c) = o(1)$ uniformly for all $k \leq m_\ast$. As a last step, note that

$$E[T_{m_\ast}] = \sum_{k=2}^{m_\ast} E[T_{m_\ast}(k)]$$

$$= \sum_{k=2}^{m_\ast} E[T_{m_\ast}(k)|A_k]P(A_k)$$

$$\geq (1 - o(1)) \sum_{k=2}^{m_\ast} E[T_{m_\ast}(k)|A_k]. \quad (4.40)$$

Now, note that (exactly as in (4.27)), conditioned on $A_k$, $T_{m_\ast}(k)$ is geometric($1 - \pi_{kk}$), and hence has expected value

$$E[T_{m_\ast}(k)|A_k] = \frac{1}{1 - \pi_{kk}} \geq \frac{1}{(k/2)c_2}, \quad \text{by (4.39).}$$

Putting this into (4.40), we get a telescoping sum

$$E[T_{m_\ast}] \geq (1 - o(1)) \sum_{k=1}^{m_\ast} \frac{2}{k(k - 1)c_2}$$

$$= (1 - o(1))c_2^{-1} \sum_{k=2}^{m_\ast} \left( \frac{1}{k - 1} - \frac{1}{k} \right)$$

$$= (1 - o(1))2c_2^{-1}(1 - 1/m_\ast)$$

$$= 2c_2^{-1}(1 - o(1)) \quad \text{since} \quad m_\ast \to \infty,$$

which completes the proof of the proposition. \qed

### 4.4 Limiting Distribution

Following Fill’s [4] work for the uniform case, we will use the first-moment expression for $T$ to derive its limiting distribution. He relied on Kingman’s result (later reproved
by Dalal and Schmutz [2]) \(E[T] = 2n(1 + o(1))\) to derive
\[
\frac{T(u)}{n} \Rightarrow \sum_{k=2}^{\infty} X_k
\]
for the uniform case, where \(X_k\) are independent and exponentially distributed with parameter \(\binom{k}{2}\). We will similarly use Theorem 2.3.1 to establish

**Theorem 4.4.1.** For \(p\) satisfying (2.16),
\[
\frac{T(p)}{c_2^{-1}} \Rightarrow \sum_{k=2}^{\infty} X_k \tag{4.41}
\]
where \(X_k, k = 2, 3, \ldots\) are independent, \(X_k \sim \text{exp} \left( \binom{k}{2} \right) \).

**Proof.** We will not provide a complete proof, as the line of reasoning follows Fill [4, Theorem 3.1] almost exactly. The main difference is that we scale \(T\) by \(c_2^{-1}\) instead of \(n\). We will however indicate here the main lines of the argument.

- As we saw in the previous section, \textbf{whp} each state \(\leq m\) is visited, if we let \(m \to \infty\) slowly enough.

- Conditioned on this event, the time spent in each state \(k \leq m\) is geometric with parameter
\[
1 - \pi_{kk} = \binom{k}{2} c_2(1 + o(1)),
\]
and all these times are independent.

- In the limit, these geometric random variables may be approximated by exponential random variables
\[
X_k = c_2^{-1} \exp \left( \binom{k}{2} \right).
\]
The time spent in states $\leq m$ is therefore distributed as the sum of infinitely many such random variables, as we let $n \to \infty$ (and thus $m \to \infty$).

- The time spent in the early phase, whatever its exact distribution, has expected value $\ll c_2^{-1}$ (by Theorem 2.3.1), and thus contributes a negligible amount to the total limiting distribution. (This part of course relies on the conditions (2.16).)
CHAPTER 5

PROCESS WITH LONG EARLY PHASE

Here we will prove Theorem 2.4.1. That is, we will show that if the conditions (2.16) are sufficiently violated, then there exists a vector \( p \) for which the main theorem does not hold. Specifically, the expected coalescence time will be \( \gg c_2^{-1} \). By analyzing the proof, we will see that this is due to the fact that the early phase no longer contributes a negligible amount of time, even if the late phase retains its qualitative behavior.

We start by restating Theorem 2.4.1 in a more precise manner.

**Theorem 5.0.2.** Let \( b_0 = n \). Suppose \( c_2 = (\ln^{-2} n)\omega(n) \), where \( \omega(n) \to \infty \) however slowly. Set \( p = \theta(c_2) \), i.e. \( p \) is the topheavy distribution \((\theta_1, \theta_2, \ldots, \theta_2)\) with

\[
\theta_1^2 + (n-1)\theta_2^2 = c_2.
\]

Then for the process evolving according to this \( p \),

\[
\tau(1) \geq c_2^{-1} \frac{\sqrt{\omega(n)}}{20} \quad \text{whp}
\]

so, in particular \( E[\tau(1)] \gg c_2^{-1} \).

**Remarks**

1. Our choice of \( p \) should be expected. Indeed, the recurrence inequality (4.4) signals, intuitively, that the coalescent process for \( p = \theta(c_2) \) is a good candidate for being the slowest among all \( p \) with \( \sum_j p_j^2 = c_2 \).
2. In [1, Theorem 5], Adler et al. had proved that for \( \max_j p_j \) bounded away from 0, and the remaining \( p_j \) uniformly small, the expected coalescence time exceeds \( c_2^{-1} \) by a factor of \( \ln n \). They therefore also use a vector of top-heavy type. Their proof is much shorter than ours, as they claim that collisions outside of box 1 may be ignored, which we however feel needs justification. Certainly in the present case, where \( \theta_1 = o(1) \), we will need to treat this issue with care.

**Proof of Theorem 5.0.2.** For simplicity we let \( c := c_2 \) throughout this chapter.

### 5.1 Heuristics

For \( p = \theta(c) \), the Markov chain is almost as simple as that for the uniform \( p \). Indeed, given \( B(t) \), the number of balls that land in box 2, \ldots, \( n \) (call it \( \hat{B}(t) \)) is binomially distributed with parameters \( B(t) \) and success probability \( 1 - \theta_1 \), i.e. \( \hat{B}(t) = \text{Bin}(B(t), 1 - \theta_1) \) for short. Conditioned on \( \hat{B}(t) \), we have a uniform allocation of \( \hat{B}(t) \) balls among \( n - 1 \) boxes 2, \ldots, \( n \). And, for \( B(t) \) sufficiently large, \textbf{whp} \( \hat{B}(t) \sim (1 - \theta_1)B(t) \). So, based on our experience with deterministic approximations earlier in the paper, we should expect that—after fusing balls that landed in the same box—these \( \hat{B}(t) \) balls give birth to about

\[
(n - 1) \left[ 1 - \exp \left( - \frac{\hat{B}(t)}{n - 1} \right) \right] \sim (n - 1) \left[ 1 - \exp \left( - \frac{(1 - \theta_1)B(t)}{n - 1} \right) \right]
\]

balls for next generation. All the balls that landed in box 1, if there are any, will coalesce into one ball. Disregarding this box for now, we expect then that the process
\{B(t)\} \textbf{whp} “closely” obeys a recurrence inequality of the form

$$B(t + 1) \geq (n - 1)\varphi \left(\frac{(1 - \theta_1)B(t)}{n - 1}\right), \quad \varphi(x) := 1 - e^{-x}.$$ 

Here is a precise claim.

### 5.2 Concentration for the Deterministic Approximation

**Lemma 5.2.1.** Let $\gamma := 1 - 2e^{1/2}$ and introduce

$$\Xi(k) := (n - 1)\xi \left(\frac{\gamma k}{n - 1}\right), \quad \xi(x) := 1.5\varphi(x) - 0.5x.$$ 

Then, for $n$ sufficiently large,

$$P(B(t + 1) < \Xi(B(t)) \mid B(t) = k) \leq e^{-ke^{1/3}} + e^{-k^{3/7}n^{2}}. \quad (5.1)$$

**Remark** $\xi(0) = 0$, $\xi(x) \leq \varphi(x)$, and $\xi(x)$ is increasing for $x \leq \ln 3$. This $\Xi$ is the counterpart to the $\Psi$ we used in Section 4.1 for the upper bound: $\Xi(B(t))$ is small enough to make conditional deviation (given $B(t)$) of $B(t + 1)$ below it sufficiently rare, but large enough so that on the events $\{B(t + 1) < \Xi(B(t))\}$, the sequence does not decrease too quickly. As for $\gamma$, notice that $\gamma < 1 - \theta_1$; think of it as a likely lower bound for the proportion of balls that reach boxes 2, . . . , $n$.

**Proof of Lemma 5.2.1.** Notice first that

$$B(t + 1) \geq \mathcal{R}(t + 1),$$ 

where $\mathcal{R}(t + 1)$ is the number of boxes among 2, . . . , $n$ that host at least one of $\hat{B}(t)$
balls. (In particular, either $B(t + 1) = R(t + 1) + 1$ or $B(t + 1) = R(t + 1)$.) Denoting $P(\{\cdot\} | B(t) = k)$ by $P_k(\{\cdot\})$, we have then

$$P_k(B(t + 1) < \Xi(k)) \leq P_k(R(t + 1) < \Xi(k)) \leq P_k(R(t + 1) < \Xi(k), \hat{B}(t) \geq \gamma k) + P_k(\hat{B}(t) < \gamma k).$$

Now, denoting the c.d.f. of $R(t + 1)$ conditioned on $\{\hat{B}(t) = j\}$ by $F_j$, we have: for $j_1 < j_2$,

$$F_{j_2}(x) \leq F_{j_1}(x), \quad \forall x \geq 0.$$  

(Informally, the fewer balls we allocate among the boxes $2, \ldots, n$, the fewer nonempty boxes we end up with.) Therefore

$$P_k(R(t + 1) < \Xi(k), \hat{B}(t) \geq \gamma k) = \sum_{j \geq \gamma k} P(R(t + 1) < \Xi(k) | \hat{B}(t) = j) P_k(\hat{B}(t) = j) \leq P(\hat{B}(t) = \lceil \gamma k \rceil) \sum_{j \geq \gamma k} P_k(\hat{B}(t) = j) \leq P(R(t + 1) < \Xi(k) | \hat{B}(t) = \lceil \gamma k \rceil).$$

Consequently

$$P_k(B(t + 1) < \Xi(k)) \leq P_1 + P_2,$$

where we define the ‘bad’ event probabilities

$$P_1 := P_k(R(t + 1) < \Xi(k) | \hat{B}(t) = \lceil \gamma k \rceil),$$

$$P_2 := P_k(\hat{B}(t) < \gamma k).$$
Informally, $P_1$ is the probability of not enough balls reaching boxes $2, \ldots, n$, and $P_2$ is the probability that there are “too many” collisions in these boxes. As long as these two quantities are sufficiently small, the sequence $\{B(t)\}$ will not decrease too rapidly.

To bound $P_2$, let’s use the Chernoff bound for the tail of binomial distribution (see Mitzenmacher and Upfal [11], for instance), we have

$$P_2 = P(\text{Bin}(k, 1 - \theta_1) < \gamma k)$$

$$\leq \exp \left( -\frac{kp\delta^2}{2} \right) \bigg|_{p=1-\theta_1, \ \delta=1-\gamma(1-\theta_1)^{-1}}$$

$$\leq e^{-kc/3}, \quad (5.2)$$

as $\theta_1 \sim c_2^{1/2}$, from the definition (4.5) of $\theta_1, \theta_2$.

Now turn to $P_1$. Applying Theorem 3.2 to the boxes set $\{2, \ldots, n\}$ and the uniform distribution $u$ on this set, we obtain

$$P_1 \leq 4\sqrt{\gamma k} \exp \left( -\frac{\Phi_u(\gamma k) - \Xi(k)}{2\gamma k} \right).$$

Here, using the definition of $\xi(x)$,

$$\Phi_u(\gamma k) - \Xi(k) = (n-1)\phi \left( \frac{\gamma k}{n-1} \right) - (n-1)\xi \left( \frac{\gamma k}{n-1} \right)$$

$$= \frac{(n-1)}{2} \left[ \frac{\gamma k}{n-1} - \phi \left( \frac{\gamma k}{n-1} \right) \right].$$

So, as

$$x - \phi(x) = e^{-x} - 1 + x \geq \frac{x^2}{2} - \frac{x^3}{6} \geq \frac{x^2}{3}, \quad x \in [0, 1],$$

we have

$$P_1 \leq 3\sqrt{\gamma k} \exp \left( -\frac{(\gamma k)^4}{72\gamma k(n-1)^2} \right) \leq e^{-k^3/72n^2}. \quad (5.3)$$
The estimates (5.2)-(5.3) imply (5.1).

\[ \square \]

## 5.3 Threshold for Early Phase

To continue, let \( \hat{k} := n^{3/4} \). This will serve as our threshold between an early phase and a late phase; it will be clear shortly why we define it this way. (Actually, we could take \( \hat{k} = n^{2/3 + \gamma} \) for any \( 0 < \gamma < 1/3 \), but for the sake of simplicity we will use \( n^{3/4} \)). Introduce two events,

\[ \Gamma := \{ \forall t, B(t) \geq \hat{k} \implies B(t + 1) \geq \Xi(B(t)) \} \]

and

\[ \Pi := \{ \exists t > c^{-3/2} : B(t) \geq \hat{k} \} \].

On the event \( \Gamma \), \( B(t) \) does not decrease “too quickly” as long as \( B(t) \) is above \( \hat{k} \). On the event \( \Pi \), it follows immediately that \( \tau(\hat{k}) \geq c^{-3/2} \), whence \( \tau(\hat{k}) \gg c^{-1} \). Then, by Lemma 5.2.1,

\[
P(\Gamma^c \cap \Pi^c) \leq P \left\{ \bigcup_{t=0}^{c^{-3/2}} \{ B(t+1) < \Xi(B(t)), B(t) \geq \hat{k} \} \right\}
\]

\[
\leq (1 + c^{-3/2}) \left( e^{-kc/4} + e^{-\hat{k}3/7n^2} \right)
\]

\[
= (1 + c^{-3/2}) \left( e^{-n^{3/4}c/4} + e^{-n^{1/4}/73} \right) \to 0,
\]

i. e.

\[
P(\Gamma \cup \Pi) \to 1.
\]

If we show that \( \tau(\hat{k}) \gg c^{-1} \) on the event \( \Gamma \) as well, we will be able then to claim that \textbf{wph} \( \tau(\hat{k}) \gg c^{-1} \), and the proof of Theorem 5.1 will be complete.
5.4 Iterating the Recurrence

Define
\[ x(t) := \frac{B(t)}{n - 1} \]
to be the scaled version of \( B(t) \), and note that \( x(0) = 1 - (n - 1)^{-1} \).

To iterate the likely recurrence for \( B(t) \) given by the event \( \Gamma \), we observe that on \( \Gamma \),
\[ x(t + 1) \geq \xi(\gamma x(t)), \quad (5.4) \]
as long as
\[ x(t) \geq \frac{\hat{k}}{n - 1} \sim n^{-1/4}. \]

**Lemma 5.4.1.** Let \( n \geq 3 \). Under the recurrence (5.4),
\[ x(t) \geq \frac{2}{3} \frac{\gamma^t}{t + 1}. \quad (5.5) \]

**Proof.** The argument runs in parallel to that for the lower bound of \( B(t) \) in (4.12).

The base case \( t = 0 \) is just
\[ x(0) = 1 - (n - 1)^{-1} \geq \frac{2}{3}, \quad n \geq 4. \]

Suppose (5.5) holds for some \( t \). Since \( \xi(x) \) is increasing for \( x \leq \ln 3 \), (5.4) implies that
\[ x(t + 1) \geq \xi \left( \frac{2}{3} \frac{\gamma^t}{t + 1} \right) = \xi \left( \frac{2}{3} \frac{\gamma^{t+1}}{t + 1} \right). \]

So, to complete the inductive step, we need to show that
\[ \xi(y) \geq y \frac{t + 1}{t + 2}, \quad y := \frac{2}{3} \frac{\gamma^{t+1}}{t + 1}, \quad (5.6) \]
Define \( z \) as a root of
\[
\xi(z) = z \frac{t + 1}{t + 2} \quad \text{or} \quad 1 - e^{-z} = z \frac{t + 4/3}{t + 2}.
\]
(5.7)

Since \( 1 - e^{-1} < 2/3 \), equation (5.7) has a (unique) root \( z = z(t) \) for \( t \geq 0 \). Using \( 1 - e^{-z} \geq z - z^2/2 \), we obtain
\[
y(t) \geq \frac{4/3}{t + 2}.
\]

Inequality (5.6) holds if \( y \leq y(t) \), which is certainly so because
\[
\frac{2/3}{t + 1} \leq \frac{4/3}{t + 2}, \quad \forall t \geq 0.
\]

Thus on the event \( \Gamma \), by definition of \( y(t) \),
\[
B(t) \geq \hat{k} \implies B(t + 1) \geq (n - 1) \frac{2}{3} \frac{\gamma^{t+1}}{t + 2}.
\]
(5.8)

Lemma 5.4.2. On the event \( \Gamma \),
\[
\tau(\hat{k}) \geq \frac{\ln^2 n}{20 \sqrt{\omega(n)}} = c^{-1} \frac{\sqrt{\omega(n)}}{20}.
\]

Proof. Let the event \( \Gamma \) hold. By (5.8), and the definition of \( \tau(\cdot) \), \( \hat{\tau} := \tau(\hat{k}) \) satisfies
\[
\hat{k} \geq B(\hat{\tau}) \geq (n - 1) \frac{2}{3} \frac{\gamma^{\hat{\tau}}}{\hat{\tau} + 1}.
\]

Recalling that \( \hat{k} = n^{3/4}, \gamma = 1 - 2e^{1/2} \), and taking logarithms, we get
\[
3e^{1/2} \hat{\tau} + \ln(\hat{\tau} + 1) \geq \frac{1}{5} \ln n,
\]
or
\[
3 \frac{\sqrt{\omega(n)}}{\ln n} \hat{\tau} + \ln(\hat{\tau} + 1) \geq \frac{1}{5} \ln n.
\]
It follows immediately that
\[
\hat{\tau} \geq \frac{1}{4 \cdot 5} \ln^2 n = c^{-1} \frac{\sqrt{\omega(n)}}{20}.
\]
\[\square\]

In summary, on the event $\Gamma \cup \Pi$,
\[
\tau(\hat{k}) \geq \min \left\{ c^{-3/2}, c^{-1} \frac{\sqrt{\omega(n)}}{20} \right\} = c^{-1} \frac{\sqrt{\omega(n)}}{20}.
\]
Recalling that $P(\Gamma \cup \Pi) \to 1$, we conclude that
\[
cE[\tau(1)] \geq cE[\tau(\hat{k})] \geq E[\tau(\hat{k}) |_{\Gamma \cup \Pi}] \geq \frac{\sqrt{\omega(n)}}{20} P(\Gamma \cup \Pi) \to \infty.
\]
This concludes the proof of Theorem 5.0.2. \[\square\]

5.5 Extremal Cases

We do not wish to leave the reader with the impression that any distribution $p$ that violates (2.16) gives rise to a process where $T \gg c_2^{-1}$ whp. Indeed, take the simple example where
\[
p = \rho := \left( \frac{1}{\ln n}, \ldots, \frac{1}{\ln n}, 0, \ldots, 0 \right)
\]
where there must be $\lfloor \ln n \rfloor$ nonzero components. We will call vectors $\rho$ of this type “restricted-uniform”. Then here
\[
c_2(\rho) = \lfloor \ln(n) \rfloor \cdot \left( \lfloor \ln n \rfloor \right)^{-2} = \frac{1}{\ln n} \gg \ln^{-2} n,
\]
so certainly (2.16) is violated. But even if we start with $B(0) = n$, it must be the case that $B(1) \leq \lfloor \ln n \rfloor$ (there are only $\lfloor \ln n \rfloor$ possible destinations for the $n$
balls), and thereafter the process is exactly uniform on these first \( \lfloor \ln n \rfloor \) boxes. (In
the biological literature, this \( \lfloor \ln n \rfloor \) would be referred to as the “effective population
size”.) Kingman’s result (2.11) then applies to this smaller set of boxes,

\[
E[T(\rho)] - 1 = 2\lfloor \ln n \rfloor (1 + o(1));
\]

since \( c_2(\rho) = \frac{1}{\lfloor \ln n \rfloor} \), this still implies that

\[
E[T(\rho)] = 2c_2(\rho)^{-1}(1 + o(1))
\]

for this \( \rho \)-process.

Moreover, as a counterpart to Proposition 3.1.2, which roughly shows that for a given
\( c_2 \), topheavy-type vectors produce the slowest coalescent processes (at least for the
early phase), we show that restricted-uniform vectors produce the fastest processes
for a given \( c_2 \), in the same sense. For the sake of simplicity, assume that \( c_2 = 1/m \)
for some integer \( m \leq n \). Define \( f(x) := (1 - e^{-x})/x \), and note that \( f \) is decreasing
and concave up.

**Proposition 5.5.1.** Let \( \mathbf{p} \) be any probability vector with \( c_2(\mathbf{p}) = 1/m \). Then for all
\( k \in [n] \),

\[
\Phi_{\mathbf{p}}(k) \geq \Phi_{\rho(1/m)}(k),
\]

(5.9)

where \( \rho(1/m) = (1/m, \ldots, 1/m, 0, \ldots, 0) \).
Proof.

\[ \Phi_p(k) = \sum_j 1 - e^{-p_j k} = k \sum_j p_j \frac{(1 - e^{-p_j k})}{kp_j} \]

\[ = k \sum_j p_j f(kp_j) \]

\[ \geq kf \left( \sum_j p_j kp_j \right) \quad \text{By Jensen’s inequality} \]

\[ = kf(kc_2) \]

\[ = c_2^{-1} (1 - e^{-kc_2}) = m (1 - e^{-k/m}) = \Phi_{p(1/m)}(k). \]
CHAPTER 6

FUTURE WORK, OPEN QUESTIONS

We have not spent much effort in obtaining precise estimates for the $o(1)$ terms that appear throughout this work. However, especially if a potential application seemed to warrant it, one could certainly take care to bound these error terms, either with the techniques presented here or others. For example, given Kingman’s nonasymptotic bound $E[T_n(u)] \leq 2n - 2$ for the uniform case, it would be desirable to have an analogue for the nonuniform case, something like $E[T_n(p)] \leq 2c_2^{-1}$, rather than simply the asymptotic upper bound of $2c_2^{-1}(1 + o(1))$.

As well, all of our results so far concern the memoryless case, where, during the allocation process, the destination of a given ball does not depend on its current location. In this framework, the current number of balls $B(t)$ encodes all the information about the process, and thus the process $\{B(t)\}_{t \geq 0}$ has the Markov property. We could eschew this requirement, and more generally stipulate that for any $i, j \in [n]$, a ball from box $i$ be sent to box $j$ with probability $p_{ij}$, where $P := (p_{ij})_{i,j \in [n]}$ is an arbitrary stochastic matrix. We would then need to keep track of the random $\{0, 1\}$–vectors

$$V(t) := (V_1(t), \ldots, V_n(t)),$$

where $V_j(t)$ is 1 or 0 depending on whether there is a (fused) ball in box $j$ or not. The sequence $\{V(t)\}_{t \geq 0}$ then has the Markov property, but is clearly much more
difficult to deal with than \( \{B(t)\}_{t \geq 0} \). If the matrix \( P \) is sufficiently ‘regular’, in that there is not too much variation among its columns, then approaches similar to the one in this work may be feasible. However, for a practical random-sampling application, \( P \) is likely to be quite sparse, and thus our current techniques might be useless. If the chain driven by \( P \) has a stationary distribution \( \pi = (\pi_1, \ldots, \pi_n) \), then we may be able to do the following: run the allocation process \textit{without fusing any balls that collide} for some predetermined number \( t_{\text{mix}} \) times until the chain is “well-mixed”, in some well-defined sense. At time \( t_{\text{mix}} + 1 \), run the process the usual way, fusing any balls that collide. Ideally, the locations of the balls at time \( t_{\text{mix}} + 1 \) have distributions “sufficiently close” to \( \pi \) and are “sufficiently independent”. Repeat this two-phase “mixing/colliding” process until there is only one ball left. By a simple coupling argument, it can be shown that the time to coalescence for this new process stochastically dominates the coalescence time for the usual process (where we fuse colliding balls at every step). By invoking our main theorem, can we then conclude that the expected coalescence time for this process is given by something like \( t_{\text{mix}} \cdot 2(\sum_{j=1}^{n} \pi_j^2)^{-1} \)?

One may also wish to prove similar results for more general nonuniform allocation models than the multinomial one considered. There is a good deal of work in this direction in the papers of Möhle [12, 13]. In light of Theorems 2.3.1 and 2.4.1, one could perform a deeper investigation into the qualitative transition that happens around \( c_2 = \ln^{-2} n \), and find the analogue of this threshold for a more general model.


