RECENT COMMON ANCESTORS OF ALL PRESENT-DAY INDIVIDUALS

JOSEPH T. CHANG,* *Yale University*

Abstract

Previous study of the time to a common ancestor of all present-day individuals has focused on models in which each individual has just one parent in the previous generation. For example, 'mitochondrial Eve' is the most recent common ancestor (MRCA) when ancestry is defined only through maternal lines. In the standard Wright–Fisher model with population size n, the expected number of generations to the MRCA is about $2n$, and the standard deviation of this time is also of order n. Here we study a two-parent analog of the Wright–Fisher model that defines ancestry using both parents. In this model, if the population size n is large, the number of generations, $\mathcal{T}_n$, back to a MRCA has a distribution that is concentrated around $\lg n$ (where lg denotes base-2 logarithm), in the sense that the ratio $\mathcal{T}_n/(\lg n)$ converges in probability to 1 as $n \to \infty$. Also, continuing to trace back further into the past, at about $1.77 \lg n$ generations before the present, all partial ancestry of the current population ends, in the following sense: with high probability for large n, in each generation at least $1.77 \lg n$ generations before the present, all individuals who have *any* descendants among the present-day individuals are actually ancestors of *all* present-day individuals.

Keywords: Coalescent; Wright–Fisher model; Galton–Watson process; genealogical models; population genetics

AMS 1991 Subject Classification: Primary 92D25
Secondary 60J85

1. Introduction

Starting with the set of all of us present-day humans, imagine tracing back in time through our mothers, our mothers' mothers, and so on. This is the maternal family tree of mankind, and we are at its leaves. Recent research has suggested that the woman at the root of this tree lived roughly 100 000 or 200 000 years ago, perhaps in Africa (Cann *et al.* (1987); Vigilant *et al.* (1991)). This woman has been dubbed 'mitochondrial Eve', since all present-day human mitochondrial DNA descended from hers. Mitochondrial Eve was undoubtedly not the only woman alive at her time, so the name 'Eve' is misleading, as has been pointed out by a number of authors; see, for example, Ayala (1995). However, this misunderstanding aside, questions of the origins of mankind and the nature of our relationships to each other are still of keen interest, and the research on mitochondrial Eve has received a great deal of publicity, generating headlines in the popular press as well as in scientific publications. Svante Pääbo (1995) explains:

> ... the recent date of our mitochondrial ancestor is in a sense the really controversial conclusion from these studies. Everyone agrees that we trace our ancestry to *Homo erectus*, who emerged in Africa and from there colonized most of Eurasia about a million years ago or even earlier. What the mitochondrial data seem to show, however, is that we have a much more recent ancestor, one who lived some 100 000 or 200 000 years ago.

What captures the imagination is not the particular choice to trace back through the maternal line, but rather it is the idea that all of present-day humanity may have a common ancestor who lived as little as 100 000 years ago, a time that seems to many to be surprisingly recent. If we retain this idea while removing the restriction to the maternal line, the question becomes: How far back in time do we need to trace the full genealogy of mankind in order to find any individual who is a common ancestor of all present-day individuals? In this paper we address this sort of question in a simple mathematical model.

The *coalescent* model of Kingman (1982a) forms the basis of many of the calculations, formal and informal, used in recent treatments of questions about mitochondrial Eve and related topics. The coalescent is a large-population limit of a number of the fundamental models of population genetics, including the Wright–Fisher process. These models are haploid, with each individual in a given generation having a single parent in the previous generation. The Wright–Fisher model assumes 'random mating', in the sense that the parent of a given individual is equally likely to be any of the individuals in the previous generation. The standard model also postulates a constant population size, which may be an 'effective population size' when modeling more general situations. A number of important properties of the coalescent model are used in applications. For example, the model implies a relationship between coalescence times and population size: the expected coalescence time (measured in generations) of a large sample is about twice the population size. Hudson (1990) gives a survey of the theory and applications of the coalescent.

Here we study a natural two-parent analog of the Wright–Fisher process. (This process was previously considered by Kämmerle (1991) and Möhle (1994); see the end of this section for a discussion of related work.) We assume the population size is constant at n. Generations are discrete and non-overlapping. The genealogy is formed by this random process: in each generation, each individual chooses two parents at random from the previous generation. The choices are made just as in the standard Wright–Fisher model—randomly and equally likely over the n possibilities—the only difference being that here each individual chooses twice instead of once. All choices are made independently. Thus, for example, it is possible that when an individual chooses his two parents, he chooses the same individual twice, so that in fact he ends up with just one parent; this happens with probability $1/n$.

This model is designed only as a simple starting point for thought; of course it is not meant to be particularly realistic. Still, one might worry that this simple model ignores considerations of sex and allows impossible genealogies. If this seems bothersome, an alternative interpretation of the same process is that each 'individual' is actually a couple, and that the population consists of n monogamous couples. Then the random choices cause no contradictions: the husband and wife each were born to a couple from the previous generation. They could even come from the same couple in the previous generation.

Our interest here is in finding individuals who are common ancestors of all present-day individuals. For convenience, we use the abbreviation 'CA' to refer to a common ancestor of all present-day individuals, and 'MRCA' stands for 'most recent common ancestor.'

It turns out that mixing occurs extremely rapidly in the two-parent model, so that CAs may be found within a number of generations that depends logarithmically on the population size. In particular, our first main result says that the number of generations back to an MRCA is about $\lg n$, where $\lg$ denotes logarithm to base 2.

Theorem 1. *Let $\mathcal{T}_n$ denote the number of generations, counting back in time from the present, to an MRCA of all present-day individuals, in a population of size n. Then*

$$\frac{\mathcal{T}_n}{\lg n} \xrightarrow{P} 1 \quad as\ n \to \infty.$$

This contrasts dramatically with the one-parent situation. For example if n is 1 million, then the one-parent MRCA ('Eve') is expected to occur about 2 million generations ago, whereas a two-parent MRCA occurs with high probability within the last 20 generations or so. Also, the variability in the one-parent situation is such that the actual time to the MRCA may easily be as small as half the expected time or as large as double the expected time, say, even in arbitrarily large populations. In contrast, the time to an MRCA for the two-parent model is much less variable. For example, if the population is large enough, it is very unlikely that a random realization of the two-parent MRCA time will differ from $\lg n$ by even one percent.

This paper also addresses a second related question. Imagine tracing back through the two-parent genealogy. According to Theorem 1, after about $\lg n$ generations, we will reach the most recent generation that contains a CA. That generation might contain just one CA, or it might contain more than one. In any case, if we continue tracing back further through successive generations, then the title 'CA' becomes much less of a prestigious distinction. For example, both parents of a CA will be CAs, and all grandparents of a CA will be CAs, and so on. Eventually, in a given generation, many (and in fact most) of the individuals will be CAs. At some point we reach a generation in which some individuals are CAs (having all present-day individuals as descendants) and some are 'extinct' (having no present-day individuals as descendants), but no individual is intermediate (having some but not all present-day individuals as descendants). That is, at this point, *everyone* who is not extinct is a CA. This condition persists forever as we trace back in time: every individual is a CA or extinct. The next result shows that this condition is reached very rapidly in the model studied here.

Theorem 2. *Let $\mathcal{U}_n$ denote the number of generations, counting back in time before the present, to a generation in which each individual is either a CA of all present-day individuals or an ancestor of no present-day individual. Let γ denote the smaller of the two numbers satisfying the equation $\gamma\,e^{-\gamma} = 2\,e^{-2}$, and let $\zeta = -1/(\lg \gamma) \approx 0.7698$. Then*

$$\frac{\mathcal{U}_n}{(1 + \zeta) \lg n} \xrightarrow{P} 1 \quad as\ n \to \infty.$$

Thus, within about $1.77 \lg n$ generations, a tiny amount of time in comparison with the order n time required to get a one-parent CA, everyone in the population is either a CA of all present-day individuals or extinct.

Figure 1 shows a small example to illustrate the definitions and statements. The population size is 5. At the bottom of the figure is generation 0, the present. Going up in the graph corresponds to going back in time, so that the top row is generation -5. For each individual I in each previous generation, we calculate the set of present-day individuals (individuals in generation 0) that are descendants of I. For example, the set of present-day descendants of individual #1 in generation -1 is $\{3, 5\}$. The calculations propagate backward in time

FIGURE 1: An example illustrating the model. Here the fourth individual in generation -2 is a CA. By generation -5, all individuals are CAs or extinct: individuals 1,4, and 5 are CAs, and individuals 2 and 3 are extinct.

according to the rule: the set of descendants of an individual I is the union of the sets of descendants of the children of I. For example, the set of present-day descendants of individual #4 in generation -2 is the union $\{3, 5\} \cup \{5\} \cup \{1, 2, 4\}$, which is the whole population S. Thus, individual #4 in generation -2 is a CA of the set S of all present-day individuals. Continuing backward in time, at generation -5 we reach the stage where each individual has as descendants either the whole population S or the empty set $\varnothing$. That is, each individual in generation -5 is either a CA or extinct, having as descendants either everybody or nobody from the set of present-day individuals, and all generations prior to generation -5 also have this property. In the example shown, $\mathcal{T}_5 = 2$ and $\mathcal{U}_5 = 5$.

What is the significance of these results? An application to the world population of humans would be an obvious misuse. For example, we would not claim that a common ancestor of every present-day human may be found within the last $\lg n$ generations. Even if we took n to be 5 billion, this would imply a CA just about 32 generations ago—perhaps 500 years or so. An important source of the inapplicability of the model to this situation is the obvious non-random nature of mating in the history of mankind. For example, parents are much more likely to live within a few miles of their children than a thousand miles away or halfway around the world. So the model studied here is too simple to be directly applicable to the evolution of mankind as a whole. In such complicated situations, the results sound a note of caution: if the logarithmic time to CAs seems patently implausible, then at least one of the assumptions of the model, such as the random mating assumption, must be causing a great deal of trouble. On the other hand, it would be interesting to know whether there are simpler real-life situations in which the assumptions of the model do apply reasonably well and the theorems provide reasonably accurate quantitative descriptions. Perhaps a relatively homogeneous population lacking discernible structures (geographic or otherwise) that interact strongly with reproduction would be a promising candidate.

The random time analysed in Theorem 2 seems of natural interest in this process and may also be pertinent to certain questions about 'species trees' or 'population trees' (as opposed to 'gene trees'). In many contexts the species tree is considered to be the real object of interest, and we use genetic data and gene trees to attempt to learn about the species tree. For example, for humans, chimpanzees, and gorillas, is the 'true species tree' $(HC)G$, $(HG)C$, or $(CG)H$? Roughly, the conceptual framework of this question is as follows. There were two 'speciation events' that split a single species ancestral to humans, chimpanzees, and gorillas into the three separate modern species. The tree $(HC)G$, for example, says that the first such split separated the subpopulation that eventually became modern gorillas from the remainder, which later split to become modern humans and chimpanzees. Unfortunately, more precise definitions of the concept of species tree that remain useful in difficult or unclear cases seem hard to come by. One might adopt the viewpoint that the proper starting point for a definition of 'species tree' is the full two-parent genealogy of all present-day individuals. Given such a definition, if we knew all details of this genealogy, then we could read off an answer to the H, C, and G question (the answer might be 'none of the 3 choices above'—that is, the species tree is not well defined or at least not bifurcating). One interpretation of the time $\mathcal{U}_n$ is as follows. Suppose we imagine a case where evolution really proceeded according to a neat succession of 'speciation events'. Under a certain reasonable definition of a species tree, if the times between those speciation events exceed $\mathcal{U}_n$, then the species tree is guaranteed to be well defined and coincide with the history of speciation events. This idea will be discussed more fully elsewhere.

A caveat to forestall potential misunderstanding: this paper is not about genetics. That is, it is not about who gets what genes; it is about something more primitive, namely, the ancestor–descendant relationship. One-parent models are appropriate in tracing the history of a sample of nonrecombining genes or small bits of DNA; a single nucleotide descends from a single nucleotide from either the mother or father, but not both. Here we are considering ancestry in the more common, demographic sense of the word, as applied to people, for example, rather than genes.

Previous genetics research that is somewhat related, although still very different from the present study, considers models incorporating recombination. This type of model has been investigated in a number of papers, including those of Hudson (1983) and Griffiths and Marjoram (1997). The history of a sample of DNA sequences may be described by a collection of genealogies, with each nucleotide position in the DNA having its own one-parent genealogy. The genealogies for two positions that experience no recombination between them will be congruent, with the paths of the two genealogies going back through the same individuals, whereas a recombination between two positions causes the genealogies of those positions to differ. Each of the genealogies in the collection will have its own MRCA (the nucleotide at its root), which may occur in different individuals. Each of these individuals will be a CA in the sense considered in this paper, but the most recent of these individuals is generally not an MRCA in our sense. Our MRCA is more recent, since the paths from ancestors to descendants consist of all potential paths for genes to be transmitted, and may include paths that did not happen to be taken by any genes. No previous results about these genetic models have been similar to the results here, for example, in getting times of order $\log n$. This is not surprising, since the asymptotics would require an assumption that the sequence lengths and the number of recombinations tend to infinity. This is another manifestation of the statement that the questions we are investigating here are not fundamentally genetics questions.

There is some previous work on the process we study here and related processes. Two papers of Kämmerle (1989, 1991) introduce a general class of two-parent (called 'bisexual' in those papers) versions of the Wright–Fisher and other processes. These papers focus on two main questions. First, they analyse the probability of extinction of a set of individuals in the present generation, that is, the probability that the set of individuals eventually has no descendants in some future generation. Second, in a two-parent version of the Moran model, they study the number $R_n(t)$ of individuals t generations ago who have at least one descendant in the present generation. Kämmerle (1989) finds that the Markov chain $\{R_n(t) : t = 0, 1, \dots \}$, suitably normalized and suitably initialized (with the initialization essentially requiring that the chain is started in steady state), converges weakly as $n \to \infty$ to a discrete-time Ornstein–Uhlenbeck process.

Möhle (1994) both generalizes and refines the results of Kämmerle. In particular, Möhle provides a detailed analysis of the extinction probabilities in a two-parent Wright–Fisher model that approximates the probabilities up to $o(1/n)$. He also establishes weak convergence in a general class of two-parent models, including an Ornstein–Uhlenbeck limit for the two-parent Wright–Fisher process. Möhle also has a number of other papers in press, including one that relaxes the assumption of constant population size.

These previous results are complementary to the results in this paper. The previous papers considered individuals who have at least one descendant in a given future generation. Here we consider CAs, who have as descendants all members of the future generation. The previous results about the process $\{R_n(t)\}$ apply to large t, that is, to the behavior of the process many generations before the present, with the process in steady state. Here we focus on the behavior of a related process at small (i.e. recent) times, starting far away from steady state. We show that at about $t = 1.77 \lg n$ generations before the present, with high probability the $R_n(t)$ individuals who have at least one present-day descendant are all in fact CAs.

2. Simulations

Table 1 presents a small simulation study consisting of 25 trials each for $n = 500$, $n = 1000$, $n = 2000$, and $n = 4000$. Two numbers are reported for each trial: $\mathcal{T}_n$, the number of generations back to an MRCA, and $\mathcal{U}_n$, the generation at which every individual is either a CA or extinct.

In these simulations the distribution of the time back to an MRCA is indeed quite concentrated around the value $\lg n$, which is nearly 9 for $n = 500$, nearly 10 for $n = 1000$, and so on. Thus, the simulation results show that the asymptotic ($n \to \infty$) statement of Theorem 1 is 'not so asymptotic', in that it describes the situation well even for rather small values of n. The behavior predicted by Theorem 2 is also reflected reasonably well in the simulations, although one might have guessed a numerical constant closer to 2 rather than 1.77 from this small study.

3. Proofs

3.1. General ideas and tools

We start with the observation that although Theorems 1 and 2 are phrased in terms of counting generations back in time from the present until some condition obtains, these results may be proved by counting forward in time from a fixed generation. For example, the event $\{\mathcal{T}_n \leq m\}$ requires that a CA of all individuals in generation 0 may be found among generations $-1, -2, \dots, -m$. This is equivalent to requiring that if we start with generation $-m$ and trace forward in time, then some individual in generation $-m$ becomes a CA of all individuals in some generation $t \in \{-m + 1, -m + 2, \dots, 0\}$.

TABLE 1: A small simulation study. For each of four population sizes n, the two times $\mathcal{T}_n$ and $\mathcal{U}_n$ are reported for 25 trials.

	$n = 500$		$n = 1000$		$n = 2000$		$n = 4000$	
1	10	18	11	19	12	24	13	24
2	9	18	11	21	12	22	13	23
3	10	21	10	20	11	23	12	24
4	9	18	11	20	12	22	13	24
5	10	19	10	20	12	21	13	24
6	9	19	11	31	12	23	13	24
7	10	19	11	20	12	21	13	24
8	10	21	10	23	12	27	13	24
9	9	19	11	20	11	24	13	24
10	9	17	11	20	12	24	13	31
11	9	19	10	21	12	26	13	23
12	9	18	11	21	11	22	13	23
13	9	19	11	21	12	21	13	25
14	9	20	11	20	12	25	13	22
15	9	19	10	20	11	24	13	24
16	10	21	10	21	11	23	13	23
17	9	19	11	19	12	24	13	23
18	9	17	10	26	12	22	13	24
19	9	19	11	21	11	23	13	25
20	10	18	10	21	12	22	13	25
21	9	19	11	19	12	22	12	25
22	10	19	11	26	12	22	13	26
23	9	19	10	20	12	24	13	27
24	10	17	11	21	12	22	13	26
25	10	19	11	23	12	23	13	25

So we will count generations forward in time, and for convenience let us renumber generations so that the initial generation is 'generation 0'. The population at generation $t \geq 0$ consists of n individuals denoted by $I_{t,1}, I_{t,2}, \ldots, I_{t,n}$. We can picture $I_{t,1}, I_{t,2}, \ldots, I_{t,n}$ as dots in an array as in Figure 1, with $I_{t,j}$ being the jth dot in row t. The association of a number j to individual $I_{t,j}$ is an arbitrary labeling of the individuals within generation t. Assigned only as a means of referring to individuals, the labels have no significance in the model, which does not order the individuals within a generation. Let $\mu_{t,1}, \nu_{t,1}, \mu_{t,2}, \nu_{t,2}, \ldots, \mu_{t,n}, \nu_{t,n}$ be independent and uniformly distributed on the set $\{1, \ldots, n\}$. We interpret $\mu_{t,j}$ and $\nu_{t,j}$ as labels of the parents of individual $I_{t,j}$; that is, the parents of $I_{t,j}$ are $I_{t-1,\mu_{t,j}}$ and $I_{t-1,\nu_{t,j}}$. Defining a sequence of random sets $\mathcal{G}_0^i, \mathcal{G}_1^i, \ldots$ recursively by $\mathcal{G}_0^i = \{i\}$ and

$$\mathcal{G}_t^i = \{j \leq n : \mu_{t,j} \in \mathcal{G}_{t-1}^i \text{ or } \nu_{t,j} \in \mathcal{G}_{t-1}^i\},$$

$\mathcal{G}_t^i$ is the set of labels of the descendants of $I_{0,i}$ in generation t. Let G_t^i denote the cardinality of $\mathcal{G}_t^i$. The conditional probability that individual $I_{t+1,j}$ has at least one parent among the G_t^i members of $\mathcal{G}_t^i$ is

$$P(\{\mu_{t+1,j} \in \mathcal{G}_t^i\} \cup \{\nu_{t+1,j} \in \mathcal{G}_t^i\} \mid \mathcal{G}_t^i) = (G_t^i/n) + (G_t^i/n) - (G_t^i/n)(G_t^i/n).$$

The process $\{G_t^i : t = 0, 1, \dots\}$ is a Markov chain with transition probabilities

$$(G_{t+1}^i \mid G_t^i) \sim \text{Bin}\left(n, \frac{2G_t^i}{n} - \left(\frac{G_t^i}{n}\right)^2\right), \tag{1}$$

where $\text{Bin}(n, p)$ denotes the binomial distribution for the number of successes in n independent trials each having success probability p.

Throughout the proof, $\{G_t\}$ will denote a Markov chain with transition probabilities as in (1), although in different parts of the proof we will consider different possible initial values G_0. For example, taking $G_0 = 1$ corresponds to following the descendants of a particular individual in generation 0. In the early stages of the process, while G_t remains small relative to n, in view of (1) the conditional distribution of G_{t+1} given G_t is nearly Poisson$(2G_t)$, that is, the Poisson distribution with mean $2G_t$. In other words, while $\{G_t\}$ remains small, it evolves nearly as a Galton–Watson branching process $\{Y_t\}$ with offspring distribution Poisson(2). Kämmerle (1991) gave a formal statement of a result of this nature. A special case of his result says that for fixed u, the joint distribution of $(G_0, G_1, \dots, G_u)$ converges to that of $(Y_0, Y_1, \dots, Y_u)$ as $n \to \infty$. For our purposes, we will use the following result that allows us to approximate probabilities for the G process by those for the Y process up to a higher order of accuracy and over longer intervals of time that may have random lengths.

Lemma 3. *Let $Y_0, Y_1, \dots$ denote a Galton–Watson branching process with offspring distribution Poisson(2). Suppose that $Y_0 = G_0 = 1$. Define $\tau_b^Y = \inf\{t : Y_t \geq b\}$ and $\tau_{0b}^Y = \inf\{t : Y_t = 0 \text{ or } Y_t \geq b\}$, with corresponding definitions for τ_b^G and τ_{0b}^G. As $n \to \infty$, if m and b satisfy $mb^2 = o(n)$, then*

$$P\{\tau_b^G > m\} = P\{\tau_b^Y > m\}(1 + o(1)) \tag{2}$$

and

$$P\{\tau_{0b}^G > m\} = P\{\tau_{0b}^Y > m\}(1 + o(1)). \tag{3}$$

Proof. A straightforward calculation bounds the likelihood ratio

$$\begin{aligned} L(y \mid x) :&= \frac{P\{G_{t+1} = y \mid G_t = x\}}{P\{Y_{t+1} = y \mid Y_t = x\}} \\ &= \frac{P\{\text{Bin}(n, (2x/n) - (x^2/n^2)) = y\}}{P\{\text{Poisson}(2x) = y\}} \leq e^{2x}(1 - (2x/n) + (x^2/n^2))^{n-y}, \end{aligned}$$

so that

$$\log L(y \mid x) \leq 2x + (n - y)\left(-\frac{2x}{n} + \frac{x^2}{n^2}\right) \leq (x^2 + 2xy)/n.$$

This holds whenever the denominator $P\{Y_{t+1} = y \mid Y_t = x\}$ is positive, that is, for all $x > 0$ and $y \geq 0$, and also for $x = y = 0$. Thus, for all such pairs of x and y satisfying $x < b$ and $y < b$, we have

$$\log L(y \mid x) \leq 3b^2/n.$$

A similar calculation gives the lower bound

$$\log L(y \mid x) \geq -5b^2/(2n)[1 + O(b/n)],$$

so that $\log L(y \mid x) \geq -3b^2/n$ for sufficiently large n. So if $x_1, \ldots, x_m$ are all less than b, then

$$
\begin{aligned}
P\{G_1 = x_1, \ldots, G_m = x_m\} &= P\{G_1 = x_1 \mid G_0 = 1\} \cdots P\{G_m = x_m \mid G_{m-1} = x_{m-1}\} \\
&= P\{Y_1 = x_1, \ldots, Y_m = x_m\} L(x_1 \mid 1) \cdots L(x_m \mid x_{m-1}) \\
&\leq P\{Y_1 = x_1, \ldots, Y_m = x_m\} e^{3mb^2/n}
\end{aligned}
$$

and

$$
P\{G_1 = x_1, \ldots, G_m = x_m\} \geq P\{Y_1 = x_1, \ldots, Y_m = x_m\} e^{-3mb^2/n}.
$$

Thus,

$$
\begin{aligned}
P\{\tau_b^G > m\} &= \sum_{0 \leq x_1 < b} \cdots \sum_{0 \leq x_m < b} P\{G_1 = x_1, \ldots, G_m = x_m\} \\
&\leq \sum_{0 \leq x_1 < b} \cdots \sum_{0 \leq x_m < b} P\{Y_1 = x_1, \ldots, Y_m = x_m\} e^{3mb^2/n} \\
&= P\{\tau_b^Y > m\} e^{3mb^2/n}
\end{aligned}
\tag{4}
$$

and, similarly, $P\{\tau_b^G > m\} \geq P\{\tau_b^Y > m\} e^{-3mb^2/n}$, so that, by the assumption that $mb^2 = o(n)$, we obtain $P\{\tau_b^G > m\} = P\{\tau_b^Y > m\}(1 + o(1))$. This proves (2). The proof of (3) uses the same reasoning, with the summations in (4) ranging over $0 < x_t < b$ rather than $0 \leq x_t < b$.

The previous result will be useful because the Poisson Galton–Watson process is simple and well understood. The next lemma records a few well known items for future reference.

Lemma 4. *Let $Y_0, Y_1, \ldots$ denote a Galton–Watson process with offspring distribution Poisson(2). Define the moment generating function $\psi(z) = E(z^{Y_1}) = e^{-2+2z}$. The extinction probability $\rho = P\{Y_t = 0 \text{ for some } t\} \approx 0.20319$ is the smaller of the two solutions of $\psi(\rho) = \rho$, and $\rho = \gamma/2$, where γ is as defined in Theorem 2. The t-fold composition $\psi_t = \psi \circ \cdots \circ \psi$ satisfies $\psi_t(z) \uparrow \rho$ for all $0 \leq z \leq \rho$.*

The relation $\rho = \gamma/2$ is confirmed by comparing the definitions of ρ and γ. Despite the simple relationship, we will keep the two different letters in our notation for conceptual clarity.

Defining $g_t = G_t/n$, we have

$$
E(g_{t+1} \mid g_t) = 2g_t - g_t^2 = g_t(2 - g_t).
\tag{5}
$$

That is, if the fraction of descendants of a given individual is currently g_t, it is expected to multiply by a factor of $2 - g_t$ in the next generation. For example, in the early stages of the process when the fraction g_t is small, it nearly doubles in expectation in the next generation. For very small g_t (of the order $1/n$, for example) the random variability is large; for example, the process could easily go extinct. This is when it is most useful to approximate the G process by the Poisson(2) Galton–Watson process.

On the other hand, for larger values of g_t, the multiplication factor g_{t+1}/g_t, although expected to be somewhat smaller, has much less variability. The deviations of this factor from its expected value are bounded probabilistically by large deviations inequalities for the binomial distribution. We will use the following inequality of Bernstein (1946) as a basic tool.

Lemma 5. *(Bernstein's inequality.) If* $X \sim \mathrm{Bin}(n, p)$ *and* $r > 0$*, then*

$$P\{X \geq np + r\} \leq \exp\left\{\frac{-r^2}{2np(1 - p) + (2/3)r}\right\}. \tag{6}$$

Since $n - X \sim \mathrm{Bin}(n, 1 - p)$, the right-hand side of (6) is also an upper bound for the probability $P\{X \leq np - r\}$.

3.2. Proof of Theorem 1

Outline. The proof will be divided into several parts. We start from generation 0 and trace forward in time.

Stage 1: By the end of Stage 1, we identify an individual I in generation 0 who has a number of descendants that is small compared to n, but large enough so that I is unlikely ever to become extinct. In particular, we look for a generation t such that some individual I in generation 0 has at least $\lg^2(n)$ descendants in generation t. With probability approaching 1, this happens in time $o(\lg n)$, negligible compared with $\lg n$; this is shown by using Lemma 3 to approximate our process by a Poisson Galton–Watson process. The rest of the proof will show that with probability approaching 1, individual I becomes a CA within $(1 + \epsilon)(\lg n)$ generations, where ϵ is an arbitrary positive number.

Stage 2: Let $\beta \in (0, 1)$. Stage 2 follows the descendants of I until reaching a generation containing at least n^β descendants. In view of (5), since n^β is a small fraction of n for large n, throughout Stage 2 the number of descendants in a generation is expected to be nearly double the number of descendants in the previous generation. And $\lg^2(n)$ is large enough so that the multiplication factor will be very close to its expected value, with high probability. So Stage 2 should not take much more than about $\lg(n^\beta) = \beta \lg(n)$ generations.

Stage 3: This stage brings the count of descendants of I up from n^β to $(1/2)n$. Since the fraction of descendants during Stage 3 stays below $\frac{1}{2}$, the expected multiplication factor is at least $2 - \frac{1}{2} = \frac{3}{2}$. Again, this multiplication factor is very reliable, so that with high probability Stage 3 takes no more than about $\log_{3/2}\{(n/2)/(n^\beta)\}$ generations. We can make this an arbitrarily small fraction of $\lg n$ by choosing β close enough to 1.

Stage 4: Now we switch to looking at the fraction B_t of individuals in a generation who are *not* descendants of individual I. This fraction is expected to square each generation. This causes B_t to decrease very quickly. Fixing $\alpha \in (\frac{1}{2}, \frac{2}{3})$, we show that Stage 4, which takes the fraction B_t from $\frac{1}{2}$ down to $n^{-\alpha}$, takes only order $\lg \lg n$ time.

Stage 5: This completes the process, ending when the B process hits 0, and individual I has become a CA. We show that this takes just one generation with high probability.

Upper bound: Combining the results of Stages 1 through 5 gives the probabilistic upper bound $\lim_{n \to \infty} P\{\mathcal{T}_n \leq (1 + \epsilon) \lg n\} = 1$.

Lower bound: Here we show that $\lim_{n \to \infty} P\{\mathcal{T}_n \geq (1 - \epsilon) \lg n\} = 1$. This is done by using Bernstein's inequality to prove an assertion of the following form: for positive r and δ, once the process of descendants of any given individual reaches a power n^r of n, it is very unlikely to increase by a factor of more than $2 + \delta$ in a generation, whereas it would have to do so in order to have $\mathcal{T}_n < (1 - \epsilon) \lg n$.

3.2.1. *Stage 1.* Here we will show that with high probability, within a number of generations negligible compared to $\lg n$, we can find a generation with at least $\lg^2 n$ individuals who share a common ancestor. For simplicity we give a crude argument that circumvents the need to

consider any dependence among the processes $\{G_t^i : t \geq 0\}$ starting from different individuals $I_{0,i}$. This could also be done along the lines of the argument in Lemma 19 below, where we need to confront this dependence.

Lemma 6. *Define* $\tau_b = \inf\{t : G_t \geq b\}$. *Assuming that* $G_0 = 1$,

$$\liminf_{n \to \infty} P\{\tau_{\lg^2 n} \leq 3 \lg \lg n\} > 0.$$

Proof. Let b and m denote $\lg^2 n$ and $\lfloor 3 \lg \lg n \rfloor$, respectively. Let $\{Y_t\}$ be a Galton–Watson process with offspring distribution Poisson(2), and define $M_t = Y_t 2^{-t}$. The process $\{M_t\}$ is a non-negative martingale that converges almost surely to a limit M_∞, say, with $P\{M_\infty = 0\} = \rho < 1$. Note that $P\{\tau_b^Y > m\} \leq P\{Y_m < b\} = P\{M_m < b2^{-m}\}$. Therefore, using Fatou's lemma and the assumption that $b2^{-m} \to 0$,

$$\limsup P\{\tau_b^Y > m\} \leq \limsup P\{M_m < b2^{-m}\} \leq P(\limsup\{M_m < b2^{-m}\})$$
$$= P\{M_m < b2^{-m} \text{ infinitely often}\} \leq P\{M_\infty = 0\} = \rho < 1.$$

By Lemma 3, $P\{\tau_b > m\} = P\{\tau_b^Y > m\}(1 + o(1))$ as $n \to \infty$. Therefore,

$$\limsup P\{\tau_b > m\} \leq \limsup P\{\tau_b^Y > m\} \leq \rho < 1.$$

So $\liminf P\{\tau_b \leq m\} \geq 1 - \rho > 0$.

Proposition 7. *Let* G_t^i *denote the number of descendants in generation t of individual* $I_{0,i}$ *(the ith individual in generation 0), and let* $G_t^* = \max_{1 \leq i \leq n}\{G_t^i\}$. *Define* $\tau_b^{G^*} = \inf\{t : G_t^* \geq b\}$. *Then* $\tau_{\lg^2 n}^{G^*} = o_P(\lg n)$.

Proof. We use a geometric trials argument. Let $m_n = \lfloor 3 \lg \lg n \rfloor$, and choose a sequence $\{k_n\}$ with $k_n \to \infty$ and $k_n m_n = o(\lg n)$. Perform a sequence of k_n trials as follows. For the first trial, start with individual $I_{0,1}$, and follow his progeny for m_n generations. We say the trial is a success if $I_{0,1}$ has at least $\lg^2 n$ descendants in generation m_n; by Lemma 6 this happens with probability at least c, say, where $c > 0$. If the trial is a failure, start a new trial, following the progeny of individual $I_{m_n,1}$ for m_n more generations. And so on. We stop at the first success, having found an individual with at least $\lg^2 n$ descendants. The probability that this sequence of trials fails to terminate by generation $k_n m_n$ is at most $(1 - c)^{k_n}$, which tends to 0.

Thus, with probability tending to 1, there is a $\kappa \in \{0, \ldots, k_n - 1\}$ such that individual $I_{\kappa m_n, 1}$ has at least $\lg^2 n$ descendants in generation $(\kappa + 1)m_n$. Let I denote any ancestor of $I_{\kappa m_n, 1}$ in generation 0. We will show in the remainder of the proof that for each $\epsilon > 0$, with probability tending to 1 as $n \to \infty$, individual I becomes a CA within $(1 + \epsilon) \lg n$ generations.

3.2.2. *Stage 2.* The following simple consequence of Bernstein's inequality will be a convenient tool.

Lemma 8. *If* $\delta \leq 3/4$ *and* $G_t \leq \delta n/20$, *then*

$$P\{G_{t+1} \leq (2 - \delta)G_t \mid G_t\} \leq \exp(-\delta^2 G_t/5).$$

The next result shows that the probability that Stage 2 takes more than $\lg n$ generations approaches 0 as $n \to \infty$. In fact, we show that this probability is $o(1/n)$; this will be used in the proof of Theorem 2.

Proposition 9. *Assume that* $G_0 \geq \lg^2 n$, *and let* $0 < \beta < 1$. *Define* $T_2 = \inf\{t : G_t \geq n^\beta\}$. *Then* $P\{T_2 > \lg n\} = o(1/n)$ *as* $n \to \infty$.

Proof. Take $0 < \delta < 3/4$ such that $\lg(2 - \delta) > \beta$, and define

$$b(n) = \left\lceil \log_{2-\delta}\left(\frac{n^\beta}{\lg^2 n}\right) \right\rceil.$$

Note that

$$b(n) \leq \frac{\beta \lg n}{\lg(2 - \delta)} \leq \lg n,$$

at least for $n \geq 3$, so that $P\{T_2 > \lg n\} \leq P\{T_2 > b(n)\}$. We will show that $P\{T_2 > b(n)\} = o(1/n)$.

The inequality $T_2 > b(n)$ implies that $G_{t+1} < (2 - \delta)G_t$ for some $0 \leq t \leq b(n) - 1$. The first such t must also satisfy $G_t \geq \lg^2 n$. Thus,

$$P\{T_2 > b(n)\} \leq P\left\{ \bigcup_{t=0}^{b(n)-1} \{G_{t+1} < (2 - \delta)G_t, \; G_t \geq \lg^2 n, \; T_2 > b(n)\} \right\}$$

$$\leq \sum_{t=0}^{b(n)-1} P\{G_{t+1} < (2 - \delta)G_t, \; \lg^2 n \leq G_t \leq n^\beta\}.$$

However, $n^\beta \leq \delta n/20$ for sufficiently large n. Therefore, on the event $\{\lg^2 n \leq G_t \leq n^\beta\}$, we may apply Lemma 8 to obtain

$$P\{G_{t+1} < (2 - \delta)G_t \mid G_t\} \leq \exp\left(-\frac{\delta^2}{5}\lg^2 n\right) = n^{-(\delta^2/5)(\lg e)(\lg n)}.$$

Thus,

$$P\{T_2 > b(n)\} \leq b(n)n^{-(\delta^2/5)(\lg e)(\lg n)} = o(1/n) \quad \text{as } n \to \infty.$$

3.2.3. Stage 3. This stage starts in a generation in which the number of descendants of I is just over n^β and ends when the number of descendants in a generation reaches $\frac{1}{2}n$. Defining $g_t = G_t/n$, we have $E(g_{t+1} \mid g_t) = g_t(2 - g_t)$. The idea is that if $g_t \leq \frac{1}{2}$, then in the next generation g_t is expected to multiply by a factor of $2 - g_t \geq \frac{3}{2}$. So with high probability, throughout Stage 3, at each generation the number of descendants will multiply by at least $\sqrt{2}$, say, since $\sqrt{2} < \frac{3}{2}$. So to get from n^β to $\frac{1}{2}n$, we should need at most $\log_{\sqrt{2}}(\frac{1}{2})n^{1-\beta} = 2[(1 - \beta)\lg n - 1]$ generations.

Proposition 10. *Assume* $G_0 \geq n^\beta$, *and define* $T_3 = \inf\{t : G_t \geq (\frac{1}{2})n\}$. *Then* $P\{T_3 > 2(1 - \beta)\lg n\} = o(\frac{1}{n})$ *as* $n \to \infty$.

Proof. The proof is similar to that of Proposition 9. For $n^\beta \leq G_t \leq n/2$, a straightforward calculation using Bernstein's inequality gives

$$P\{G_{t+1} \leq \sqrt{2}G_t \mid G_t\} \leq \exp(-0.001G_t) \leq \exp(-0.001n^\beta).$$

Note that $\log_{\sqrt{2}}\{(n/2)/n^\beta\} = 2(1 - \beta)\lg n - 2$. So if $T_3 > 2(1 - \beta)\lg n$, then we must have $G_{t+1} \leq \sqrt{2}G_t$ for some $t < 2(1 - \beta)\lg n$ satisfying $n^\beta \leq G_t \leq n/2$. Thus,

$$P\{T_3 > 2(1 - \beta)\lg n\} \leq 2(1 - \beta)(\lg n)\exp(-0.001n^\beta) = o(1/n) \quad \text{as } n \to \infty.$$

3.2.4. *Stage 4.* Let B_t denote $1 - G_t/n$, the *fraction* of individuals in generation t who are *not* descendants of the chosen individual I. Then

$$(B_{t+1} \mid B_t, B_{t-1}, \ldots) \sim \frac{1}{n} \operatorname{Bin}(n, B_t^2), \tag{7}$$

since an individual is not a descendant of I when both of his parents fail to be descendants of I. Fix $\alpha \in (\frac{1}{2}, \frac{2}{3})$. Stage 4 takes the B_t process from $\frac{1}{2}$ down to $n^{-\alpha}$. The idea is this. Since $E(B_{t+1} \mid B_t) = B_t^2$, we expect B_t to square each generation. We will show that the probability $P\{B_{t+1} \geq B_t^{3/2}\}$ is small throughout Stage 4 (note $\frac{3}{2} < 2$). This will be good enough, since if $B_{t+1} < B_t^{3/2}$ holds throughout Stage 4, then Stage 4 is completed in order $\lg\lg n$ time.

Proposition 11. *Consider a process* $B_0, B_1, \ldots$ *satisfying* (7), *and suppose* $B_0 \leq \frac{1}{2}$. *Let* $\alpha \in (\frac{1}{2}, \frac{2}{3})$ *and define* $T_4 = \inf\{t : B_t \leq n^{-\alpha}\}$. *Then* $P\{T_4 \geq 2\lg\lg n\} = o(1/n)$ *as* $n \to \infty$.

Proof. By Bernstein's inequality,

$$P\{B_{t+1} \geq B_t^{3/2} \mid B_t\} = P\{\operatorname{Bin}(n, B_t^2) \geq nB_t^{3/2} \mid B_t\}$$

$$\leq \exp\left\{\frac{-n^2 B_t^3(1 - B_t^{1/2})^2}{2nB_t^2(1 - B_t^2) + (2/3)nB_t^{3/2}(1 - B_t^{1/2})}\right\}$$

$$= \exp\left\{\frac{-nB_t(1 - B_t^{1/2})^2}{2(1 - B_t^2) + (2/3)B_t^{-1/2}(1 - B_t^{1/2})}\right\}.$$

If $n^{-\alpha} \leq B_t \leq \frac{1}{2}$, then $(1 - B_t^{1/2})^2 \geq 1.5 - \sqrt{2} \geq 0.08$, and

$$\frac{nB_t(1 - B_t^{1/2})^2}{2(1 - B_t^2) + (2/3)B_t^{-1/2}(1 - B_t^{1/2})} \geq \frac{0.08n^{1-\alpha}}{2 + (2/3)n^{\alpha/2}} \geq 0.08n^{1-(3/2)\alpha}$$

(the last inequality holding for $n \geq 6^{2/\alpha}$), so that

$$P\{B_{t+1} \geq B_t^{3/2} \mid B_t\} \leq \exp\{-0.08n^{1-(3/2)\alpha}\}.$$

For $n \geq 2$, if $B_0 \leq \frac{1}{2}$ and $B_{t+1} \leq B_t^{3/2}$ for $t = 0, 1, \ldots, \lceil 2\lg\lg n\rceil - 1$, then $B_{\lceil 2\lg\lg n\rceil} \leq n^{-1} \leq n^{-\alpha}$. Therefore,

$$\{T_4 > \lceil 2\lg\lg n\rceil\} \subseteq \{B_{\lceil 2\lg\lg n\rceil} > n^{-\alpha}\}$$

$$\subseteq \bigcup_{t=0}^{\lceil 2\lg\lg n\rceil - 1} \{B_{t+1} > B_t^{3/2}, n^{-\alpha} < B_t \leq 1/2\},$$

so that

$$P\{T_4 > \lceil 2 \lg \lg n \rceil\} \leq \sum_{t=0}^{\lceil 2 \lg \lg n \rceil - 1} P\{B_{t+1} \geq B_t^{3/2}, n^{-\alpha} < B_t \leq 1/2\}$$

$$\leq \lceil 2 \lg \lg n \rceil \exp\{-0.08 n^{1-(3/2)\alpha}\} = o(1/n).$$

3.2.5. Stage 5. This stage starts with the $\{B_t\}$ process below $n^{-\alpha}$ and ends when it hits 0. We show that with high probability this takes just one generation.

Proposition 12. *Suppose $B_0 \leq n^{-\alpha}$. Then $P\{B_1 = 0\} \to 1$ as $n \to \infty$.*

Proof. Since $B_1 \sim (1/n)\mathrm{Bin}(n, B_0^2)$ and $2\alpha > 1$, we have $P\{B_1 = 0\} = (1 - B_0^2)^n \geq (1 - n^{-2\alpha})^n \to 1$. ∎

3.2.6. Upper bound.

Proposition 13. *For each $\epsilon > 0$, $P\{\mathcal{T}_n > (1 + \epsilon) \lg n\} \to 0$ as $n \to \infty$.*

Proof. Define T_1 to be the time at which Stage 1 ends. Then we know that T_1 is finite with probability 1, and, for arbitrary positive ξ, $P\{T_1 > \xi \lg n\} \to 0$ as $n \to \infty$. At the end of Stage 1 we have found an individual I, say, in generation 0 who has at least $\lg^2(n)$ descendants in generation T_1. Let G_t denote the number of descendants of I in generation t, and let $\tau(b)$ denote $\inf\{t : G_t \geq b\}$. Our previous results have shown that

$$P\{\tau(n^\beta) - T_1 > \lg n\} = o(1/n),$$
$$P\{\tau(n/2) - \tau(n^\beta) > 2(1 - \beta) \lg n \mid \tau(n^\beta) < \infty\} = o(1/n),$$
$$P\{\tau(n - n^{1-\alpha}) - \tau(n/2) > 2 \lg \lg n \mid \tau(n/2) < \infty\} = o(1/n),$$
$$P\{\tau(n) - \tau(n - n^{1-\alpha}) > 1 \mid \tau(n - n^{1-\alpha}) < \infty\} = o(1).$$

Thus,

$$P\{\mathcal{T}_n > \xi \lg n + \lg n + 2(1 - \beta) \lg n + 2 \lg \lg n + 1\}$$
$$\leq P\{T_1 > \xi \lg n\} + P\{T_1 < \infty, \tau(n^\beta) - T_1 > \lg n\}$$
$$+ P\{\tau(n^\beta) < \infty, \tau(n/2) - \tau(n^\beta) > 2(1 - \beta) \lg n\}$$
$$+ P\{\tau(n/2) < \infty, \tau(n - n^{1-\alpha}) - \tau(n/2) > 2 \lg \lg n\}$$
$$+ P\{\tau(n - n^{1-\alpha}) < \infty, \tau(n) - \tau(n - n^{1-\alpha}) > 1\}$$
$$= o(1) + o(1/n) + o(1/n) + o(1/n) + o(1) = o(1).$$

Given $\epsilon > 0$, taking ξ and β such that $\xi + 2(1 - \beta) < \epsilon$, we see that $P\{\mathcal{T}_n > (1 + \epsilon) \lg n\} \to 0$. ∎

3.2.7. Lower bound. We will use Bernstein's inequality in the following form.

Lemma 14. *For $\delta \leq \frac{3}{2}$, $P\{G_{t+1} \geq (2 + \delta)G_t \mid G_t\} \leq \exp[-\delta^2 G_t/5]$.*

Proposition 15. *For each $\epsilon > 0$, $P\{\mathcal{T}_n < (1 - \epsilon) \lg n\} \to 0$.*

Proof. Fix $\epsilon \in (0, 1)$. Proceeding forward in time from generation 0, we want to show that the probability that none of the individuals in generation 0 becomes a CA before generation $\lfloor (1 - \epsilon) \lg n \rfloor$ tends to 1 as $n \to \infty$. Define $G_0 = 1$ and $(G_{t+1} \mid G_t, \ldots, G_0) \sim$ $\mathrm{Bin}(n, 2G_t/n - (G_t/n)^2)$. Here we think of G_t as the number of descendants of individual $I_{0,1}$ in generation t. Fix $r \in (0, \epsilon)$ so that $2^{(1-r)/(1-\epsilon)} \in (2, 3.5)$. Let $\{\tilde{G}_t\}$ evolve like $\{G_t\}$ except that it is truncated (or 'reflected') below at the value $\lfloor n^r \rfloor$. That is,

$$(\tilde{G}_{t+1} \mid \tilde{G}_t, \ldots, \tilde{G}_0) \sim \max \left\{ \mathrm{Bin}\left(n, \frac{2\tilde{G}_t}{n} - \left(\frac{\tilde{G}_t}{n}\right)^2\right), \lfloor n^r \rfloor \right\}.$$

Defining $\tau_n^G = \inf\{t : G_t = n\}$ and $\tau_n^{\tilde{G}} = \inf\{t : \tilde{G}_t = n\}$, obviously $\mathrm{P}\{\tau_n^G \geq u\} \geq \mathrm{P}\{\tau_n^{\tilde{G}} \geq u\}$ for all u. Since $\tilde{G}_0 = \lfloor n^r \rfloor$, if $\tau_n^{\tilde{G}} \leq \lfloor (1 - \epsilon) \lg n \rfloor$, then we must have $\tilde{G}_{t+1} \geq \tilde{G}_t 2^{(1-r)/(1-\epsilon)}$ for some $t < \lfloor (1 - \epsilon) \lg n \rfloor$. Defining $\delta = 2^{(1-r)/(1-\epsilon)} - 2 \in (0, \frac{3}{2})$, by Lemma 14 the probability of this is at most

$$\lfloor (1 - \epsilon) \lg n \rfloor \exp(-\delta^2 \lfloor n^r \rfloor / 5),$$

which is $o(1/n)$ as $n \to \infty$. Thus, we have shown that the probability that individual $I_{0,1}$ has become a CA by generation $\lfloor (1 - \epsilon) \lg n \rfloor$ is $o(1/n)$. So the event that at least one of the n individuals in generation 0 becomes a CA by generation $\lfloor (1 - \epsilon) \lg n \rfloor$ is a union of n such events of probability $o(1/n)$, and hence has probability that tends to 0 as $n \to \infty$.

3.3. Proof of Theorem 2

Idea. The idea of the proof is as follows. Define $t_n = \lfloor (\zeta - \epsilon) \lg n \rfloor$ and $u_n = \lfloor (\zeta + \epsilon) \lg n \rfloor$. For each $i = 1, \ldots, n$, the process $\{G_t^i : t = 0, 1, \ldots\}$ follows the descendants of individual $I_{0,i}$. We are waiting until all n of the processes $\{G_t^1\}, \ldots, \{G_t^n\}$ have reached either 0 or n (some will reach 0 and some will reach n). The key ingredient of the argument is this assertion: with high probability, there are many i's such that $G_{t_n}^i \in [1, \lg^2(n)]$ and there is no i such that $G_{u_n}^i \in [1, \lg^2(n)]$. This follows from Lemma 3 together with an analysis of the Galton–Watson process with offspring distribution Poisson(2). For an upper bound, consider the situation at time u_n. Some of the processes have become extinct and reached 0, and we are just waiting for the other, non-extinct processes to reach 0 or n. The key assertion says that with high probability, all of the non-extinct processes have reached values above $\lg^2(n)$. This level is high enough so that with high probability these processes will all increase predictably and reach n within $(1 + \epsilon) \lg n$ additional generations; this was shown in the proof of Theorem 1. So with high probability, $\mathcal{U}_n \leq u_n + (1 + \epsilon) \lg n$. For a lower bound, the key assertion states that with high probability many of the n processes are in the interval $[1, \lg^2(n)]$ at time t_n. It is very unlikely that all of these will go extinct. Furthermore, since these processes are starting from at most $\lg^2(n)$ at time t_n, with high probability it will take more than $(1 - \epsilon) \lg n$ additional generations for any of them to reach n. So $\mathcal{U}_n > t_n + (1 - \epsilon) \lg n$ with high probability.

3.3.1. *A branching process result.*

Lemma 16. *Let $\{Y_t\}$ be a Galton–Watson process whose offspring distribution is Poisson with mean 2, starting at $Y_0 = 1$. Define γ as in Theorem 2, and let $b_1, b_2, \ldots$ be positive integers satisfying $\lg(b_t) = o(t)$ as $t \to \infty$. Then*

$$\lim_{t \to \infty} \frac{1}{t} \lg \mathrm{P}\{1 \leq Y_t \leq b_t\} = \lg(\gamma) \approx -1.29911.$$

Proof. We use a number of results from chapter 1 of Athreya and Ney (1972). First, the Monotone Ratio Lemma says that for each k there is a $\lambda_k < \infty$ such that

$$\frac{P\{Y_t = k\}}{P\{Y_t = 1\}} \uparrow \lambda_k \text{ as } t \to \infty.$$

Also,

$$\Lambda(s) := \sum_{k=1}^{\infty} \lambda_k s^k < \infty \quad \text{for all } s \in (0, 1).$$

Finally, using the notation and facts collected in Lemma 4, we have

$$P\{Y_t = 1\} = \psi'_t(0) = \psi'[\psi_{t-1}(0)]\psi'_{t-1}(0) = \psi'[\psi_{t-1}(0)]P\{Y_{t-1} = 1\},$$

so that

$$\frac{P\{Y_t = 1\}}{P\{Y_{t-1} = 1\}} \uparrow \psi'(\rho) = 2\rho = \gamma.$$

In particular, $(1/t) \lg P\{Y_t = 1\} \uparrow \lg \gamma$.

For $s \in (0, 1)$,

$$\sum_{k=1}^{b_t} P\{Y_t = k\} \leq P\{Y_t = 1\} \sum_{k=1}^{b_t} \lambda_k$$

$$\leq P\{Y_t = 1\}s^{-b_t} \sum_{k=1}^{b_t} \lambda_k s^k$$

$$\leq P\{Y_t = 1\}s^{-b_t} \Lambda(s). \tag{8}$$

If we take s close to 1 (e.g. $s = 1 - b_t^{-1}$, say), then the term s^{-b_t} will remain bounded and present no difficulty. So we would like to know how $\Lambda(s)$ grows as $s \uparrow 1$.

Define φ to be the inverse function ψ^{-1}, and φ_k to be the k-fold composition $\varphi \circ \cdots \circ \varphi$. By equation (6) on page 12 of Athreya and Ney (1972), for each $s \in (\rho, 1)$,

$$\Lambda(\varphi(s)) = \gamma^{-1}[\Lambda(s) - \Lambda(e^{-2})] \leq \gamma^{-1}\Lambda(s).$$

Therefore, since $\rho < \frac{1}{2} < \varphi(\frac{1}{2}) < \varphi_2(\frac{1}{2}) < \cdots < 1$,

$$\Lambda(\varphi_k(\tfrac{1}{2})) \leq \gamma^{-k}\Lambda(\tfrac{1}{2}).$$

However, since $\psi'(1) = 2$, we may choose a number Φ so that

$$\varphi_k(\tfrac{1}{2}) \geq 1 - (1.9)^{-k}\Phi$$

and, therefore,

$$\Lambda(1 - (1.9)^{-k}\Phi) \leq \gamma^{-k}\Lambda(\tfrac{1}{2})$$

hold for all sufficiently large k. From this, it follows that

$$\Lambda(1 - y) \leq \Lambda(\tfrac{1}{2})(y/\Phi)^{(\lg \gamma)/(\lg 1.9)} \leq y^{2\lg \gamma}$$

holds for all sufficiently small positive y.

Now substituting $s = 1 - b_t^{-1}$ in (8), there is a finite constant C such that

$$\sum_{k=1}^{b_t} P\{Y_t = k\} \leq C\gamma^t \Lambda(1 - b_t^{-1}) \leq C\gamma^t b_t^{2\lg(1/\gamma)}.$$

Thus, as long as b_t grows subgeometrically, that is, $\lg(b_t) = o(t)$, we have

$$\overline{\lim}_{t\to\infty} \frac{1}{t} \lg P\{1 \leq Y_t \leq b_t\} \leq \lg(\gamma).$$

Combining this with the fact that $\lim_{t\to\infty}(1/t) \lg P\{Y_t = 1\} = \lg(\gamma)$ completes the proof.

3.3.2. *Upper bound.*

Lemma 17. *Let $I_{0,i}$ denote individual i in generation 0. Define G_t^i to be the number of descendants of $I_{0,i}$ in generation t; in particular, $G_0^i = 1$ for all $i = 1, \ldots, n$. Also define*

$$\tau_{0,b}^i = \inf\{t : G_t^i = 0 \text{ or } G_t^i \geq b\},$$

and let

$$A_n = \bigcup_{i=1}^{n} \{\tau_{0,\lg^2 n}^i > (\zeta + \epsilon) \lg n\}. \tag{9}$$

Then $P(A_n) \to 0$ as $n \to \infty$.

Proof. Define $\tau_{0b}^Y = \inf\{t : Y_t = 0 \text{ or } Y_t \geq b\}$. Since $\{\tau_{0b}^Y > t\} \subseteq \{1 \leq Y_t < b\}$, Lemma 16 and (3) give

$$\overline{\lim} \frac{1}{t} \lg P\{\tau_{0b}^i > t\} \leq \lg(\gamma) \quad \text{if } \lg(b) = o(t) \text{ and } tb^2 = o(n). \tag{10}$$

Letting $\epsilon > 0$ and applying (10) to $t = (\zeta + \epsilon)(\lg n)$ and $b = \lg^2(n)$ gives

$$\lg P\{\tau_{0,\lg^2(n)}^i > (\zeta + \epsilon)(\lg n)\} \leq (\lg(\gamma) + \delta)(\zeta + \epsilon)(\lg n)$$

for all δ and all sufficiently large n. Taking δ sufficiently small, from the definition of ζ we see that

$$P\{\tau_{0,\lg^2(n)}^i > (\zeta + \epsilon)(\lg n)\} = o(1/n) \quad \text{as } n \to \infty,$$

so that $P(A_n) = o(1)$.

We have shown that, with high probability, all individuals in generation 0 have either no descendants or more than $\lg^2(n)$ descendants in generation $\lfloor(\zeta + \epsilon) \lg n\rfloor$ for $\epsilon > 0$. Next we will show that for any given $\epsilon > 0$, with high probability, each individual having more than $\lg^2(n)$ descendants in generation $\lfloor(\zeta + \epsilon) \lg n\rfloor$ becomes a CA within $(1 + \epsilon) \lg n$ additional generations. Most of the work required to prove this has already been done in the proof of Theorem 1; the extra ingredient is the following lemma, which takes a closer look at 'Stage 5'. We retain the definition $B_t = 1 - (G_t/n)$ from above.

Lemma 18. *Let $\alpha \in (\frac{1}{2}, \frac{2}{3})$ and take $k(\alpha) > 1/(2\alpha - 1)$. Suppose that $B_0 \leq n^{-\alpha}$ and define $T_5 = \inf\{t : B_t = 0\}$. Then $P\{T_5 > k(\alpha)\} = o(1/n)$ as $n \to \infty$.*

Proof. Since $(B_{t+1} \mid B_t) \sim (1/n)\mathrm{Bin}(n, B_t^2)$, on the event $\{B_t \le n^{-\alpha}\}$ we have

$$
\begin{aligned}
\mathrm{P}\{B_{t+1} > 0 \mid B_t\} &= 1 - (1 - B_t^2)^n \\
&\le 1 - (1 - 2nB_t^2) = 2nB_t^2 \le 2n^{1-2\alpha},
\end{aligned}
$$

where the first inequality holds for sufficiently large n (since $\alpha > \frac{1}{2}$ implies that nB_t^2 is arbitrarily small for sufficiently large n). In particular,

$$
\mathrm{P}\{0 < B_{t+1} \le n^{-\alpha} \mid 0 < B_t \le n^{-\alpha}\} \le 2n^{1-2\alpha} \quad \text{for sufficiently large } n. \tag{11}
$$

Next, by Bernstein's inequality, on the event $\{B_t \le n^{-\alpha}\}$,

$$
\begin{aligned}
\mathrm{P}\{B_{t+1} > n^{-\alpha} \mid B_t\} &= \mathrm{P}\left\{\frac{1}{n}\mathrm{Bin}(n, B_t^2) > B_t^2 + (n^{-\alpha} - B_t^2) \mid B_t\right\} \\
&\le \exp\left[\frac{-n^2(n^{-\alpha} - B_t^2)^2}{2nB_t^2(1 - B_t^2) + (2/3)n(n^{-\alpha} - B_t^2)}\right] \\
&\le \exp\left[\frac{-n^{2-2\alpha}}{2n^{1-2\alpha} + (2/3)n^{1-\alpha}}\right].
\end{aligned}
$$

Since the exponent is asymptotic to $-(\frac{3}{2})n^{1-\alpha}$, clearly

$$
\mathrm{P}\{B_{t+1} > n^{-\alpha} \mid B_t\} \le \exp[-n^{1-\alpha}] \quad \text{on} \quad \{B_t \le n^{-\alpha}\}
$$

for sufficiently large n. Assuming that $B_0 \le n^{-\alpha}$,

$$
\bigcup_{t=0}^{k}\{B_t > n^{-\alpha}\} \subseteq \bigcup_{t=0}^{k-1}\{B_t \le n^{-\alpha}, B_{t+1} > n^{-\alpha}\}.
$$

Therefore, since

$$
\mathrm{P}\{B_t \le n^{-\alpha}, B_{t+1} > n^{-\alpha}\} = \mathrm{E}[\{B_t \le n^{-\alpha}\}\mathrm{P}\{B_{t+1} > n^{-\alpha} \mid B_t\}] \le \exp[-n^{1-\alpha}],
$$

we obtain

$$
\mathrm{P}\left[\bigcup_{t=0}^{k}\{B_t > n^{-\alpha}\}\right] \le k\exp[-n^{1-\alpha}]. \tag{12}
$$

Thus, using (11) and (12),

$$
\begin{aligned}
\mathrm{P}\{T_5 > k\} &\le \mathrm{P}\left(\bigcup_{t=0}^{k}\{B_t > n^{-\alpha}\}\right) + \mathrm{P}\left(\bigcap_{t=0}^{k}\{0 < B_t \le n^{-\alpha}\}\right) \\
&\le k\exp[-n^{1-\alpha}] + (2n^{1-2\alpha})^k.
\end{aligned}
$$

Applying this to $k = k(\alpha) > 1/(2\alpha - 1)$ gives the desired result.

Proof of the upper bound in Theorem 2. Let $\mathcal{U}_n$ denote the time at which everyone from generation 0 has become a CA or extinct. Recall the definition of A_n from (9), and let $\tau^i(b) = \inf\{t : G_t^i \geq b\}$. Since

$$\{\mathcal{U}_n > (1 + \zeta + 2\epsilon)\lg n\} \subseteq A_n \cup [A_n^c \cap \{\mathcal{U}_n > (1 + \zeta + 2\epsilon)\lg n\}]$$

$$\subseteq A_n \cup \bigcup_{i=1}^n \{\tau^i(\lg^2 n) \leq (\zeta + \epsilon)\lg n, \tau^i(n) > (1 + \zeta + 2\epsilon)\lg n\},$$

to show that $P\{\mathcal{U}_n > (1 + \zeta + 2\epsilon)\lg n\} = o(1)$, by Lemma 17 it suffices to show that

$$P\{\tau^1(\lg^2 n) \leq (\zeta + \epsilon)\lg n, \tau^1(n) > (1 + \zeta + 2\epsilon)\lg n\} = o(1/n).$$

To see this, observe that the results of Stages 2 through 4 from the proof of Theorem 1 show that

$$P\{\tau^1(n^\beta) - \tau^1(\lg^2 n) > \lg n \mid \tau^1(\lg^2 n) < \infty\} = o(1/n),$$
$$P\{\tau^1(n/2) - \tau^1(n^\beta) > 2(1 - \beta)\lg n \mid \tau^1(n^\beta) < \infty\} = o(1/n),$$
$$P\{\tau^1(n - n^{1-\alpha}) - \tau^1(n/2) > 2\lg\lg n \mid \tau^1(n/2) < \infty\} = o(1/n),$$

and Lemma 18 gives

$$P\{\tau^1(n) - \tau^1(n - n^{1-\alpha}) > k(\alpha) \mid \tau^1(n - n^{1-\alpha}) < \infty\} = o(1/n).$$

Consequently,

$$P\{\tau^1(n) - \tau^1(\lg^2 n) > \lg n + 2(1 - \beta)\lg n + 2\lg\lg n + k(\alpha) \mid \tau^1(\lg^2 n) < \infty\} = o(1/n).$$

Choosing β sufficiently close to 1, we see that for any given $\epsilon > 0$,

$$P\{\tau^1(n) - \tau^1(\lg^2 n) > (1 + \epsilon)\lg n \mid \tau^1(\lg^2 n) < \infty\} = o(1/n).$$

Thus,

$$P\{\tau^1(\lg^2 n) \leq (\zeta + \epsilon)\lg n, \tau^1(n) > (1 + \zeta + 2\epsilon)\lg n\}$$
$$\leq P\{\tau^1(\lg^2 n) < \infty, \tau^1(n) - \tau^1(\lg^2 n) > (1 + \epsilon)\lg n\} = o(1/n),$$

as desired.

3.3.3. *Lower bound.* The proof goes as follows. First we show that at time $t_n = \lfloor(\zeta - \epsilon)\lg n\rfloor$, there are many individuals i who have $G_{t_n}^i \in [1, \lg^2 n]$. The probability that all of these individuals eventually become extinct is negligibly small. In the probable event that not all of these individuals become extinct, the time $\mathcal{U}_n$ must wait for at least one of them to become a CA. From the previous results we know that this will take an additional $(1 - \epsilon)\lg n$ generations.

Here is some notation that will be used throughout the proof. Let t_n denote $\lfloor(\zeta - \epsilon)\lg n\rfloor$. For $1 \leq i \leq n$, define J_i to be the event $\{G_t^i \in [1, \lg^2 n]$ for all $t \leq t_n\}$; we will also denote by J_i the indicator random variable corresponding to this event. Thus, $J_i = 1$ means that individual i in generation 0 does not become extinct by time t_n and that the number of descendants of this individual also remains relatively small (no more than $\lg^2(n)$) up to time t_n.

At time t_n these individuals still have a chance to become CAs, but they have not yet made much progress toward doing so. The number of such individuals is $N_n = \sum_{i=1}^{n} J_i$.

The next lemma shows that there is little dependence between the numbers of descendants of different individuals in the early stages of the process. The lemma gives an upper bound on a probability; a similar lower bound may be obtained, but it is not needed in the remainder of the proof.

Lemma 19. $P(J_1 J_2) \leq [P(J_1)]^2 (1 + o(1))$ *as* $n \to \infty$.

Proof. Consider individuals $I_{0,1}$ and $I_{0,2}$, that is, individuals 1 and 2 in generation 0. Let A_t denote the number of individuals in generation t who are descendants of $I_{0,1}$ but not of $I_{0,2}$. Let C_t denote the number of individuals in generation t who are descendants of $I_{0,2}$ but not of $I_{0,1}$. Let B_t denote the number of individuals in generation t who are descendants of both $I_{0,1}$ and $I_{0,2}$. This notation is local to this proof; in particular, B_t has a different meaning here than it did in the proof of Theorem 1. So $G_t^1 = A_t + B_t$ and $G_t^2 = C_t + B_t$. Letting $H_t = (A_t, B_t, C_t)$, the process $\{H_t\}$ is a Markov chain. For convenience we use the notation $P_H(a_t, b_t, c_t)$ for $P\{A_t = a_t, B_t = b_t, C_t = c_t\}$, $P_H(a_{t+1}, b_{t+1}, c_{t+1} \mid a_t, b_t, c_t)$ for $P\{A_{t+1} = a_{t+1}, B_{t+1} = b_{t+1}, C_{t+1} = c_{t+1} \mid A_t = a_t, B_t = b_t, C_t = c_t\}$, and so on.

We begin by observing that

$$P(J_1 J_2) \sim P(J_1 J_2 \{B_t = 0 \text{ for all } t \leq t_n\}). \tag{13}$$

This is easy to see intuitively: If A_t and C_t are both bounded by $\lg^2(n)$ and $B_t = 0$, then the conditional probability that $B_{t+1} > 0$ is at most $2 A_t C_t / n = O(\lg^4(n)/n)$. This suggests that for each $s \leq t_n$, given the event $J_1 J_2$, the conditional probability that B_t is positive for the first time at $t = s$ is $O(\lg^4(n)/n)$. Adding these probabilities over all $s \leq t_n = O(\lg n)$ would then give

$$P(B_t > 0 \text{ for some } t \leq t_n \mid J_1 J_2) = O(\lg^5(n)/n).$$

This is correct, and the relation

$$P(J_1 J_2) = P(J_1 J_2 \{B_t = 0 \text{ for all } t \leq t_n\}) \left[1 + O\left(\frac{\lg^5(n)}{n}\right)\right]$$

follows from a rather tedious calculation whose details we omit. The calculation bounds ratios of binomial probabilities in similar way to an argument that is given later in this proof.

Letting $\mathcal{L}_n$ denote the interval $[1, \lg^2 n]$, we want an upper bound on the probability

$$
\begin{aligned}
&P(J_1 J_2 \{B_t = 0 \text{ for all } t \leq t_n\}) \\
&= \sum_{a_1 \in \mathcal{L}_n} \cdots \sum_{a_{t_n} \in \mathcal{L}_n} \sum_{c_1 \in \mathcal{L}_n} \cdots \sum_{c_{t_n} \in \mathcal{L}_n} P_H(a_1, 0, c_1, a_2, 0, c_2, \ldots, a_{t_n}, 0, c_{t_n}) \\
&= \sum_{a_1, c_1 \in \mathcal{L}_n} P_H(a_1, 0, c_1) \sum_{a_2, c_2 \in \mathcal{L}_n} P_H(a_2, 0, c_2 \mid a_1, 0, c_1) \cdots \\
&\qquad\qquad \sum_{a_{t_n}, c_{t_n} \in \mathcal{L}_n} P_H(a_{t_n}, 0, c_{t_n} \mid a_{t_n-1}, 0, c_{t_n-1}).
\end{aligned}
$$

Defining

$$\alpha_s = \frac{2a_s}{n} - \frac{a_s(a_s + 2c_s)}{n^2},$$

$$\beta_s = \frac{2a_s c_s}{n^2},$$

and

$$\gamma_s = \frac{2c_s}{n} - \frac{c_s(c_s + 2a_s)}{n^2},$$

we may write

$$P_H(a_t, 0, c_t \mid a_{t-1}, 0, c_{t-1}) = P\{\text{Bin}(n, \alpha_{t-1}) = a_t\} P\left\{\text{Bin}\left(n - a_t, \frac{\gamma_{t-1}}{1 - \alpha_{t-1}}\right) = c_t\right\}$$

$$\times P\left\{\text{Bin}\left(n - a_t - c_t, \frac{\beta_{t-1}}{1 - \alpha_{t-1} - \gamma_{t-1}}\right) = 0\right\}.$$

We want to compare this to the analogous probability for two independent $\{G_t\}$ processes, that is, to

$$P_G(a_t \mid a_{t-1}) P_G(c_t \mid c_{t-1}) = P\{\text{Bin}(n, \alpha_{t-1} + \beta_{t-1}) = a_t\} P\{\text{Bin}(n, \gamma_{t-1} + \beta_{t-1}) = c_t\}.$$

The ratio

$$\frac{P_H(a_t, 0, c_t \mid a_{t-1}, 0, c_{t-1})}{P_G(a_t \mid a_{t-1}) P_G(c_t \mid c_{t-1})} \tag{14}$$

is the product of three terms:

$$\frac{P\{\text{Bin}(n, \alpha_{t-1}) = a_t\}}{P\{\text{Bin}(n, \alpha_{t-1} + \beta_{t-1}) = a_t\}}, \tag{15}$$

$$\frac{P\{\text{Bin}(n - a_t, \gamma_{t-1}/(1 - \alpha_{t-1})) = c_t\}}{P\{\text{Bin}(n, \gamma_{t-1} + \beta_{t-1}) = c_t\}}, \tag{16}$$

and

$$P\left\{\text{Bin}\left(n - a_t - c_t, \frac{\beta_{t-1}}{1 - \alpha_{t-1} - \gamma_{t-1}}\right) = 0\right\}. \tag{17}$$

We bound the third term (17) simply by 1; in fact it is close to 1. The term (15) is

$$\frac{\alpha_{t-1}^{a_t}(1 - \alpha_{t-1})^{n-a_t}}{(\alpha_{t-1} + \beta_{t-1})^{a_t}(1 - \alpha_{t-1} - \beta_{t-1})^{n-a_t}} \le \left(1 + \frac{\beta_{t-1}}{1 - \alpha_{t-1} - \beta_{t-1}}\right)^n$$

$$= 1 + O\left(\frac{\lg^4(n)}{n}\right),$$

since

$$\frac{\beta_{t-1}}{1 - \alpha_{t-1} - \beta_{t-1}} \sim \beta_{t-1} \le \frac{2 \lg^4(n)}{n^2}$$

for $a_{t-1}, c_{t-1} \leq \lg^2(n)$. By a similar calculation, (16) is also $1 + O\left(n^{-1}\lg^4(n)\right)$. Multiplying, we obtain

$$\frac{P_H(a_t, 0, c_t \mid a_{t-1}, 0, c_{t-1})}{P_G(a_t \mid a_{t-1})P_G(c_t \mid c_{t-1})} = 1 + O\left(\frac{\lg^4(n)}{n}\right).$$

Thus,

$$P(J_1 J_2 \{B_t = 0 \text{ for all } t \leq t_n\})$$

$$= \sum_{a_1, c_1 \in \mathcal{L}_n} P_H(a_1, 0, c_1) \sum_{a_2, c_2 \in \mathcal{L}_n} P_H(a_2, 0, c_2 \mid a_1, 0, c_1) \cdots$$

$$\sum_{a_{t_n}, c_{t_n} \in \mathcal{L}_n} P_H(a_{t_n}, 0, c_{t_n} \mid a_{t_n-1}, 0, c_{t_n-1})$$

$$\leq \sum_{a_1, c_1 \in \mathcal{L}_n} P_G(a_1 \mid 1)P_G(c_1 \mid 1) \sum_{a_2, c_2 \in \mathcal{L}_n} P_G(a_2 \mid a_1)P_G(c_2 \mid c_1) \cdots$$

$$\sum_{a_{t_n}, c_{t_n} \in \mathcal{L}_n} P_G(a_{t_n} \mid a_{t_n-1})P_G(c_{t_n} \mid c_{t_n-1}) \left[1 + O\left(\frac{\lg^4(n)}{n}\right)\right]^{t_n}$$

$$= \sum_{a_t, \dots, a_{t_n} \in \mathcal{L}_n} P_G(a_1 \mid 1)P_G(a_2 \mid a_1) \cdots P_G(a_{t_n} \mid a_{t_n-1})$$

$$\sum_{c_t, \dots, c_{t_n} \in \mathcal{L}_n} P_G(c_1 \mid 1)P_G(c_2 \mid c_1) \cdots P_G(c_{t_n} \mid c_{t_n-1}) \left[1 + O\left(\frac{\lg^4(n)}{n}\right)\right]^{t_n}$$

$$= \left[P\{G_t^1 \in \mathcal{L}_n \text{ for all } t \leq t_n\}\right]^2 \left[1 + O\left(\frac{\lg^5(n)}{n}\right)\right]$$

$$= [P(J_1)]^2 \left[1 + O\left(\frac{\lg^5(n)}{n}\right)\right].$$

This completes the proof.

Lemma 20. $N_n \to \infty$ *in probability as* $n \to \infty$.

Proof. We will show that the mean and standard deviation of N_n satisfy $EN_n \to \infty$ and $SD(N_n) = o(E(N_n))$. To see that $EN_n = nPJ_1 \to \infty$, begin with (3), which gives

$$P(J_1) \sim P\{Y_t \in [1, \lg^2(n)] \text{ for all } t \leq t_n\}.$$

This last probability is very close to $P\{Y_{t_n} \in [1, \lg^2(n)]\}$. Indeed, the difference

$$P\{Y_{t_n} \in [1, \lg^2(n)]\} - P\{Y_t \in [1, \lg^2(n)] \text{ for all } t \leq t_n\}$$

$$= P\{Y_{t_n} \in [1, \lg^2(n)], Y_t > \lg^2(n) \text{ for some } t < t_n\}, \tag{18}$$

is the probability that the Y process exceeds $\lg^2(n)$ some time before t_n but then decreases to be below $\lg^2(n)$ at time t_n. Since the Bernstein inequality applied to the Poisson distribution gives

$$P\{Y_{t+1} \leq Y_t \mid Y_t\} \leq \exp[-(\tfrac{3}{14})Y_t] \leq \exp[-(\tfrac{3}{14})\lg^2(n)] \quad \text{on} \quad \{Y_t > \lg^2(n)\},$$

the difference (18) is bounded by $t_n \exp[-(3/14)\lg^2(n)] = o(1/n)$. By Lemma 16,

$$\frac{\lg \mathrm{P}\{Y_{t_n} \in [1, \lg^2(n)]\}}{(\zeta - \epsilon)\lg n} \sim \frac{1}{t_n} \lg \mathrm{P}\{Y_{t_n} \in [1, \lg^2(n)]\} \to \lg \gamma = \frac{-1}{\zeta} \,,$$

which implies that $n\mathrm{P}\{Y_{t_n} \in [1, \lg^2(n)]\} \to \infty$. Thus,

$$n\mathrm{P}(J_1) \sim n\mathrm{P}\{Y_t \in [1, \lg^2(n)] \text{ for all } t \le t_n\} = n[\mathrm{P}\{Y_{t_n} \in [1, \lg^2(n)]\} + o(1/n)] \to \infty.$$

Finally, to see that $\mathrm{SD}(N_n) = o(\mathrm{E}(N_n))$, we apply Lemma 19 to obtain

$$\begin{aligned}
\mathrm{Var}(N_n) = \mathrm{E}(N_n^2) - (\mathrm{E}N_n)^2 &= n\mathrm{P}J_1 + n(n-1)\mathrm{P}(J_1 J_2) - (n\mathrm{P}J_1)^2 \\
&\le n\mathrm{P}J_1 + n(n-1)[\mathrm{P}(J_1)]^2(1 + o(1)) - (n\mathrm{P}J_1)^2 \\
&= o(n^2(\mathrm{P}J_1)^2) = o((\mathrm{E}N_n)^2).
\end{aligned}$$

Proof of the lower bound in Theorem 2. Defining $W_n = \{i : G_{t_n}^i \in [1, \lg^2(n)]\}$, we have

$$\begin{aligned}
\mathrm{P}\{\mathcal{U}_n &\le t_n + (1-\epsilon)\lg(n)\} \\
&\le \mathrm{P}\{W_n = \varnothing\} + \mathrm{P}\{\text{eventual extinction for all } i \in W_n\} \\
&\quad + \mathrm{P}\{G_{\lfloor t_n + (1-\epsilon)\lg(n)\rfloor}^i = n \text{ for some } i \in W_n\}. \qquad (19)
\end{aligned}$$

The cardinality of the set W_n is N_n. Since $N_n \overset{P}{\to} \infty$, clearly the probability that all individuals $\{I_{0,i} : i \in W_n\}$ eventually become extinct converges to 0; this is an easy consequence of results about extinction probabilities of Kämmerle (1991) or Möhle (1994). So it remains to show that the last probability in (19) tends to 0. To see this, taking $i \in W_n$, observe that for the event $\{G_{\lfloor t_n + (1-\epsilon)\lg(n)\rfloor}^i = n\}$ to occur the $\{G_t^i\}$ process must go from below $\lg^2(n)$ at time t_n to n at time $\lfloor t_n + (1-\epsilon)\lg(n)\rfloor$. That is, the process must go from below $\lg^2(n)$ to n within a time span of at most $(1-\epsilon)\lg(n)$ generations. However, by the proof of Proposition 15, we know that this has probability $o(1/n)$, so that, taking the union over $i \in W_n$ gives a total probability of $o(1)$.

4. Discussion

A motivation behind this study was the interest surrounding the idea of all of mankind having a recent common ancestor. In thinking about a mathematical treatment of that idea, it seemed natural to remove the restriction to the maternal line and consider a two-parent model.

We have seen that CAs occur very recently in the two-parent model studied here. The most recent CA occurs, with high probability, about $\lg n$ generations ago. Within $1.77 \lg n$ generations, with high probability, all individuals who are not extinct are CAs. These results describe the behavior of populations satisfying certain assumptions of random mating and so on. If our world really satisfied such assumptions, the anthropological excitement about the recentness of mitochondrial Eve would be misplaced: in only a tiny fraction of the time back to mitochondrial Eve, common ancestors of mankind would abound, and in fact a randomly chosen individual would be a CA with probability about 0.8.

If we wish to understand analogous questions in more complicated models that could better address phenomena such as the evolution of mankind, further study is required. For example,

the absence of geographic structure is a key feature limiting the applicability of the model studied here to such situations.

Conclusions based on analyses of simple models that ignore geographic considerations are commonly seen in the scientific discourse about the evolution of mankind. As a typical example, the abstract of Ayala's (1995) paper states

> The theory of gene coalescence suggests that, throughout the last 60 million years, human ancestral populations have had an effective size of 100 000 individuals or greater.

The investigation of '*Y*-chromosome Adam' by Dorit *et al.* (1995) is another interesting example. Such analyses rely strongly on the basic predictions of standard coalescent theory: n generations for the expected coalescence time of a pair of genes among a population of n genes, and $2n$ generations for the expected coalescence of the whole population.

On the other hand, it is doubtful that anyone would seriously entertain the two-parent answer of $\lg n$ for a CA in the context of the evolution of mankind. This raises conceptual questions. On what basis do we draw insight from the analysis of a one-parent model, when analysis of an analogous two-parent model leads to results we find implausible? Whereas the answer $2n$ given by a haploid model for the 'Eve' coalescence time may not be so obviously inapplicable in a given situation, the two-parent model's CA time of $\lg n$ may well be.

A possible source of comfort when confronting doubts about the realism of the assumptions underlying the standard coalescent model is the body of results about 'robustness' of the coalescent. Kingman (1982b) showed that the coalescent arises as a limiting genealogy in a whole class of models that includes Wright–Fisher and other classical models. However, this class of models assumes symmetries (related to exchangeability) that are typically violated in models incorporating population subdivision or geographic structure.

The two-parent results highlight the strong consequences that can follow from assuming models of the Wright–Fisher type, and in particular from assumptions of random mating. The mathematical conclusions of these models may not be as robust as one might hope, and models that ignore violations of assumptions such as random mating can easily lead to absurd estimates. Such unrealistically simple assumptions form a natural starting point for this first investigation of MRCAs in two-parent models, as it seems appropriate to begin with a direct analog of a classical model that lies at the foundation of the one-parent theory. But there is much that could be done to generalize this work.

In the context of one-parent models, a substantial literature investigates departures from the simplest assumptions of the classical models. In particular, models allowing population size to vary over time and models incorporating various forms of population subdivision and geographic structure are all under continuing investigation. See, for example, the recent collection of papers edited by Donnelly and Tavaré (1997), which gives a fine overview of recent work.

Möhle (1997) has considered the effect of varying population size in some genetic models distinguishing males and females. Möhle's focus is on the genetic question of ultimate fixation of an allele, so that the two-sex, diploid aspect of the models does not fundamentally affect the nature of the results, although it complicates the proofs; the results match earlier results of Donnelly (1986) about variable population size versions of the standard one-parent models. Aside from this work, generalizations of the standard assumptions remain to be investigated in two-parent models.

Acknowledgements

References

ATHREYA, K. B. AND NEY, P. (1972). *Branching Processes*. Springer, New York.

AYALA, F. J. (1995). The myth of Eve: molecular biology and human origins. *Science* **270,** 1930–1936.

BERNSTEIN, S. (1946). *The Theory of Probabilities*. Gastehizdat, Moscow.

CANN, R. L., STONEKING, M. AND WILSON, A. C. (1987). Mitochondrial DNA and human evolution. *Nature* **325,** 31–36.

DONNELLY, P. (1986). Genealogical approach to variable-population-size models in population genetics. *J. Appl. Prob.* **23,** 283–296.

DONNELLY, P. AND TAVARÉ, S. (eds.) (1997). *Progress in Population Genetics and Human Evolution*. Springer, New York.

DORIT, R. L., AKASHI, H. AND GILBERT, W. (1995). Absence of polymorphism at the ZFY locus on the human Y chromosome. *Science* **268,** 1183–1185.

GRIFFITHS, R. AND MARJORAM, P. (1997). An ancestral recombination graph. In *Progress in Population Genetics and Human Evolution*, eds. S. Tavaré and P. Donnelly, Springer, New York, pp. 257–270.

HUDSON, R. R. (1983). Properties of a neutral allele model with intragenic recombination. *Theoret. Popn. Biol.* **23,** 183–201.

HUDSON, R. R. (1990). Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* **7,** 1–44.

KÄMMERLE, K. (1989). Looking forwards and backwards in a bisexual model. *J. Appl. Prob.* **27,** 880–885.

KÄMMERLE, K. (1991). The extinction probability of descendants in bisexual models of fixed population size. *J. Appl. Prob.* **28,** 489–502.

KINGMAN, J. F. C. (1982a). Exchangeability and the evolution of large populations. In *Exchangeability in Probability and Statistics*, eds. G. Koch and F. Spizzichino, North-Holland, New York, pp. 97–112.

KINGMAN, J. F. C. (1982b). On the genealogy of large populations. In *Essays in Statistical Science,* eds. J. Gani and E. J. Hannan (J. Appl. Prob. **19A,**). Applied Probability Trust, Sheffield, UK, pp. 27–43.

MÖHLE, M. (1994). Forward and backward processes in bisexual models with fixed population sizes. *J. Appl. Prob.* **31,** 309–332.

MÖHLE, M. (1997). Fixation in bisexual models with variable population sizes. *J. Appl. Prob.* **34**, 436–448.

PÄÄBO, S. (1995). The Y chromosome and the origin of all of us (men). *Science* **268,** 1141–1142.

VIGILANT, L., STONEKING, M., HARPENDING, H., HAWKES, K. AND WILSON, A. C. (1991). African populations and the evolution of human mitochondrial DNA. *Science* **253,** 1503–1507.