# Ballot theorems and the heights of trees

Dana Addario-Berry
School of Computer Science & Department of Mathematics and Statistics
McGill University, Montreal, Canada
laddar@cs.mcgill.ca

September 2006

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of Doctor of Philosophy.

*Your file Votre référence*
*ISBN: 978-0-494-27748-5*
*Our file Notre référence*
*ISBN: 978-0-494-27748-5*

## Abstract

A *ballot theorem* is a theorem that yields information about the conditional probability that a random walk stays above its mean, given its value $S_t$ after some specified amount of time $t$. In the first part of this thesis, ballot theorems are proved for all walks whose steps consist of independent, identically distributed random variables that are in the range of attraction of the normal distribution. With a mild assumption on the moments of the steps, the results are strengthened; the latter results are shown to be within a constant factor of optimal when the value of the random walk at time $t$ is of order $\sqrt{t}$. Further results are proved for random walks whose value after time $t$ is of order $\Omega(t)$.

In the second part of the thesis, two questions about the *heights of random trees* are studied. The random trees that are studied are of interest from both a purely probabilistic, and an algorithmic perspective. It turns out that in two seemingly very distinct settings, the height of a random tree turns out to be closely linked to the behavior of a random walk, in particular to the probability that a random walk stays above its mean. The tools developed in the first part of the thesis, together with additional results, are then used to derive information about the moments of the height of these random trees. We also demonstrate that this information can be used to bound the moments of the minima of certain branching random walks.



## Résumé

Un théorème de ballottage est un théorème qui apporte de l'information sur la probabilité conditionnelle qu'une marche aléatoire soit toujours au dessus de sa moyenne, étant donné sa valeur après quelques pas.

Dans la première partie de cette thèse, on prouve un théorème de ballottage pour toutes marches aléatoires telles que les pas sont les variables aléatoires indépendantes et identiquement distribuées et telles qu'il existe une séquence $\{a_n\}_{n=1}^{\infty}$ telle que $S_n/a_n$ tend vers une distribution de Gauss. Si on suppose que les pas ont un moment d'ordre $2+\alpha$, $\alpha > 0$, alors on prouve un résultat plus fort, ce qui est le plus fort possible modulo un facteur constant.

Dans la deuxième partie de la thèse, on étudie deux questions qui s'adressent aux hauteurs des arbres aléatoires. Les arbres aléatoires qu'on étudie sont intéressants du point de vue de la probabilité ainsi que du point de vue de l'informatique. Ces deux questions qui ne semblent pas avoir rien à faire l'une avec l'autre peuvent toutes les deux être étudiées en utilisant une marche aléatoire et en utilisant les résultats de la première partie de la thèse. En étudiant ces arbres du point de vue des marches aléatoires, on arrive à borner les moments de leurs hauteurs. On démontre aussi que ces bornes impliquent des bornes pour les moments des minimums de certaines marches aléatoires de branchement.

# Acknowledgement

I would like to thank the support and systems staff in the department of computer science for their great help and consistent good cheer. I thank my labmates Ketan, Erin, Andrew, Sean, and Nicolas, for a remarkably fruitful working environment and for many blackboards full of exciting ideas. Special thanks are also due to Nicolas for reading much of my thesis and providing valuable suggestions. I owe a great debt to Luc Devroye, who taught me what beauty can be found in probability, and whose encyclopedic knowledge of his field has saved me countless hours of bibliographic research. Finally, I would like to thank my supervisor, Bruce Reed, who is an academic inspiration, an excellent supervisor and teacher, and a friend for life. If I ever teach a student half of what Bruce has taught me, I will consider it a job well done.

# Contributions of Authors

Chapters 3 and 4 arise from a joint paper with Bruce Reed and Nicolas Broutin (Addario-Berry et al., submitted).

For Rosie

# Contents

# List of Tables

# List of Figures

# Chapter 1

# A history of ballot theorems

*"There is a big difference between a fair game and a game it's wise to play."*

-Bertrand (1887b).

## 1.1 Discrete time ballot theorems

We begin by sketching the development of the classical ballot theorem as it first appeared in the Comptes Rendus de l'Academie des Sciences. The statement that is fairly called the first Ballot Theorem was due to Bertrand:

**Theorem 1** (Bertrand (1887c)). *We suppose that two candidates have been submitted to a vote in which the number of voters is $\mu$. Candidate $A$ obtains $n$ votes and is elected; candidate $B$ obtains $m = \mu - n$ votes. We ask for the probability that during the counting of the votes, the number of votes for $A$ is at all times greater than the number of votes for $B$. This probability is $(2n - \mu)/\mu = (n - m)/(n + m)$.*

Bertrand's "proof" of this theorem consists only of the observation that if $P_{n,m}$ counts the number of "favourable" voting records in which $A$ obtains $n$ votes, $B$ obtains $m$ votes and $A$ always leads during counting of the votes, then

$$P_{n+1,m+1} = P_{n+1,m} + P_{n,m+1},$$

the two terms on the right-hand side corresponding to whether the last vote counted is for candidate $B$ or candidate $A$, respectively. This "proof" can be easily formalized as follows. We first note that the binomial coefficient $B_{n,m} = (n+m)!/n!m!$ counts the total number of possible voting records in which $A$ receives $n$ votes and $B$ receives $m$, Thus, the theorem equivalently states that for any $n \geq m$, $B_{n,m} - P_{n,m}$, which we denote by $\Delta_{n,m}$, equals $2mB_{n,m}/(n+m)$. This certainly holds in the case $m = 0$ as $B_{n,0} = 1 = P_{n,0}$, and in the case $m = n$, as $P_{n,n} = 0$. The binomial coefficients also satisfy the recurrence $B_{n+1,m+1} = B_{n+1,m} + B_{n,m+1}$, thus so does the difference $\Delta_{n,m}$. By induction,

$$\begin{aligned} \Delta_{n+1,m+1} &= \Delta_{n+1,m} + \Delta_{n,m+1} \\ &= \frac{2m}{n+m+1} B_{n+1,m} + \frac{2(m+1)}{n+m+1} B_{n,m+1} = \frac{2(m+1)}{n+m+2} B_{n+1,m+1}, \end{aligned}$$

as is easily checked; it is likely that this is the proof Bertrand had in mind.

After Bertrand announced his result, there was a brief flurry of research into ballot theorems and coin-tossing games by the probabilists at the Academie des Sciences. The first formal proof of Bertrand's Ballot Theorem was due to André and appeared only weeks later (André, 1887). Andr'e exhibited a bijection between unfavourable voting records starting with a vote for $A$ and unfavourable voting records starting with a vote for $B$. As the latter number is clearly $B_{n,m-1}$, this immediately establishes that $B_{n,m} - P_{n,m} = 2B_{n,m-1} = 2mB_{n,m}/(n+m)$.

A little later, Barbier (1887) asserted but did not prove the following generalization of the classical Ballot Theorem: if $n > km$ for some integer $k$, then the probability candidate $A$

always has more than $k$-times as many votes as $B$ is precisely $(n-km)/(n+m)$. In response to the work of André and Barbier, Bertrand had this to say:

> "Though I proposed this curious question as an exercise in reason and calculation, in fact it is of great importance. It is linked to the important question of duration of games, previously considered by Huygens, [de] Moivre, Laplace, Lagrange, and Ampere. The problem is this: A gambler plays a game of chance in which in each round he wagers $\frac{1}{n}$'th of his initial fortune. What is the probability he is eventually ruined and that the he spends his last coin in the $(n+2\mu)$'th round?" (Bertrand, 1887a)

He notes that by considering the rounds in reverse order and applying Theorem 1 it is clear that the probability that ruin occurs in the $(n+2\mu)$'th round is nothing but $\frac{n}{n+2\mu}\binom{n+2\mu}{\mu}2^{-(2\mu+n)}$. By informal but basic computations, he then derives that the probability ruin occurs *before* the $(n+2\mu)$'th round is approximately $1-\frac{\sqrt{2/\pi n}}{\sqrt{n+2\mu}}$, so for this probability to be large, $\mu$ must be large compared to $n^2$. (Bertrand might have added Pascal, Fermat, and the Bernoullis (Hald, 1990, pp. 226-228) to his list of notable mathematicians who had considered the game of ruin; (Balakrishnan, 1997, pp. 98-114) gives an overview of prior work on ruin with an eye to its connections to the ballot theorem.)

Later in the same year, he proved that in a *fair game* (a game in which, at each step, the average change in fortunes is nil) where at each step, one coin changes hands, the expected number of rounds before ruin is infinite. He did so using the fact that by the above formula, the probability of ruin in the $t$'th round (for $t$ large) is of the order $1/t^{3/2}$, so the expected time to ruin behaves as the sum of $1/t^{1/2}$, which is divergent. He also stated that in a fair game in which player $A$ starts with $a$ dollars and player $B$ starts with $b$ dollars, the expected time until the game ends (until one is ruined) is precisely $ab$ (Bertrand, 1887b). This fact is easily proved by letting $e_{a,b}$ denote the expected time until the game ends and using the recurrence $e_{a,b} = 1 + (e_{a-1,b} + e_{a,b-1})/2$ (with boundary conditions $e_{a+b,0} = e_{0,a+b} = 0$). Expanding on Bertrand's work, Rouché provided an alternate proof of the above formula for the probability of ruin (Rouché, 1888a). He also provided an exact formula for the the expected number of rounds in a biased game in which player $A$ has $a$ dollars and bets $a_0$ dollars each round, player $B$ has $b$ dollars and bets $b_0$ dollars each round, and in each round player $A$ wins with probability $p$ satisfying $a_0 p > b_0(1-p)$ (Rouché, 1888b).

All the above questions and results can be restated in terms of a *random walk* on the set of integers $\mathbb{Z}$. For example, let $S_0 = 0$ and, for $i \geq 0$, $S_{i+1} = S_i \pm 1$, each with probability 1/2 and independently of the other steps - this is called a *symmetric simple random walk.* (For the remainder of this section, we will phrase our discussion in terms of random walks instead of votes, with $X_{i+1} = S_{i+1} - S_i$ constituting a step of the random walk.) Then Theorem 1 simply states that given that $S_t = h > 0$, the probability that $S_i > 0$ for all $i = 1, 2, \ldots, t$ (i.e. the random walk is favourable for $A$) is precisely $h/t$). Furthermore, the time to ruin when player $A$ has $a$ dollars and player $B$ has $b$ dollars is the *exit time* of the random walk $S$ from the interval $[a, -b]$. The research sketched above constitutes the first detailed examination of the properties of a random walk $S_0, S_1, \ldots, S_n$ conditioned on the value $S_n$, and the use of such information to study the asymptotic properties of such a walk.

In 1923, Aeppli proved Barbier's generalized ballot theorem by an argument similar to that used by André's. This proof is presented in Balakrishnan (1997, pp.101-102), where it is also observed that Barbier's theorem can be proved using Bertrand's original recurrence in the same fashion as above. A simple and elegant technique was used by Dvoretzky and Motzkin (1947) to prove Barbier's theorem; we use it to prove Bertrand's theorem as an example of its application, as it highlights an interesting perspective on ballot-style results.

We think of $\mathcal{X} = (X_1, \ldots, X_{n+m}, X_1)$ as being arranged clockwise around a cycle (so that $X_{n+m+1} = X_1$). There are precisely $n + m$ walks corresponding to this set, obtained by choosing a first step $X_i$, so to establish Bertrand's theorem it suffices to show that however $X_1, \ldots, X_{n+m}$ are chosen such that $S_n = n - m > 0$, precisely $n - m$ of the walks $X_{i+1}, \ldots, X_{n+m}, X_1, \ldots, X_i$ are favourable for $A$. Let $S_{ij} = X_{i+1} + \ldots + X_j$ (this sum includes $X_{n+m}$ if $i < j$). We say that $X_i, \ldots, X_j$ is a *bad run* if $S_{ij} = 0$ and $S_{i'j} < 0$ for all $i' \in \{i+1, \ldots, j\}$ (this set includes $n+m$ if $i > j$). In words, this condition states that $i$ is the first index for which the reversed walk starting with $X_j$ and ending with $X_{i+1}$ is nonnegative. It is immediate that if two bad runs intersect then one is contained in the other, so the maximal bad runs are pairwise disjoint. (An example of a random walk and its bad runs is shown in Figure 1.1).

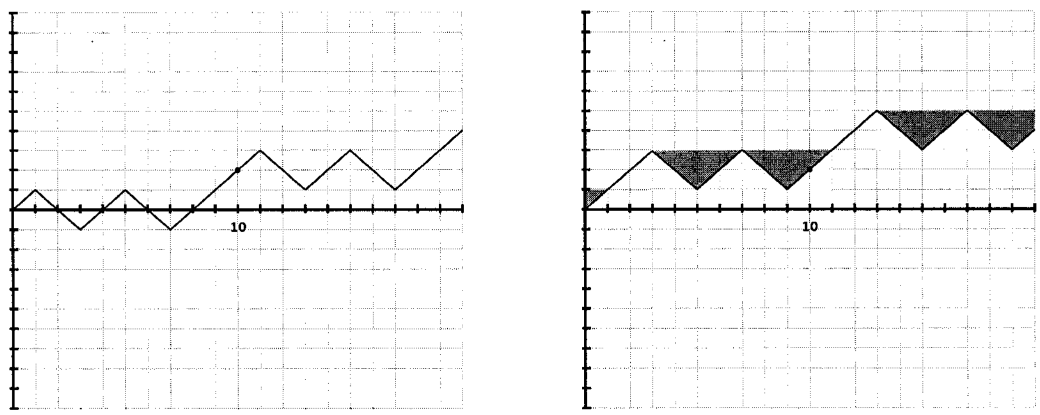


Figure 1.1: On the left appears the random walk corresponding to the voting sequence $(1, -1, -1, 1, 1, -1, -1, 1, 1, 1, 1)$, doubled to indicate the cyclic nature of the argument. On the right is the reversal of the random walk; the maximal bad runs are shaded grey.

If $X_i = 1$ and $S_{ij} = 0$ for some $j$ then $X_i$ begins a bad run, and since $S_n = \sum_{i=1}^{n} X_i > 0$, if $X_i = -1$ then $X_i$ ends a bad run. As $S_{ij} = 0$ for a maximal bad run and $X_i = 1$ for every $X_i$ not in a bad run, there must be precisely $n - m$ values of $i$ for which $X_i$ is not in a bad run. If the walk starting with $X_i$ is favourable for $A$ then for all $i \neq j$, $S_{ij}$ is positive, so $X_i$ is not in a bad run. Conversely, if $X_i$ is not in a bad run then $X_i = 1$ and for all $j \neq i$, $S_{ij} > 0$, so the walk starting with $X_i$ is favourable for $A$. Thus there are precisely $(n - m)$ favourable walks corresponding to $\mathcal{X}$, which is what we set out to prove.

With this technique, the proof of Barbier's theorem requires nothing more than letting the positive steps have value $1/k$ instead of 1. This proof is notable as it is the first time the idea of cyclic permutations was applied to prove a ballot-style result. This "rotation principle" is closely related to the *strong Markov property* of the random walk: for any integer $t \geq 0$, the random walk $S_t - S_t, S_{t+1} - S_t, S_{t+2} - S_t, \ldots$ has identical behavior to the walk $S_0, S_1, S_2$ and is independent of $S_0, S_1, \ldots, S_t$. (Informally, if we have examined the behavior of the walk up to time $S$, we may think of *restarting* the random walk at time $t$, starting from a height of $S_t$; this will be important in the generalized ballot theorems to be presented in Chapter 2.) This proof can be rewritten in terms of *lattice paths* by letting votes for $A$ be unit steps in the positive $x$-direction and votes for $B$ be unit steps in the positive $y$-direction. When conceived of in this manner, this proof immediately yields several natural generalizations (Dvoretzky and Motzkin, 1947; Grossman, 1950; Mohanty, 1966).

Starting in 1962, Lajos Takács proved a sequence of increasingly general ballot-style results and statements about the distribution of the maxima when the ballot is viewed as a random walk (Takács, 1962a,b,c, 1963, 1964a,b, 1967). I highlight two of these theorems below; I have not chosen the most general statements possible, but rather theorems which I believe capture key properties of ballot-style results.

A family of random variables $X_i, \ldots, X_n$ is *interchangeable* if for all $(r_1, \ldots, r_n) \in \mathbb{R}^n$ and all permutations $\sigma$ of $\{1, \ldots, n\}$, $\mathbf{P}\{X_i \leq r_i \forall 1 \leq i \leq n\} = \mathbf{P}\left\{X_i \leq r_{\sigma(i)} \forall 1 \leq i \leq n\right\}$. We say $X_1, \ldots, X_n$ is *cyclically interchangeable* if this equality holds for all *cyclic* permutations $\sigma$. A family of interchangeable random variables is cyclically interchangeable, but the converse is not always true. The first theorem states:

**Theorem 2.** *Suppose that $X_1, \ldots, X_n$ integer-valued, cyclically interchangeable random variables with maximum value* 1*, and for $1 \leq i \leq n$, let $S_i = X_1 + \ldots + X_i$. Then for any integer $0 \leq k \leq n$,*

$$\mathbf{P}\{S_i > 0 \ \forall 1 \leq i \leq n | S_n = k\} = \frac{k}{n}.$$

This theorem was proved independently by Tanner (1961) and Dwass (1962) – we note that it can also be proved by Dvoretzky and Motzkin's approach. (As a point of historical curiosity, Takacs' proof of this result in the special case of interchangeable random variables, and Dwass' proof of the more general result above, appeared in the same issue of Annals of Mathematical Statistics.) Theorem 2 and the "bad run" proof of Barbier's ballot theorem both suggest that the notion of cyclic interchangeability or something similar may lie at the heart of all ballot-style results.

**Theorem 3** (Takács (1967), p. 12). *Let $X_1, X_2, \ldots$ be an infinite sequence of iid integer random variables with mean $\mu$ and maximum value 1 and for any $i \geq 1$, let $S_i = X_1 + \ldots + X_i$. Then*

$$\mathbf{P}\{S_n > 0 \textit{ for } n = 1, 2, \ldots\} = \begin{cases} \mu & \textit{if } \mu > 0, \\ 0 & \textit{if } \mu \leq 0. \end{cases}$$

The proof of Theorem 3 proceeds by applying Theorem 2 to finite subsequences $X_1, X_2, \ldots, X_n$, so this theorem also seems to be based on cyclic interchangeability. Takács has generalized these theorems even further, proving similar statements for step functions with countably many discontinuities and in many cases finding the exact distribution of $\max_{i=1}^{n}(S_i - i)$.

(Takacs originally stated his results in terms of non-negative integer random variables – his original formulation results if we consider the variables $(1 - X_1), (1 - X_2), \ldots$ and the corresponding random walk.) Theorem 3 implies the following classical result about the probability of ever returning to zero in a biased simple random walk:

**Theorem 4** (Feller (1968), p. 274). *In a biased simple random walk $0 = R_0, R_1, \ldots$ in which $\mathbf{P}\{R_{i+1} - R_i = 1\} = p > 1/2$, $\mathbf{P}\{R_{i+1} - R_i = -1\} = 1 - p$, the probability that there is no $n \geq 1$ for which $R_n = 0$ is $2p - 1$.*

*Proof.* Observe that the expected value of $R_i - R_{i-1}$ is $2p - 1 > 0$, so if $R_1 = -1$ then with probability 1, $R_i = 0$ for some $i \geq 2$. Thus,

$$\mathbf{P}\{R_n \neq 0 \text{ for all } n \geq 1\} = \mathbf{P}\{R_n > 0 \text{ for all } n \geq 1\}.$$

The latter probability is equal to $2p - 1$ by Theorem 3. □

We close this section by presenting the beautiful "reflection principle" proof of Bertrand's theorem. We think of representing the symmetric simple random walk as a sequence of points $(0, S_0), (1, S_1), \ldots, (n, S_n)$ and connecting neighbouring points. If $S_1 = 1$ and the walk is unfavourable, then letting $k$ be the smallest value for which $S_k = 0$ and "reflecting" the random walk $S_0, \ldots, S_k$ in the $x$-axis yields a walk from $(0, 0)$ to $(n, t)$ whose first step is negative – this is shown in Figure 1.1. This yields a bijection between walks that are unfavourable for $A$ and start with a positive step, and walks that are unfavourable for $A$ and start with a negative step. As all walks starting with a negative step are unfavourable for $A$, all that remains is rote calculation. This proof is often incorrectly attributed to André (1887), who established the same bijection in a different way - its true origins remain unknown.

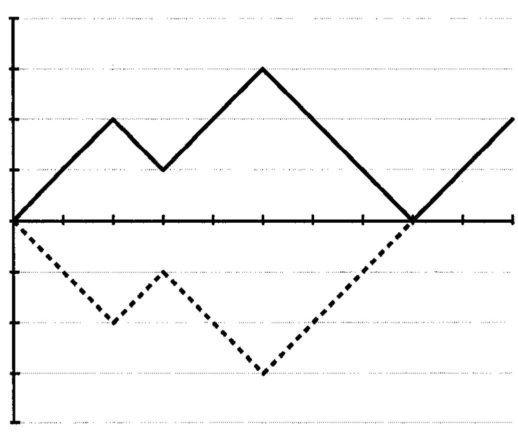

Figure 1.2: The dashed line is the reflection of the random walk from (0,0) to the first visit of the $x$-axis.

## 1.2 Continuous time ballot theorems

The theorems which follow are natural given the results presented in Section 1.1; however, their statements require slightly more preliminaries. A *stochastic process* is simply a nonempty set of real numbers $T$ and a collection of random variables $\{X_t, t \in T\}$ defined on some probability space. The collection of random variables $\{X_1, \ldots, X_n\}$ seen in Section 1.1 is an example of a stochastic process for which $T = \{1, 2, \ldots, n\}$. In this section we are concerned with stochastic processes for which $T = [0, r]$ for some $0 < r < \infty$ or else $T = [0, \infty)$.

A stochastic process $\{X_t, 0 \le t \le r\}$ has *(cyclically) interchangeable increments* if for all $n = 2, 3, \ldots,$, the finite collection of random variables $\{X_{rt/n} - X_{r(t-1)/n}, t = 1, 2, \ldots, n\}$ is (cyclically) intechangeable. A process $\{X_t, t \ge 0\}$ has *interchangeable increments* if for all $r > 0$ and $n > 0$, $\{X_{rt/n} - X_{r(t-1)/n}, t = 1, 2, \ldots, n\}$ is interchangeable, and is *stationary* if this latter collection is composed of independent identically distributed random variables. As in the discrete case, these are natural sorts of prerequisites for a ballot-style theorem to apply.

There is an unfortunate technical restriction which applies to all the ballot-style results we will see in this section. The stochastic process $\{X_t, t \in T\}$ is said to be *separable* if there are almost-everywhere-unique measurable functions $X^+, X_-$ such that almost surely $X_- \le X_t \le X^+$ for all $t \in T$, and there are countable subsets $S_-, S^+$ of $T$ such that almost surely $X^+ = \sup_{t \in S^+} X_t$ and $X_- = \inf_{t \in S_-} X_t$. The results of this section only hold for separable stochastic processes. In defense of the results, we note that there are nonseparable stochastic processes $\{X_t, 0 \le t \le r\}$ for which $\sup\{X_t - t, 0 \le t \le r\}$ is non-measurable. As the distribution of this random variable is one of the key issues with which we are concerned, the assumption of separability is natural and in some sense necessary in order for the results to be meaningful. Moreover, in very general settings it is possible to construct a separable stochastic process $\{Y_t | t \in T\}$ such that for all $t \in T$, $Y_t$ and $X_t$ are almost surely equal (see, e.g., Gikhman and Skorokhod, 1969, Sec.IV.2); in this case it can be fairly said that the assumption of separability is no loss.

The following theorem is the first example of a continous-time ballot theorem. A *sample function* of a stochastic process is a function $x_\omega : T \to \mathbb{R}$ given by fixing some $\omega \in \Omega$ and letting $x_\omega(t) = X_t(\omega)$.

**Theorem 5** (Takács (1965a)). *If $\{X_t, 0 \leq t \leq r\}$ is a separable stochastic process with cyclically interchangeable increments whose sample functions are almost surely nondecreasing step functions, then*

$$\mathbf{P}\{X_t - X_0 \leq t \text{ for } 0 \leq t \leq r | X_r - X_0 = s\} = \begin{cases} \frac{t-s}{t} & \text{if } 0 \leq s \leq t, \\ 0 & \text{otherwise.} \end{cases}$$

This theorem is a natural continuous equivalent of Theorem 2 of Section 1.1; it can be used to prove a theorem in the vein of Theorem 3 which applies to stochastic processes $\{X_t, t \geq 0\}$. Takács' other ballot-style results for continuous stochastic processes are also essentially step-by-step extensions of his results from the discrete setting; see Takács (1965a,b, 1967, 1970b).

In 1957, Baxter and Donsker derived a double integral representation for $\sup\{X_t - t, t \geq 0\}$ when this process has stationary independent increments. Their proof proceeds by analyzing the zeros of a complex-valued function associated to the process. They are able to use their representation to explicitly derive its distribution when the process is a Gaussian process, a coin-tossing process, or a Poisson process. This result was rediscovered by Takács (1970a), who also derived the joint distribution of $X_r$ and $\sup\{X_t - t, 0 \leq t \leq r\}$ for $r$ finite, using a generating function approach. Though these results are clearly related to the continuous ballot theorems, they are not as elegant, and neither their statements nor their proofs bring to mind the ballot theorem. It seems that considering separable stationary processes in their full generality does not impose enough structure for it to be possible to prove these results via straightforward generalization of the discrete equivalents.

A beautiful perspective on the ballot theorem appears by considering *random measures* instead of stochastic processes. Given an almost surely nondecreasing separable stochastic process $\{X_t, 0 \le t \le r\}$, fixing any element $\omega$ of the underlying probability space $\Omega$ yields a sample function $x_\omega$. By our assumptions on the stochastic process, almost every sample function $x_\omega$ yields a measure $\mu_\omega$ on $[0,r]$, where $\mu_\omega[0,b] = x_\omega(b) - x_\omega(a)$. This allows us to define a "random" measure $\mu$ on $[0,r]$; $\mu$ is a function with domain $\Omega$, $\mu(\omega) = \mu_\omega$, and for almost all $\omega \in \Omega$, $\mu(\omega)$ is a measure on $[0,r]$. If $x_\omega$ is a nondecreasing step function, then $\mu_\omega$ has countable support, so is *singular* with respect to the Lebesgue measure (i.e. the set of points which have positive $\mu_\omega$-measure has *Lebesgue* measure 0); if this holds for almost all $\omega$ then $\mu$ is an "almost surely singular" random measure.

We have just seen an example of a random measure; we now turn to a more precise definition. Given a probability space $\mathcal{S} = (\Omega, \Sigma, \mathbf{P})$, a random measure on a possibly infinite interval $T \subset \mathbb{R}$ is a function $\mu$ with domain $\Omega \times T$ satisfying that for all $r \in T$, $\mu(\cdot, r)$ is a random variable in $\mathcal{S}$, and for almost all $\omega \in \Omega$, $\mu(\omega, \cdot)$ is a measure on $T$. A random measure $\mu$ is almost surely singular if for almost all $\omega \in \Omega$, $\mu(\omega, \cdot)$ is a measure on $T$ singular with respect to the Lebesgue measure. (This definition hides some technicality; in particular, for the definition to be useful it is key that the set of $\omega$ for which $\mu$ is singular is itself a measurable set! See Kallenberg (1999) for details.) A random measure $\mu$ on $\mathbb{R}^+$, say, is stationary if for all $t$, letting $X_{t,i} = \mu(\cdot, (i+1)/t) - \mu(\cdot, i/t)$, the family $\{X_{t,i} | i \in \mathbb{N}\}$ is composed of independent identically distributed random variables; stationarity for finite intervals is defined similarly.

This perspective can be used to generalize Theorem 5. Konstantopoulos (1995) has done so, as well as providing a beautiful proof using a continuous analog of the reflection principle.

The most powerful theorem along these lines to date is due to Kallenberg. To a given stationary random measure $\mu$ defined on $T \subseteq \mathbb{R}^+$ we associate a random variable $I$ called the *sample intensity* of $\mu$. (Intuitively, $I$ is the random average number of points in an arbitrary measurable set $B \subset T$ of positive finite measure, normalized by the measure of $B$. For a formal definition, see (Kallenberg, 2003, p. 189).)

**Theorem 6** (Kallenberg (1999)). *Let $\mu$ be an almost surely singular, stationary random measure on $T = \mathbb{R}^+$ or $T = (0,1]$ with sample intensity $I$ and let $X_t = \mu(\cdot, t) - \mu(\cdot, 0)$ for $t \in T$. Then there exists a uniform $[0,1]$ random variable $U$ independent from $I$ such that*

$$\sup_{t\in T} \frac{X_t}{t} = \frac{I}{U} \qquad \textit{almost surely.}$$

It turns out that if $T = (0,1]$ then conditional upon the event that $X_1 = m$, $I = m$ almost surely. It follows that in this case $\mathbf{P}\left\{\sup_{t\in T} \frac{X_t}{t} \leq 1 | X_1\right\} = \max\{1 - X_1, 0\}$. Similarly, if $T = \mathbb{R}^+$ and $\frac{X_t}{t} \to m$ almost surely as $t \to \infty$, then $I = m$ almost surely, so in this case $\mathbf{P}\left\{\sup_{t\in T} \frac{X_t}{t} \leq 1\right\} = \max\{1 - m, 0\}$. This theorem can thus be seen to include continuous generalizations of both Theorem 2 and Theorem 3.

Kallenberg has also proved the following as a corollary of Theorem 6 (this is a slight reformulation of his original statement, which applied to infinite sequences):

**Theorem 7.** *If $X$ is a real random variable with maximum value 1 and $\{X_1, X_2, \ldots, X_n\}$ are iid copies of $X$ with corresponding partial sums $\{0 = S_0, S_1, \ldots, S_n\}$, then*

$$\mathbf{P}\{S_i > 0 \forall 1 \leq i \leq n | S_n\} \geq \frac{S_n}{n}.$$

It is worth comparing this theorem with Theorem 2; the theorems are almost identical, but Theorem 7 relaxes the integrality restriction at the cost of replacing the equality of Theorem 2 with an inequality. In Chapter 2.4 we prove that such an inequality holds for a broad class of random variables that need not satisfy a one-side boundedness condition, and prove upper bounds of the same order.

## 1.3 Outline

To date, Theorem 7 is the only ballot-style result which has been proved for random walks that may take non-integer values. Paraphrasing Harry Kesten (1993a), the goal of the first part of this thesis is to move towards making ballot theorems part of "the general theory of random walks" – part of the body of results that hold for *all* random walks (with independent identically distributed steps), regardless of the precise distribution of their steps. In Chapter 2, we succeed in proving ballot-style theorems that hold for a broad class of random walks, including all random walks that can be renormalized to converge in distribution to a normal random variable. A *truly* general ballot theorem, however, remains beyond our grasp.

The work of Chapter 2 is in pursuit of more general ballot theorems for their own sake. It turns out, however, that ballot theorems can and have been used for understanding many probabilistic and algorithmic questions, often seemingly unrelated to the ballot theorem (for example, the work of Takács (1962c, 1967, 1989) explores the applications of ballot theorems to the theory of queues, and Dwass (1969) exhibited a connection between ballot theorems and branching processes; this is a connection we will also explore).

In the second part of the thesis, we will see that in several, seemingly unrelated settings, ballot theorems provide useful and novel insight into the behavior of random structures. In Chapter 3, we use a ballot theorem-inspired approach to study the growth of the components of the random graph $G_{n,p}$. In Chapter 4, we build on the work of Chapter 3 to study the height of a random tree closely linked to the $G_{n,p}$ graph process. Finally, in Chapter 5, we use random walks and ballot theorems to study the moments of the maxima of branching processes (which, we shall see, can equivalently be interpreted as answering a question about the moments of the heights of a certain class of random trees). Though the settings of Chapters 4 and 5 are seemingly very distinct, in both cases the height of a random tree turns out to be intimately linked to the behavior of a random walk, its first return to 0, and whether it stays positive (or negative) before that time.

# Chapter 2

# A general ballot theorem

The aim of this chapter is to prove analogs of the discrete-time ballot theorems of the previous chapter for more general random variables. The Theorems of Section 1.1 all have two restrictions: (1) they apply only to integer-valued random variables, and (2) they apply only to random variables that are bounded from one or both sides. (In the continuous-time setting, the restriction appearing in Section 1.2 that the stochastic processes are almost surely integer-valued, increasing step functions is of the same flavour.) In this chapter we investigate what ballot-style theorems can be proved when such restrictions are removed.

The restrictions (1) and (2) are necessary for the results of Section 1.1 to hold. Suppose, for example, that we relax the condition of Theorem 2 requiring that the variables are bounded from above by $+1$. If $X$ takes every value in $\mathbb{N}$ with positive probability, then $\mathbf{P}\{S_i > 0 \forall 1 \le i \le n | S_n = n\} < 1$, so the conclusion of the theorem fails to hold. For a more explicit example, let $X$ be any random variable taking values $\pm 1, \pm 4$ and define the corresponding cyclically interchangeable sequence and random walk. For $S_3 = 2$ to occur, we must have $\{X_1, X_2, X_3\} = \{4, -1, -1\}$. In this case, for $S_i > 0$, $i = 1, 2, 3$ to occur, $X_1$ must equal 4. By cyclic interchangeability, this occurs with probability $1/3$, and not $2/3$, as

Theorem 2 would suggest. This shows that the boundedness condition (2) is required. If we relax the integrality condition (1), we can construct a similar example where the conclusions of Theorem 2 do not hold.

Since the results of Section 1.1 can not be directly generalized to a broader class of random variables, we seek conditions on the distribution of $X$ so that the bounds of that section have the correct order, i.e., so that $\mathbf{P}\{S_i > 0 \ \forall\ 1 \leq i \leq n | S_n = k\} = \Theta(k/n)$. (When we consider random variables that are not necessarily integer-valued, the right conditioning will in fact be on an event such as $\{k \leq S_n < k+1\}$ or something similar.) How close we can come to this conclusion will depend on what restrictions on $X$ we are willing to accept. It turns out that a statement of this flavour holds for the *mean zero* random walk $S_n^0 = S_n - n\mathbf{E}X$ as long as there is a sequence $\{a_n\}_{n\geq 0}$ for which $(S_n - n\mathbf{E}X)/a_n$ converges to a non-degenerate normal distribution (in this case, we say that $X$ is in the *range of attraction of the normal distribution* and write $X \in \mathcal{D}$; for example, the classical central limit theorem states that if $\mathbf{E}\{X^2\} < \infty$ then we may take $a_n = \sqrt{n}$ for all $n$.)

From this point on, we restrict our attention to sums of mean zero random variables. We note this condition is in some sense necessary in order for the results we are hoping for to hold. If $\mathbf{E}X \neq 0$ – say $\mathbf{E}X > 0$ – then it is possible that $X$ is non-negative, so the only way for $S_n = 0$ to occur is that $X_1 = \ldots = X_n = 0$, and so $\mathbf{P}\{S_i > 0 \ \forall\ 1 \leq i \leq n | S_n = 0\} = 0$, and not $\Theta(1/n)$ as we would hope from the results of the previous chapter.

In Section 2.1 we demonstrate the approach of the chapter in a restricted setting. This allows us to highlight the key ideas behind the general ballot theorems of this chapter without too much notational and technical burden. In Sections 2.2 through 2.4, we develop and prove generalized ballot theorems which hold when $X$ has finite variance; these results are strongest when $S_n = O(\sqrt{n})$. In Section 2.5 we prove a ballot-style result which is interesting when $S_n = \Theta(n)$. Finally, in Section 2.6 we address the limits of our approach and potential avenues of research.

## 2.1 Ballot theorems for closely fought elections

One of the most basic questions a ballot theorem can be said to answer is: given that an election resulted in a *tie*, what is the probability that one of the candidates had the lead at every point aside from the very beginning and and the very end. In the language of random walks, the question is: given that $S_n = 0$, what is the probability that $S$ does not return to 0 or change sign between time 0 and time $n$? Erik Sparre Andersen has studied the conditional behavior of random walks given that $S_n = 0$ in great detail, in particular deriving beautiful results on the distribution of the maximum, the minimum, and the amount of time spent above zero. Much of the next five paragraphs can be found in Andersen (1953), for example, in slightly altered terminology.

We call the event that $S_n$ does not return to zero or change sign before time $n$, $Lead_n$. We can easily bound $\mathbf{P}\{Lead_n | S_n = 0\}$ using the fact that $X_1, \ldots, X_n$ are interchangeable. If we condition on the multiset of outcomes $\{X_1, \ldots, X_n\} = \{x_{\sigma(1)}, \ldots, x_{\sigma(n)}\}$, and then choose a uniformly random cyclic permutation $\sigma$ and a uniform element $i$ of $\{1, \ldots, n\}$, then the interchangeability of $X_1, \ldots, X_n$ implies that $(x_{\sigma(i)}, \ldots, x_{\sigma(n)}, x_{\sigma(1)}, \ldots, x_{\sigma(i-1)})$ has the same distribution as if we had sampled directly from $(X_1, \ldots, X_n)$.

Letting $s_j = \sum_{k=1}^{j-1} x_{\sigma(k)}$, in order for $Lead_n$ to occur given that $S_n = 0$, it must be the case that $s_i$ is either the unique maximum or the unique minimum among $\{s_1, \ldots, s_n\}$. The probability that this occurs is at most $2/n$ as it is exactly $2/n$ if there are unique maxima and minima, and less if either the maximum or minimum is not unique. Therefore,

$$\mathbf{P}\{Lead_n | S_n = 0\} \leq \frac{2}{n}. \tag{2.1}$$

On the other hand, the sequence certainly has *some* maximum (resp. minimum) $s_i$, and if $X_1 = x_i$ then $S_j$ is always non-positive (resp. non-negative). Denoting this event by $Nonpos_n$ (resp. $Nonneg_n$), we therefore have

$$\mathbf{P}\left\{Nonpos_n | S_n = 0\right\} \geq \frac{1}{n} \quad \text{and} \quad \mathbf{P}\left\{Nonneg_n | S_n = 0\right\} \geq \frac{1}{n} \tag{2.2}$$

If $S_n = 0$ then the $(n-1)$ renormalized random variables given by $X_i' = X_{i+1} + X_1/(n-1)$ satisfy $(n-1)S_{n-1}' = (n-1)\sum_{i=1}^{n-1} X_i' = (n-1)\sum_{i=1}^{n} X_i = 0$. If $X_1 > 0$ and none of the *renormalized* partial sums are negative, then $Lead_n$ occurs. The renormalized random variables are still interchangeable (see Andersen (1953, Lemma 2) for a proof of this easy fact), so we may apply the second bound of (2.2) to obtain

$$\mathbf{P}\left\{Lead_n | S_n = 0, X_1 > 0\right\} \geq \frac{1}{n-1}.$$

An identical argument yields the same bound for $\mathbf{P}\left\{Lead_n | S_n = 0, X_1 < 0\right\}$, and combining these bounds yields

$$\begin{aligned} \mathbf{P}\left\{Lead_n | S_n = 0\right\} &\geq \mathbf{P}\left\{Lead_n | S_n = 0, X_1 \neq 0\right\} \mathbf{P}\left\{X_1 \neq 0 | S_n = 0\right\} \\ &\geq \frac{1 - \mathbf{P}\left\{X_1 = 0 | S_n = 0\right\}}{n-1}. \end{aligned}$$

As long as $\mathbf{P}\left\{X_1 = 0 | S_n = 0\right\} < 1$, this yields that $\mathbf{P}\left\{Lead_n | S_n = 0\right\} \geq \alpha/n$ for some $\alpha > 0$. By interchangeability, $\mathbf{P}\left\{X_1 = 0 | S_n = 0\right\} < 1$ as long as $S_n = 0$ does not imply that $X_1 = \ldots = X_n = 0$. (Note, however, that there *are* cases where $\mathbf{P}\left\{X_1 = 0 | S_n = 0\right\} = 1$, for example if the $X_i$ only take values in the non-negative integers and in the negative multiples of $\sqrt{2}$.) In later chapters, we refer to the above chain of reasoning as *the standard rotation argument.*

We would like to derive similar information about $\mathbf{P}\left\{Lead_n | S_n = r\right\}$ for arbitrary $r$. In this preliminary discussion we restrict our attention to "small" $r$ (the meaning of which will become clear), which led to the phrase "closely fought elections" in the section title. Furthermore, instead of studying $Lead_n$, for the time being we will focus on the closely related event $Lead_n^+$ that $S_i > 0$ for all $0 < i < n$ (which is equivalently the event $Lead_n \cap \{X_1 > 0\}$). We can apply a similar argument to that seen above to derive a weak lower bound on $\mathbf{P}\left\{Lead_n^+|S_n = r\right\}$. By again renormalizing the random walk, this time letting $X_i' = X_i - r/n$, we obtain a new random walk $S'$, and the event that $S_n = r$ is the event that $S_n' = 0$. The random variables $X_i'$ are still interchangeable, and if the partial sums $S_i'$ are all non-negative then $S_i > 0$ for $i = 1, 2, \ldots, n$, so $Lead_n^+$ occurs. (2.2) thus yields

$$\mathbf{P}\left\{Lead_n^+|S_n = r\right\} \geq \frac{1}{n} \tag{2.3}$$

for all integers $r \geq 0$.

In fact, as we saw at the end of Section 1.2, for many random variables $X$ a lower bound of order $r/n$ holds in (2.3), and we shall show that such bounds hold for an even larger class of random variables, for certain $r$. The primary goal of this chapter is to prove upper bounds of the same order. Most generally speaking, the question we seek an answer to is this: what are sufficient conditions on the structure of a multiset $S$ of $n$ numbers to ensure that if the elements of the multiset sum to $r$, then in a uniformly random permutation of the set, all partial sums are positive with probability of order $r/n$?

One way to construct a uniformly random permutation of a multiset $\{x_1, \ldots, x_n\}$ is as follows.

- Sample uniformly from $S$ with replacement to obtain a sequence $X_1, X_2, \ldots, X_t$, stopping at the first time $t$ that every element of $S$ has appeared (we emphasize that $x_i$ and $x_j$ must both appear if $i \neq j$, even if $x_i = x_j$).
- Throw out all but the first occurrence of each $x_i$ in the sequence $X_1, \ldots, X_t$.

The sequence $X_1, \ldots, X_t$ is a random walk whose steps are given by sampling uniformly from $S$. The subsequence of $X_1, \ldots, X_t$ resulting from throwing away all but the first occurrence of each $x_i$ is a uniformly random permutation. For fixed $1 \leq k \leq n$, the probability $X_{k+1} = X_i$ for some $i \leq k$ is at most $k/n$. Fixing $t$ and summing this bound over $k < t$, it follows that the probability there is a repeated element in $X_1, \ldots, X_t$ is at most $t^2/n$. When $t = o(\sqrt{n})$, this probability is $o(1)$. In other words, for such $t$, *with high probability the first $t$ elements of a uniformly random permutation of $S$ look like the first $t$ steps of a random walk whose steps are given by sampling uniformly and independently from $S$.*

In this chapter, we focus our attention on sets $S$ whose elements are sampled independently from a mean-zero probability distribution, i.e., they are the steps of a mean-zero random walk. Using the connection described in the previous paragraph, it is possible to apply parts of our analysis to sets $S$ that do not obey this restriction. We do not, however, pursue this direction in detail.

We now return to the discussion that led to (2.3). Our basic approach is to try to find a way to express $\mathbf{P}\left\{Lead_n^+ | S_n = r\right\}$ in terms of $\mathbf{P}\left\{Lead_n | S_n = 0\right\}$, which we now understand quite well. For the purposes of this preliminary discussion, we assume that $X$ is a non-zero symmetric integer random variable with maximum value $A$. We will make several other simplifying assumptions as we proceed, to facilitate the exposition. As the reader can verify, all assumptions are in some sense natural given that $X$ is in the range of attraction of the normal distribution. We suggest the reader imagine that $X$ takes only values $+1$ and $-1$ (though all assumptions are valid much more generally) in reading what follows.

In order for $Lead_n^+$ and $\{S_n = r\}$ to occur, it is necessary that (1) letting $T$ be the first time $t > 0$ that $S_t \leq 0$ or $S_t \geq r$, we have $S_T \geq r$ (we denote this event $Pos_r$), (2) $T < n$, (3) $X_{T+1} + \ldots + X_n = r - S_T$, and (4) for all $T < i < n$, $X_{T+1} + \ldots + X_i > -(r + A)$. These events are shown in Figure 2.1.

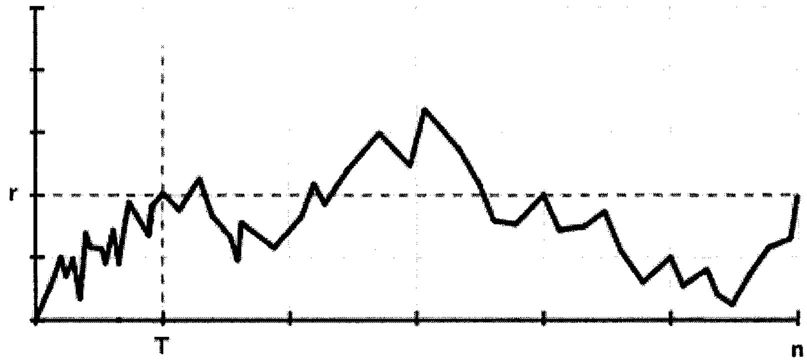


Figure 2.1: In the above random walk $Pos_r$ occurs as the walk reaches height at least $r$ before its first return to 0. The time $T$ at which this occurs is less than $n/4$. The remainder of the walk is always greater than $-(r+A)$, and $S_n = S_T = r$. Thus, $Lead_n^+$ and $\{S_n = r\}$ both occur.

For a random walk $S$ and $a > 0$, we say that $S_n$ *stays above* $-a$ if, for all $1 < i < n$, $S_i > -a$, and denote this event by $S_n$ *abo* $-a$. (So in particular, the event $S_n$ *abo* 0 is just the event $Lead_n^+$.) Let $S'$ be the random walk $S$ restarted at time $T$, so $S'_i = X_{T+1} + \ldots + X_{T+i}$ – then by the strong Markov property, we have

$$\begin{aligned}\mathbf{P}\left\{Lead_n^+, S_n = r\right\} \leq & \sum_{t=1}^{n-1} \mathbf{P}\left\{Pos_r, T = t\right\} \cdot \\ & \mathbf{P}\left\{S'_{n-t} = r - S_t, S'_{n-t}\ abo\ -(r+A) \mid Pos_r, T = t\right\} \quad (2.4)\end{aligned}$$

As we shall see later, for $r = o(\sqrt{n/\log n})$ (which we hereafter refer to as "small" $r$), $T$ is $o(n)$ with high enough probability that in fact,

$$\begin{aligned}\mathbf{P}\left\{Lead_n^+, S_n = r\right\} = & O\Bigg( \sum_{t=1}^{\lfloor n/2 \rfloor} \mathbf{P}\left\{Pos_r, T = t\right\} \cdot \\ & \mathbf{P}\left\{S'_{n-t} = r - S_t, S'_{n-t}\ abo\ -(r+A) | Pos_r, T = t\right\} \Bigg). \quad (2.5)\end{aligned}$$

We now make three additional simplifying assumptions about the random walk $S$:

(A1) for $n/2 \leq n' \leq n$, for small $r$, and for integers $0 \leq |k| \leq A$, $\mathbf{P}\{S_{n'} = k, S_{n'}\ abo\ -(r+A)\} = \Theta(\mathbf{P}\{S_n = k, S_n\ abo\ -r\}) = \Theta(\mathbf{P}\{S_n = 0, S_n\ abo\ -r\})$,

(A2) for small $r$, $\mathbf{P}\{Pos_r, T \leq \lfloor n/2 \rfloor\} = \Theta(\mathbf{P}\{Pos_r\})$, and

(A3) for small $r$, $\mathbf{P}\{S_n = 0\} = \Theta(\mathbf{P}\{S_n = r\})$.

Since if $T = t$, then $-A \leq r - S_t \leq 0$, by the strong Markov property and by applying first (A1) then (A2) in (2.5), we have that for small $r$,

$$\begin{aligned}\mathbf{P}\{Lead_n^+, S_n = r\} &= \Theta\left(\mathbf{P}\{Pos_r, T \leq \lfloor n/2 \rfloor\}\mathbf{P}\{S_n = 0, S_n\ abo\ -r\}\right). \\ &= \Theta\left(\mathbf{P}\{Pos_r\}\mathbf{P}\{S_n = 0, S_n\ abo\ -r\}\right). \qquad (2.6)\end{aligned}$$

We can also relate the events $Lead_n^+$ and $S_n\ abo\ -r$ in the following manner. In order for $\{Lead_n^+, S_n = 0\}$ to occur, it suffices that (A) $Pos_r$ occurs and $T < n/4$, (B) letting $T'$ be the last time $t < n$ that $S_t \geq r$ or $S_t = 0$, we have $S_t \geq r$ (we denote this event $Pos'_r$) and $n - T' < n/4$, and (C) letting $S_i^* = S_{T+i} - S_T$, and letting $a = -S_T - (S_n - S_{n-T'})$, we have $S_{T'-T}^* = a$ and $S_{T'-T}^*\ abo\ -r$. These events are shown in Figure 3.2.

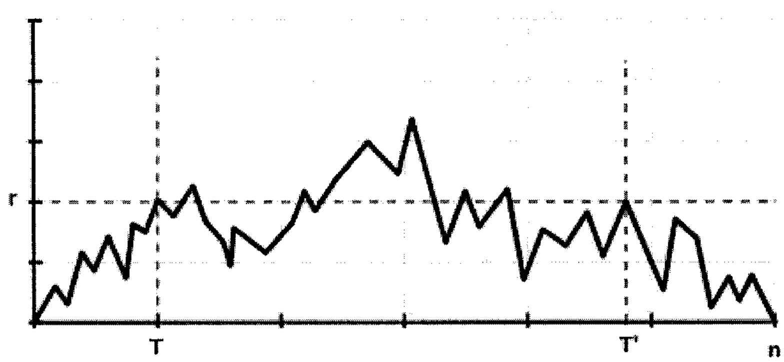


Figure 2.2: A random walk for which the events (A),(B), and (C) all occur. For the random walk depicted above, $a = 0$ as $S_T = -(S_n - S_{n-T}) = r$.

By our assumption that $r$ is small, we will be able to show that

$$\mathbf{P}\{Pos_r, Pos'_r, T < n/4, n - T' < n/4\} = \Theta(\mathbf{P}\{Pos_r\}\mathbf{P}\{Pos'_r\}). \tag{2.7}$$

Applying the assumption (A1) to the event (C), we therefore have

$$\begin{aligned}\mathbf{P}\left\{Lead_n^+, S_n = 0\right\} &= \Omega\Bigg(\mathbf{P}\{Pos_r\}\mathbf{P}\{Pos'_r\}\cdot \\ &\qquad \mathbf{P}\left\{S^*_{T'-T} = 0, S^*_{T'-T}\ abo\ -r | T < \frac{n}{4}, T' > \frac{3n}{4}\right\}\Bigg) \\ &= \Omega(\mathbf{P}\{Pos_r\}\mathbf{P}\{Pos'_r\}\mathbf{P}\{S_n = 0, S_n\ abo\ -r\}). \end{aligned} \tag{2.8}$$

Combining (2.6) and (2.8) yields that

$$\mathbf{P}\left\{Lead_n^+, S_n = r\right\} = O\left(\frac{\mathbf{P}\left\{Lead_n^+, S_n = 0\right\}}{\mathbf{P}\{Pos'_r\}}\right),$$

and finally, assumption (A3) and (2.3) yield that

$$\mathbf{P}\left\{Lead_n^+ | S_n = r\right\} = O\left(\frac{\mathbf{P}\left\{Lead_n^+ | S_n = 0\right\}}{\mathbf{P}\{Pos'_r\}}\right) = O\left(\frac{1}{\mathbf{P}\{Pos'_r\}\cdot n}\right). \tag{2.9}$$

We can thus derive upper bounds on $\mathbf{P}\left\{Lead_n^+ | S_n = r\right\}$ if we can bound $\mathbf{P}\{Pos'_r\}$, and if we can justify our assumptions (A1)-(A3). The primary work in doing so will be to justify our claims about the upper tail of $T$ and our subsequent probability manipulations. Bounding the upper tail of $T$ reduces to understanding the expected time $T$ until the random walk first exits $[0, r]$, and the probability such a time exceeds its expected value by very much.

**Observation 1:** The intuition given by the symmetric simple random walk, as discussed in Chapter 1, is that $T$ should have expected value about $r^2$. We will prove this, and using a renewal argument, we will then easily show that $\mathbf{P}\{T > 2k\mathbf{E}\{T\}\} = O(1/2^k)$ for $k \geq 1$. Thus, if $r = o(\sqrt{n/\log n})$, the probability that $T$ exceeds $\epsilon n$, (for some fixed $\epsilon > 0$) is much smaller than $1/n^2$ once $n$ is large enough.

**Observation 2:** *Assuming* $\mathbf{E}T$ is finite, by Wald's Identity we therefore have that $\mathbf{E}S_T = \mathbf{E}T\mathbf{E}X_1 = 0$. We may then write

$$0 = \mathbf{E}S_T = \mathbf{E}\left\{S_T | Pos_r\right\}\mathbf{P}\left\{Pos_r\right\} + \mathbf{E}\left\{S_T | \overline{Pos_r}\right\}\mathbf{P}\left\{\overline{Pos_r}\right\}. \tag{2.10}$$

This equation yields an easy upper bound on $\mathbf{P}\left\{Pos_r\right\}$. By definition $\mathbf{E}\left\{S_T | Pos_r\right\} \geq r$, and by our assumption that the steps have absolute value at most $A$, we have $\mathbf{E}\left\{S_T | \overline{Pos_r}\right\} \geq -A$. Therefore

$$0 \geq r\mathbf{P}\left\{Pos_r\right\} - A\mathbf{P}\left\{\overline{Pos_r}\right\} = r\mathbf{P}\left\{Pos_r\right\} - A(1 - \mathbf{P}\left\{Pos_r\right\}),$$

and rearranging the latter inequality yields that $\mathbf{P}\left\{Pos_r\right\} \leq A/(r+A)$.

We may derive an upper bound of the same order for $\mathbf{P}\left\{Pos_r\right\}$ in a similar fashion; we first observe that $\mathbf{E}\left\{S_T | Pos_r\right\} < r+A$. Furthermore, since if $X_1 \leq 0$ then $\overline{Pos_r}$ occurs and $T = 1$, we have

$$\mathbf{E}\left\{S_T | \overline{Pos_r}\right\} \leq \mathbf{E}\left\{S_T \mathbf{1}_{[X_1 \leq 0]} | \overline{Pos_r}\right\} \leq \mathbf{E}\left\{S_1 \mathbf{1}_{[X_1 \leq 0]}\right\} = \mathbf{E}\left\{X_1 \mathbf{1}_{[X_1 \leq 0]}\right\}.$$

By our assumption that $X_1$ is symmetric, integer valued, and is never 0, this yields

$$\mathbf{E}\left\{S_T | \overline{Pos_r}\right\} \leq \mathbf{E}\left\{X_1 \mathbf{1}_{[X_1 < 0]}\right\} = \mathbf{E}\left\{X_1 | X_1 < 0\right\}\mathbf{P}\left\{X_1 < 0\right\} \leq -\mathbf{P}\left\{X_1 < 0\right\} = -\frac{1}{2}.$$

Combining (2.10) with our bounds on $\mathbf{E}\left\{S_T | Pos_r\right\}$ and $\mathbf{E}\left\{S_T | \overline{Pos_r}\right\}$, we thus have,

$$0 < (r+A)\mathbf{P}\left\{Pos_r\right\} - (1/2)\mathbf{P}\left\{\overline{Pos_r}\right\} = (r+A)\mathbf{P}\left\{Pos_r\right\} - (1/2)(1 - \mathbf{P}\left\{Pos_r\right\}),$$

which after rearrangement gives $\mathbf{P}\left\{Pos_r\right\} \geq 1/2(A + r + 1/2)$.

The events $Pos'_r$ and $Pos_r$ are closely related. More precisely, letting $S'$ be the "negative reverse" of $S$ defined by $S'_0 = 0$, $S'_{i+1} = S'_i - X_{n-i}$, the event $Pos'_r$ is just the event $Pos_r$ but with respect to the random walk $S'$. Since the steps of $S$ are symmetric, the random walks $S'$ and $S$ are identically distributed, so our bound on $\mathbf{P}\{Pos_r\}$ immediately gives

$$\mathbf{P}\{Pos'_r\} \geq \frac{1}{2(A+r+1/2)}. \tag{2.11}$$

Obervation 1 justifies moving from (2.4) to (2.5) and together, the bounds given in Observations 1 and 2 justify equation (2.7), establishing the validity of our derivation of (2.9). Finally, plugging the bound (2.11) into (2.9) yields that

$$\mathbf{P}\left\{Lead_n^+ | S_n = r\right\} = O\left(\frac{r}{n}\right), \tag{2.12}$$

as desired. Assuming we accept all the restrictions we have placed on our random variables and on the random walk $S$, the only step remaining to make the argument leading to (2.12) rigorous is a proof of the claims of Observation 1, i.e., that it takes order $r^2$ time to exit a strip of width $r$ and that there are strong upper tail bounds for this time.

In what follows we do *not* make the strong assumptions of this section on the distribution of $X$ and on the random walk $S$. The only general requirement from this point on is that $\mathbf{E}X = 0$ and $\mathbf{Var}\{X\} > 0$, i.e., $X$ is not a point mass; all other assumptions will be stated explicitly. We now turn our attention to the details.

## 2.2 The time to exit a strip.

For $r > 0$, we consider the first time $t$ for which $|S_t| \geq r$, denoting this time $T_r$. We prove

**Lemma 8.** *There is $A$ such that for all $r \geq 1$, $\mathbf{E}T_r \leq Ar^2$ and for all integers $k \geq 1$, $\mathbf{P}\{T_r \geq kAr^2\} \leq 1/2^k$.*

This is an easy consequence of a classical result on how "spread out" sums of independent identically distributed random variables become. The version we present can be found in Kesten (1972):

**Theorem 9.** *For any family of independent identically distributed real random variables $X_1, X_2, \ldots$ with positive, possibly infinite variance and associated partial sums $S_1, S_2, \ldots,$ there is a constant $c$ depending only on the distribution of $X_1$ such that for all $n$,*

$$\sup_{x \in \mathbb{R}} \mathbf{P}\{x \leq S_n \leq x+1\} \leq c/\sqrt{n}.$$

*Proof of Lemma 8.* Observe that the expectation bound follows directly from the probability bound, since if the probability bound holds then we have

$$\mathbf{E}T_r \leq \sum_{j=0}^{\infty} \mathbf{P}\{T_r \geq j\} \leq \sum_{i=0}^{\infty} \lceil Ar^2 \rceil \mathbf{P}\{T_r > i\lceil Ar^2 \rceil\} \leq \sum_{i=0}^{\infty} \frac{\lceil Ar^2 \rceil}{2^i} = 2\lceil Ar^2 \rceil,$$

which establishes the expectation bound with a slightly changed value of $A$. It thus remains to prove the probability bound. By Theorem 9, there is $c > 0$ (and we can and will assume $c > 1$) such that

$$\begin{aligned} \mathbf{P}\left\{|S_{\lceil 128c^2r^2 \rceil}| \leq 2r\right\} &\leq \sum_{i=\lfloor -2r \rfloor}^{\lfloor 2r \rfloor} \mathbf{P}\left\{i \leq S_{\lceil r^2 \rceil} \leq i+1\right\} \\ &\leq (4r+1)\frac{c}{\sqrt{\lceil 128c^2r^2 \rceil}} < \frac{1}{2}, \end{aligned} \tag{2.13}$$

the last inequality holding as $c > 1$ and $r > 1$. Let $t^* = \lceil 128c^2r^2 \rceil$ - then $\mathbf{P}\{T_r > t^*\} \leq 1/2$. We use this fact to show that for any positive integer $k$, $\mathbf{P}\{T_r > kt^*\} \leq 1/2^k$, which will establish the claim with $A = 128c^2 + 1$, for example. We proceed by induction on $k$, having just proved the claim for $k = 1$. We have

$$\begin{aligned}\mathbf{P}\{T_r > (k+1)t^*\} &= \mathbf{P}\{T_r > (k+1)t^* \cap T > kt\} \\ &= \mathbf{P}\{T_r > (k+1)t^* | T_r > kt^*\}\,\mathbf{P}\{T_r > kt\} \\ &= \frac{1}{2^k}\cdot \mathbf{P}\{T_r > (k+1)t^* | T_r > kt^*\},\end{aligned}$$

by induction. It remains to show that $\mathbf{P}\{T_r > (k+1)t^* | T_r > kt^*\} \leq 1/2$. If $T_r > kt^*$ then by the strong Markov property we may think of restarting the random walk at time $kt^*$. Whatever the value of $S_{kt^*}$, if the restarted random walk exits $[-2r, 2r]$ then the original random walk exits $[-r, r]$, so this inequality holds by (2.13). This proves the lemma. □

We may derive a stronger bound by taking two facts into consideration. First, by the assumption of positive variance, there are $v > 0$, $\epsilon > 0$ such that $\mathbf{P}\{|X| \geq v\} > \epsilon$. Second, the bound of Lemma 8 is crude in the sense that when we restart the random walk, it is not really necessary that the restarted walk exit the strip $[-2r, 2r]$ in order for the original random walk to exit $[-r, r]$. If $S_t = s$, say, and $-r < s < r$, then the restarted random walk need only reach the boundary of $[r-s, -(r+s)]$ in order for the original walk to exit $[-r, r]$. To make use of the latter fact, it is useful to study the stopping time $T_{r,s}$, the first time $t$ that $|S_t + s| \geq r$, for arbitrary $-r \leq s \leq r$ – bounds on this stopping time imply bounds on $T_r$ by taking $s = 0$. We can show:

**Lemma 10.** *If, for a given $v > 0$, $\epsilon > 0$, $\mathbf{P}\{|X| > v\} \geq \epsilon$,then for all $r \geq 0$ and all $s$ for which $0 \leq |s| \leq r$,*

$$\mathbf{E}T_{r,s} \leq \max\left\{\left(\frac{4r}{v}\right)^2, 1\right\}\cdot\frac{1}{\epsilon}.$$

The intuition of our proof is contained in the case where $S$ is a symmetric simple random walk. Let $T_i$ be the first time $|S_{T_i}| = 2^i$. In particular, we have $T_0 = \mathbf{E}T_0 = 1 = (2^0)^2$. For

$i > 0$, we have the simple recurrence

$$\mathbf{E}T_i = 2\mathbf{E}T_{i-1} + (1/2)\mathbf{E}T_i,$$

given by considering whether the walk returns to 0 after the first time it reaches absolute value $2^{i-1}$ and before it reaches absolute value $2^i$. Solving the recurrence yields $\mathbf{E}T_i = 4\mathbf{E}T_{i-1}$, so by induction $\mathbf{E}T_i = (2^i)^2$. For general $n$, if $U_n$ is the first time $|S_{U_n}| = n$ then letting $i$ be the smallest integer for which $2^i > n$, we have

$$\mathbf{E}U_n \leq \mathbf{E}T_i \leq (2^i)^2 \leq (2n)^2 = 4n^2. \tag{2.14}$$

We establish the general result in much the same manner, but because the steps of our random walk do not necessarily have size exactly 1, we must be more careful about what event will correspond to "returning to zero before first having absolute value $2^i$."

*Proof of Lemma 10.* First suppose $r \leq v/2$. For any $0 \leq |s| \leq r$, if $|S_{i-1}+s| < r$ and $X_i \geq v$ then $|S_i + s| \geq r$. It follows that

$$\mathbf{P}\{T_{r,s} > i\} \leq \mathbf{P}\{|X_j| < v \text{ for } 1 \leq j \leq i\} \leq (1-\epsilon)^i,$$

for all $i \geq 0$, so

$$\mathbf{E}T_{r,s} = \sum_{i=0}^{\infty} \mathbf{P}\{T_{r,s} > i\} \leq \sum_{i=0}^{\infty} (1-\epsilon)^i = 1/\epsilon, \tag{2.15}$$

which establishes the claim for $r \leq v/2$.

We next suppose the claim holds for $r$ and establish it for $2r > v/2$, i.e., we show that for such $r$, for any $0 \leq |s| \leq r$,

$$\mathbf{E}T_{2r,2s} \leq \left(\frac{4(2r)}{v}\right)^2 \cdot \frac{1}{\epsilon}.$$

Fix $0 \leq |s| \leq r$ arbitrarily. We consider the first time $t$ for which $|S_t + s| \geq r$, denoting this time $T_{r,s}^{(1)}$. We then consider the first time $t$ greater than $T_{r,s}^{(1)}$ for which $|S_t - S_{T_{r,s}^{(1)}} + s| \geq r$, denoting this time $T_{r,s}^{(2)}$. We let $E^{++}$ be the event that $S_{T_{r,s}^{(1)}} + s > 0$ and $S_{T_{r,s}^{(2)}} - S_{T_{r,s}^{(1)}} + s > 0$, and define $E^{+-}$, $E^{-+}$, and $E^{--}$ similarly (the superscripts indicating the sign of $S_{T_{r,s}^{(1)}} + s$ and of $S_{T_{r,s}^{(2)}} - S_{T_{r,s}^{(1)}} + s$, respectively). If $E^{++}$ or $E^{--}$ occurs, then $T_{2r,2s} \leq T_{r,s}^{(2)}$. We may thus write

$$\begin{aligned}
\mathbf{E}T_{2r,2s} &= \mathbf{E}\left\{T_{2r,2s}\mathbf{1}_{[E^{++}\cup E^{--}]}\right\} + \mathbf{E}\left\{T_{2r,2s}\mathbf{1}_{[E^{+-}\cup E^{-+}]}\right\} \\
&\leq \mathbf{E}\left\{T_{r,s}^{(2)}\mathbf{1}_{[E^{++}\cup E^{--}]}\right\} + \mathbf{E}\left\{T_{r,s}^{(2)}\mathbf{1}_{[E^{+-}\cup E^{-+}]}\right\} + \mathbf{E}\left\{(T_{2r,2s} - T_{r,s}^{(2)})\mathbf{1}_{[E^{+-}\cup E^{-+}]}\right\} \\
&= \mathbf{E}T_{r,s}^{(2)} + \mathbf{E}\left\{T_{2r,2s} - T_{r,s}^{(2)} | E^{+-}\cup E^{-+}\right\}\mathbf{P}\left\{E^{+-}\cup E^{-+}\right\}. \qquad (2.16)
\end{aligned}$$

By definition and by the strong Markov property,

$$\mathbf{E}T_{r,s}^{(2)} = \mathbf{E}T_{r,s}^{(1)} + \mathbf{E}\left\{T_{r,s}^{(2)} - T_{r,s}^{(1)}\right\} = 2\mathbf{E}T_{r,s}^{(1)} = 2\mathbf{E}T_{r,s}.$$

Also by the strong Markov property,

$$\mathbf{P}\left\{E^{+-}\right\} = \mathbf{P}\left\{E^{-+}\right\} = \mathbf{P}\left\{S_{T_{r,s}} + s \geq r\right\} \cdot \mathbf{P}\left\{S_{T_{r,s}} + s \leq -r\right\} \leq 1/4,$$

so $\mathbf{P}\left\{E^{+-}\cup E^{-+}\right\} \leq 1/2$. Combining these facts with (2.16) gives

$$\mathbf{E}T_{2r,2s} \leq 2\mathbf{E}T_{r,s} + (1/2)\mathbf{E}\left\{T_{2r,2s} - T_{r,s}^{(2)} | E^{+-}\cup E^{-+}\right\}. \qquad (2.17)$$

If $S_{T_{r,s}^{(2)}} = 2s'$, for some $0 \leq |s'| \leq r$, then the expected time before the random walk restarted at time $T_{r,s}^{(2)}$ reaches $r$ or $-r$ is $\mathbf{E}T_{2r,2s'}$. Therefore

$$\begin{aligned}
\mathbf{E}\left\{T_{2r,2s} - T_{r,s}^{(2)} | E^{+-}\cup E^{-+}\right\} &\leq \sup_{0\leq|s'|\leq r} \mathbf{E}\left\{T_{2r,2s} - T_{r,s}^{(2)} | S_{T_{r,s}^{(2)}} = 2s', E^{+-}\cup E^{-+}\right\} \\
&\leq \sup_{0\leq|s'|\leq r} \mathbf{E}T_{2r,2s'}. \qquad (2.18)
\end{aligned}$$

By (2.17) and (2.18) we have

$$\mathbf{E}T_{2r,2s} \leq 2\mathbf{E}T_{r,s} + (1/2) \sup_{0\leq|s'|\leq r} \mathbf{E}T_{2r,2s'}.$$

A supremum over $s$ gives

$$\sup_{0\leq|s|\leq r} \mathbf{E}T_{2r,2s} \leq 2 \sup_{0\leq|s|\leq r} \mathbf{E}T_{r,s} + (1/2) \sup_{0\leq|s'|\leq r} \mathbf{E}T_{2r,2s'},$$

which, after rearrangement, is

$$\sup_{0\leq|s|\leq r} \mathbf{E}T_{2r,2s} \leq 4 \sup_{0\leq|s|\leq r} \mathbf{E}T_{r,s} \leq 4\max\left\{\left(\frac{4r}{v}\right)^2, 1\right\} \cdot \frac{1}{\epsilon},$$

the second inequality holding by induction. Since $2r > v/2$, we have that $(4r/v)^2 > 1$. Therefore, $4\max\{(4r/v)^2, 1\} = (4(2r)/v)^2$. This completes the proof. □

For a specific variable $X$, we can optimize this bound by choosing $v$ and $\epsilon$ carefully. For example, if $X$ is a non-zero integer valued random variable then we may take $v = 1, \epsilon = 1$; the resulting bound is identical to the bound (2.14) we proved for the simple random walk.

We can use Lemma 10 and a renewal argument to derive bounds on the probability $T_{r,s}$ is large that correspond to those of Lemma 8:

**Corollary 11.** *For any integer $k > 0$,*

$$\mathbf{P}\left\{T_{r,s} > 2k\left\lceil \max\left\{\left(\frac{4r}{v}\right)^2, 1\right\} \cdot \frac{1}{\epsilon}\right\rceil\right\} \leq \frac{1}{2^k}.$$

*Proof.* We write $T$ and $t$ in place of $T_{r,s}$ and $\lceil \max\{(4r/v)^2, 1\}/\epsilon\rceil$ for simplicity. We proceed by induction on $k$, the claim holding for $k = 1$ by Lemma 10 and Markov's inequality. We suppose the claim holds for $k$. We have

$$\begin{aligned}\mathbf{P}\{T > 2(k+1)t\} &= \mathbf{P}\{T > 2(k+1)t \cap T > 2kt\} \\ &= \mathbf{P}\{T > 2(k+1)t | T > 2kt\}\,\mathbf{P}\{T > 2kt\} \\ &= \frac{1}{2^k}\cdot \mathbf{P}\{T > 2(k+1)t | T > 2kt\},\end{aligned}$$

by induction. Furthermore,

$$\mathbf{P}\{T > 2(k+1)t | T > 2kt\} \leq \sup_{0\leq |s'|\leq r} \mathbf{P}\{T > 2(k+1)t | T > 2kt, S_{2kt} = s'\},$$

and by restarting the random walk at time $2kt$ just as in the proof of Lemma 8, for any such $s'$ we have

$$\mathbf{P}\{T > 2(k+1)t | T > 2kt, S_{2kt} = s'\} = \mathbf{P}\{T_{r,s'} \geq 2t\} \leq 1/2.$$

Thus $\mathbf{P}\{T > 2(k+1)t | T > 2kt\} \leq 1/2$, so $\mathbf{P}\{T > 2(k+1)t\} \leq 1/2^{k+1}$. This completes the proof. □

This fact yields an elementary and rigorous justification of Bertrand's intuition that in a fair game in which a gambler starts with $r$ coins ($r$ is an integer and $r \geq 1$), the gambler will not likely last for much longer than $O(r^2)$ time. The event that the gambler is not ruined at time $t$, is nothing but the probability that a symmetric simple random walk $S$ started from 0 has not visited $-r$ by time $t$. Let the time that $S$ first visits $-r$ be called $T^*$ Let $S$ be an SSRW and consider the smallest value $T_1$ for which $|S_{T_1}| \geq r$ - as each step has absolute value 1, necessarily $S_{T_1} = \pm r$. By symmetry, $S_{T_1} = -r$ with probability 1/2. For an SSRW we observed that we may take $v = \epsilon = 1$, so by Corollary 11, for any positive integer $i$,

$$\mathbf{P}\left\{T_1 > 32ir^2)\right\} \leq 1/2^i.$$

Let $T_2$ be the first time after $T_1$ that $|S_{T_2} - S_{T_1}| \geq 2r$. An identical calculation yields that $\mathbf{P}\{T_2 - T_1 > 32i(2r)^2\} \leq 1/2^i$. Furthermore, $T^* > T_2$ precisely if $S_{T_1} = +r$ and $S_{T_2} - S_{T_1} = +2r$; both the latter events have probability 1/2, and they are independent by the strong Markov property. We define $T_3, \ldots, T_i$ in this manner, with $T_{j+1}$ the first time after $T_j$ that $|S_{T_{j+1}} - S_{T_j}| \geq 2^j r$. If $T^* > 32ir^2(\sum_{j=0}^{i-1} 2^{2j})$ then either

- $T_{j+1} - T_j > (2^{2j})32ir^2$ for some $0 \leq j \leq i-1$, or
- $S_{T_{j+1}} - S_{T_j} = 2^j r$ for *all* $0 \leq j \leq i-1$ (we let $T_0$ be 0 for convenience).

Since $32ir^2(\sum_{j=0}^{i-1} 2^{2j}) < 32i(4^i)r^2$, it is immediate from the above bounds that

$$\begin{aligned}\mathbf{P}\{T^* > 32i(4^i)r^2\} &\leq \mathbf{P}\{\cup_{j=0}^{i-1}\{T_{j+1} - T_j > (2^{2j})32ir^2\}\} + \mathbf{P}\{\cap_{j=0}^{i-1}\{S_{T_{j+1}} - S_{T_j} = 2^j r\}\} \\ &\leq \frac{i}{2^i} + \frac{1}{2^i} = \frac{i+1}{2^i}.\end{aligned}$$

This is not a strong bound! For example, if $i = \log r$ then we have only shown that $\mathbf{P}\{T^* > 32r^4 \log r\} = O(\log r/r)$. One way we could strengthen this bound is by noting that the variance of $T_{j+1} - T_j$ grows as $j$ grows, so it is much more likely that the later differences are large than the earlier ones. We could also strengthen it by being more careful about the precise contribution of each difference $T_{j+1} - T_j$ to $T^*$, rather than using the crude union bound above.

However, we should not hope for a much stronger bound: if $\mathbf{E}T^*$ were finite then by Wald's identity we would have $\mathbf{E}S_{T^*} = 0$, which is clearly nonsense as $S_{T^*}$ is deterministically equal to $-r$. Thus the expected value of $T^*$ is infinite, so $\sum_{k=0}^{\infty} \mathbf{P}\{T^* \geq i\}$ is infinite, thus these tail probabilities can not decay too quickly. More strongly, if Bertrand's back-of-the-envelope calculation from Section 1.1 is to be believed (and it is) then our upper bound is within a $(\log r)^{3/2}$ factor of the true asymptotics.

## 2.3 Bounding the exit probabilities

The doubling argument used above to prove Lemma 10 and in the discussion of Bertrand's gambler will be key in our proof of the more general ballot theorem, but we will bound the associated probabilities more carefully. For a general sequence of random variables $X_1, X_2, \ldots$ and associated random walk $S$, we can still define the times $T_1, T_2, \ldots$ just as above. Before, however, we were able to exploit the symmetry of the SSRW so that at each step, the probability we went below $2^i$ was exactly $1/2$. For the general random variables we are considering, this is not necessarily true. However, it does still hold by Wald's identity that $\mathbf{E}S_{T_1} = 0$, so

$$0 = \mathbf{E}S_{T_1} = \mathbf{E}\{S_{T_1}|S_{T_1} \geq r\}\mathbf{P}\{S_{T_1} \geq r\} + \mathbf{E}\{S_{T_1}|S_{T_1} \leq -r\}\mathbf{P}\{S_{T_1} \leq -r\}. \qquad (2.19)$$

As discussed in Section 2.1, if we know that $\mathbf{E}\{S_{T_1}|S_{T_1} \geq r\}$ and $-\mathbf{E}\{S_{T_1}|S_{T_1} \leq -r\}$ are *close to* $r$, then it follows that $|\mathbf{P}\{S_{T_1} \geq r\} - \mathbf{P}\{S_{T_1} \leq -r\}|$ is small, so both are close to $1/2$.

The quantity $\mathbf{E}\{|S_{T_1}| - r\}$ is called the *overshoot* at $r$. Griffin and McConnell (1992) have considered the size of the overshoot in a very general setting; we proceed to explain those of their results which concern us. A random variable $X$ for for which $\mathbf{E}X = 0$ is in the *domain of attraction of the normal distribution* ($X \in \mathcal{D}$ for short) if there is a sequence $\{a_n\}_{n>0}$ for which $S_n/a_n$ converges in probability to a normally distributed random variable. We say $X$ is a *weak* $L^p$ random variable ($X \in WL^p$ for short) if $\mathbf{P}\{|X| > r\} = O(1/r^p)$. For $r > 0$, let $T_r = \min\{t \mid |S_t| \geq r\}$.

**Theorem 12** (Griffin and McConnell (1992)). *If* $X \in \mathcal{D}$ *and* $\mathbf{E}X = 0$ *then* $\mathbf{E}\{|S_{T_r}| - r\} = o(r)$. *If in addition,* $X \in WL^{2+\alpha}$ *for some* $0 < \alpha < 1$ *then* $\mathbf{E}\{|S_{T_r}| - r\} = O(r^{1-\alpha})$. *Finally, if in addition* $\mathbf{E}\{|X|^3\} < \infty$ *then* $\mathbf{E}\{|S_{T_r}| - r\} = O(1)$.

Griffin and McConnell also show that all the hypotheses above are necessary for the respective conclusions to hold. This paper also contains a proof of the fact that if $X \notin \mathcal{D}$ then $\mathbf{E}\{|S_{T_r}| - r\} \neq O(r)$, so, surprisingly, no matter what the distribution of $X$, either $\mathbf{E}\{|S_{T_r}| - r\} = o(r)$ or $\mathbf{E}\{|S_{T_r}| - r\} \neq O(r)$.

For $0 \leq |s| < r$ letting $T_{r,s}$ be the first time that $|S_t + s| > r$ as previously, the proofs from Griffin and McConnell (1992) yield without modification the conclusions of Theorem 12 with $\mathbf{E}\left\{|S_{T_{r,s}} + s| - r\right\}$ in place of $\mathbf{E}\{|S_{T_r}| - r\}$, and uniformly in $s$.

To avoid reference to the three separate conditions appearing in Theorem 12 in what follows, we introduce the following notation $O_X(r, f)$:

$$O_X(r,f) = \begin{cases} \infty & \text{if } X \notin \mathcal{D} \text{ or } \mathbf{E}X \neq 0 \\ o(rf) & \text{if } X \in \mathcal{D} \text{ and } \mathbf{E}X = 0, \\ O(r^{1-\alpha}f) & \text{if } 0 < \alpha < 1,\ X \in WL^{2+\alpha} \text{ and } \mathbf{E}X = 0, \text{ and} \\ O(f) & \text{if } \mathbf{E}\{|X|^3\} < \infty \text{ and } \mathbf{E}X = 0. \end{cases}$$

Combining Theorem 12 with (2.19), we can now easily prove bounds on the probability that $S + s$ exits $[-r, r]$ in the negative direction.

**Theorem 13.** *For* $0 \leq |s| < r$,

$$\mathbf{P}\left\{S_{T_{r,s}} < 0\right\} = \frac{r - s + O_X(r,1)}{2r}.$$

*Proof.* We have

$$\begin{aligned} \mathbf{E}\left\{S_{T_{r,s}} \mathbf{1}_{[S_{T_{r,s}} > 0]}\right\} &= \mathbf{E}\left\{(S_{T_{r,s}} + s)\mathbf{1}_{[S_{T_{r,s}} > 0]}\right\} - s\mathbf{P}\left\{S_{T_{r,s}} > 0\right\} \\ &= \mathbf{E}\left\{(|S_{T_{r,s}} + s| - r)\mathbf{1}_{[S_{T_{r,s}} > 0]}\right\} + (r - s)\mathbf{P}\left\{S_{T_{r,s}} > 0\right\} \\ &= O_X(r,1) + (r - s)\mathbf{P}\left\{S_{T_{r,s}} > 0\right\}. \end{aligned}$$

by Theorem 12. Similarly,

$$\begin{aligned}
\mathbf{E}\left\{S_{T_{r,s}}\mathbf{1}_{[S_{T_{r,s}}<0]}\right\} &= \mathbf{E}\left\{(S_{T_{r,s}}+s)\mathbf{1}_{[S_{T_{r,s}}<0]}\right\} - s\mathbf{P}\left\{S_{T_{r,s}}<0\right\} \\
&= -\mathbf{E}\left\{|S_{T_{r,s}}+s|\mathbf{1}_{[S_{T_{r,s}}<0]}\right\} - s\mathbf{P}\left\{S_{T_{r,s}}<0\right\} \\
&= -\mathbf{E}\left\{(|S_{T_{r,s}}+s|-r)\mathbf{1}_{[S_{T_{r,s}}<0]}\right\} - (r+s)\mathbf{P}\left\{S_{T_{r,s}}<0\right\} \\
&= O_X(r,1) - (r+s)\mathbf{P}\left\{S_{T_{r,s}}<0\right\}.
\end{aligned}$$

Thus, by Wald's Identity,

$$\begin{aligned}
0 &= \mathbf{E}S_{T_{r,s}} = \mathbf{E}\left\{S_{T_{r,s}}\mathbf{1}_{[S_{T_{r,s}}>0]}\right\} + \mathbf{E}\left\{S_{T_{r,s}}\mathbf{1}_{[S_{T_{r,s}}<0]}\right\} \\
&= (r-s)\mathbf{P}\left\{S_{T_{r,s}}>0\right\} - (r+s)\mathbf{P}\left\{S_{T_{r,s}}<0\right\} + O_X(r,1) \\
&= (r-s)\left(1-\mathbf{P}\left\{S_{T_{r,s}}>0\right\}\right) - (r+s)\mathbf{P}\left\{S_{T_{r,s}}<0\right\} + O_X(r,1) \\
&= (r-s) - 2r\mathbf{P}\left\{S_{T_{r,s}}>0\right\} + O_X(r,1).
\end{aligned}$$

The claim follows. □

When $r-|s|$ is much larger than $O_X(r,1)$, the value $(r-s+O_X(r,1))/2r$ of Theorem 13 is $\Theta((r-s)/2r)$, which agrees with the intuition given by the symmetric simple random walk. The doubling argument will allow us to strengthen Theorem 13 when $|s|$ is closer to $r$ – if $\mathbf{E}X=0$ and $X \in WL^{2+\alpha}$ for some $\alpha>0$, we can prove bounds of order $(r-s)/2r$ as long as $r-|s| = \Omega(1)$. When $\mathbf{E}X=0$ and $X \in \mathcal{D}$, we will achieve such bounds as long as $s/r = o(1)$.

**Lemma 14.** *Suppose $X \in WL^{2+\alpha}$ for some $0<\alpha\leq 1$. Then there are $c>1$, $r_0>0$ such that for all $r \geq r_0$, for all $0 \leq s \leq r-r_0$,*

$$\mathbf{P}\left\{S_{T_{r,s}}<0\right\} \geq \frac{(r-s)}{16r},$$

*and for all* $r - r_0 < s < r$

$$\mathbf{P}\left\{S_{T_{r,s}} < 0\right\} \geq \frac{1}{cr}.$$

We prove this lemma via a straightforward application of the doubling argument, using Theorem 13 to bound the probability we go positive at each step. We will eventually show that upper bounds of the same order hold; the upper bounds also follow from a doubling argument, but are a little more work to prove.

*Proof of Lemma 14.* We first consider the case that $0 \leq s \leq r - r_0$ (we will choose the value $r_0$ in the course of the proof). We recall that $T_{r,s}$ is the first time $t$ that $|S_t + s| \geq r$, i.e. it is the first time that either $S_t \geq r - s$ or that $S_t \leq -(r+s)$. We let $u = r - s$ and $d = r + s$; note that $u > r_0$ by assumption and $d > u$ since $s > 0$.

Let $T_0 = 0$. For $i \geq 0$, let $T_{i+1}$ be the smallest $t > T_i$ for which $|S_t - S_{T_i}| \geq 2^i u$, and let $E_{i+1}$ be the event that $S_{T_{i+1}} < S_{T_i}$. Finally, let $j^* = \lceil \log(d/u) \rceil + 1$. If $\cap_{i=1}^{j^*} E_i$ occurs then $S_i < u$ for all $i \leq T_{j^*}$, and

$$S_{T_{j^*}} \leq -\sum_{i=1}^{j^*}(2^{i-1}u) = -(2^{j^*} - 1)u \leq -(2d/u - 1)u < -d = -(r+s),$$

so $T_{r,s} \leq T_j^*$ and $S_{T_{r,s}} \leq -d < 0$. It follows that $\mathbf{P}\left\{S_{T_{r,s}} < 0\right\} \geq \mathbf{P}\left\{\cap_{i=1}^{j^*} E_i\right\} = \prod_{i=1}^{j^*} \mathbf{P}\left\{E_i\right\}$, the latter equality holding as the $E_i$ are determined on disjoint sections of the random walk. For a given $i \geq 1$, $E_i$ is the event that the walk $S$ restarted at time $T_{i-1}$ dips below $-2^{i-1}u$ before exceeding $2^{i-1}u$. By Theorem 13, it follows that

$$\mathbf{P}\left\{E_i\right\} = \frac{2^{i-1}u + O_X(2^{i-1}u, 1)}{2^i u} \geq \frac{1}{2} - \frac{c_1}{(2^i u)^\alpha}, \tag{2.20}$$

for some constant $c_1 > 0$. We thus have

$$\mathbf{P}\left\{S_{T_{r,s}} < 0\right\} \geq \prod_{i=1}^{j^*} \mathbf{P}\left\{E_i\right\} \geq \prod_{i=1}^{j^*} \left(\frac{1}{2} - \frac{c_1}{(2^i u)^\alpha}\right) \geq \prod_{i=1}^{j^*} \left(\frac{1}{2} - \frac{c_1}{2^{\alpha i} \cdot r_0^\alpha}\right) \tag{2.21}$$

$$\geq \frac{1}{2^{j^*+1}} \geq \frac{u}{8d} \geq \frac{r-s}{16r}, \tag{2.22}$$

(2.22) following from (2.21) as long as we choose $r_0$ large enough. This proves the first claim of the theorem. Observe that in particular, this implies that for all $-r < s \leq r - r_0$,

$$\mathbf{P}\left\{S_{T_{r,s}} < 0\right\} \geq \mathbf{P}\left\{S_{T_{r,r-r_0}} < 0\right\} \geq \frac{r_0}{16r}. \tag{2.23}$$

If $r - r_0 < s < r$, Let $T$ be the first time $t$ that $S_t + s \geq r$ or $S_t + s \leq r - r_0$. By restarting the random walk at time $T$ and applying (2.23), we have

$$\mathbf{P}\left\{S_{T_{r,s}} < 0\right\} \geq \mathbf{P}\left\{S_T < 0\right\} \mathbf{P}\left\{S_{T_{r,s}} < 0 | S_T < 0\right\} \geq \mathbf{P}\left\{S_t < 0\right\} \cdot \frac{r_0}{16r}. \tag{2.24}$$

Finally, since $\mathbf{E}X = 0$ and $\mathbf{Var}\{X\} > 0$, there are $v > 0$, $\epsilon > 0$ such that $\mathbf{P}\{X < -v\} \geq \epsilon$. Thus

$$\mathbf{P}\left\{S_T < 0\right\} \geq \mathbf{P}\left\{\cap_{i=1}^{\lceil r_0/v \rceil} \{X_i < -v\}\right\} \geq \epsilon^{\lceil r_0/v \rceil}. \tag{2.25}$$

The second claim of the theorem follows from (2.24) and (2.25) by taking $c = r_0 \epsilon^{\lceil r_0/v \rceil}/16$. □

In the above proof, the only place where we use the fact that $X \in WL^{2+\alpha}$ is in our bound for $\mathbf{P}\{E_i\}$. If we replace the assumption that $X \in WL^{2+\alpha}$ by the assumption that $X \in \mathcal{D}$, then $O_X(r, 1) = o(r)$, so in (2.20) we can only conclude that $\mathbf{P}\{E_i\} \geq 1/2 - \epsilon$, where we can make $\epsilon$ arbitrarily small by choosing $r_0$ large. Following this chain of reasoning we can prove the following lemma for the case $X \in \mathcal{D}$; the details are omitted.

**Lemma 15.** *Suppose $X \in \mathcal{D}$ for some $0 < \alpha \leq 1$. Then for all $\epsilon > 0$ there is $r_0 > 0$ such* *that for all* $r \geq r_0$*, for all* $0 \leq s \leq r - r_0$*,*

$$\mathbf{P}\left\{S_{T_{r,s}} < 0\right\} \geq \frac{(r-s)}{r^{1+\epsilon}},$$

*and for all* $r - r_0 \leq s < r$*,*

$$\mathbf{P}\left\{S_{T_{r,s}} < 0\right\} \geq \frac{1}{r^{1+\epsilon}}.$$

In proving corresponding upper bounds for Lemmas 14 and 15, we must consider the overshoot much more carefully because it has a potentially significant effect on the doubling argument. Suppose we have defined $T_1$ to be the first time $t$ that $|S_t| \geq k$ (for some $k$) and $T_2$ to be the first time $t > T_1$ that that $|S_t - S_{T_1}| > 2k$. If it happens that $S_{T_1} \geq 2k$ - i.e., $S$ exits $[-k, k]$ in the negative direction and has overshoot at least $k$ - then the fact that $S_{T_2} - S_{T_1}$ is positive does *not* imply that $S_{T_2}$ is positive. Similarly, if the *cumulative* overshoot after many steps eventually exceeds $k$, then we can no longer conclude that a positive step of the doubling process implies that $S$ exceeds zero.

To deal with this difficulty, we need to modify our stopping times. There are two natural candidates for the definition of $T_2$, for example, which ensure that if $S_{T_2} - S_{T_1} > 0$ then $S_{T_2} \geq k$: we could let $T_2$ be the first time $t > T_1$ that $|S_t - S_{T_1}| \geq |S_{T_1}| + k$; or, we could let $T_2$ be the first time $t \geq T_1$ that $|S_{T_2} - k| \geq 2k$. Either approach introduces technicalities to our proof. In the first case, the strip boundaries become random and we do not know precisely how many doublings there will be before the upper boundary of the strip has height $r$. In the second case, when we begin a restarted random walk we do not necessarily start from the center of the strip, and it is possible that there are "degenerate" doublings, in the sense that $T_{i+1} = T_i$.

Which approach we adopt changes the precise technicalities but not the essence of the proof – we choose the latter as we have already developed some tools for analyzing exit times for "non-centered" random walks. We prove:

**Lemma 16.** *Suppose $X \in WL^{2+\alpha}$ for some $0 < \alpha \le 1$. Then there are $r_0 > 0$, $c > 0$ such that for all $r \ge r_0$, for all $0 \le s \le r - r_0$,*

$$\mathbf{P}\left\{S_{T_{r,s}} < 0\right\} \le \frac{4(r-s)}{r},$$

*and for all $r - r_0 < s < r$,*

$$\mathbf{P}\left\{S_{T_{r,s}} < 0\right\} \le \frac{c}{r}.$$

In proving this lemma, we will use the following fact, which is an easy consequence of Theorem 12; we omit the proof.

**Fact 17.** *Suppose $\mathbf{E}\{X\} = 0$ and $X \in WL^{2+\alpha}$ for some $0 < \alpha < 1$. Then for all $\epsilon, \delta > 0$ and $0 < \beta < \alpha$ there is $r_0 > 0$ such that for all $r \ge r_0$ and for all $0 \le |s| < r$,*

$$\mathbf{P}\left\{|S_{T_{r,s}} + s| - r \in [\delta r^{1-\beta}, r^{1-\beta}]\right\} \le \epsilon, \tag{2.26}$$

*and for all $i \ge 0$,*

$$\mathbf{P}\left\{|S_{T_{r,s}} + s| - r \in [2^i r^{1-\beta}, 2^{i+1} r^{1-\beta}]\right\} \le \epsilon/2^i \tag{2.27}$$

*Proof of Lemma 16.* As in the proof of Lemma 14, the second claim of Lemma 16 follows easily from the first; we therefore restrict our attention to proving the first claim.

We will use the doubling argument inductively; we begin with a sketch of our approach. We let $u = r - s$ and $d = r + s$; note that $d > u$ since $s > 0$. For $i \ge 1$, let $m_i = 2^{i-1}u$ and let $v_i = (2^i - 1)u = \sum_{j=1}^{i} m_i$. We define a sequence of stopping times by setting $T_0 = 0$ and, for $i \ge 1$, letting $T_i$ be the first time $t > T_{i-1}$ that $S_t < -v_i$ or $S_t \ge u$. (Notice that if ever $S_{T_i} \ge u$ then $T_j = T_i$ for all $j \ge i$. It is also possible that $T_{i+1} = T_i$ if, for example, the first time that $S_t > v_i$ is also the first time $S_t > v_{i+1}$.)

Let $F_0$ be the event that $S_0 = 0$, and for $i > 0$ let $F_i$ be the event that $S_{T_i} < -v_i$. (The event $F_i$ is a rough analog of the event $\cap_{j=1}^{i} E_j$ from the proof of Lemma 14). Since if ever $S_{T_i} \geq u$ then $S_{T_j} \geq u$ for all $j > i$, the $F_i$ form a decreasing sequence of events.

We find an increasing sequence $\Delta_i$ for which $\Delta_i \leq 4$ for all $i$, and prove inductively that $\mathbf{P}\{F_i\} \leq \Delta_i/2^i$. Letting $i^* = \lfloor \log(d/u) \rfloor$, if $S_{T_{r,s}} < 0$ then $F_{i^*}$ must occur; it follows that

$$\mathbf{P}\left\{S_{T_{r,s}} < 0\right\} \leq \frac{\Delta_{i^*}}{2^{i^*}} \leq \frac{4u}{d} \leq \frac{4(r-s)}{r}.$$

Of course, we would like to simply take $\Delta_j = 4$ for all $j$, but we will have to choose a little more carefully to make the induction work. The difficulty is that if $F_1$ occurs, say, and $|S_{T_1}| - v_1$ is extremely large then the probability of $F_2$ is much larger than $1/2$. In order to better control this event, we have to use the bounds provided by Fact 17 on the probability of a large overshoot.

For the reason discussed above, we will in fact apply induction not only to bound the probability $F_i$ occurs, but to the bound the probability that $F_i$ occurs and *in addition* the overshoot is large. To make this statement more precise, we fix $0 < \beta < \alpha$, then fix some $0 < \delta < 1$ and define the following events:

$$\begin{aligned}
A_i \quad &\text{is the event} \quad F_i \cap \{|S_{T_i}| - v_i \in (0, \delta m_i^{1-\beta}]\} \\
B_i \quad &\text{is the event} \quad F_i \cap \{|S_{T_i}| - v_i \in (\delta m_i^{1-\beta}, m_i^{1-\beta}]\} \\
D_{i,j} \quad &\text{is the event} \quad F_i \cap \{|S_{T_i}| - v_i \in (2^j m_i^{1-\beta}, 2^{j+1} m_i^{1-\beta}]\} \text{ for } j \geq 0
\end{aligned}$$

(We mention that the restriction $0 < \delta < 1$ is the *only* restriction on $\delta$.) We observe that if $F_i$ occurs then either $A_i$ or $B_i$ or one of the $D_{i,j}$ must occur. We let $\gamma = 1/10$; continuing to postpone the definition of the sequence $\Delta_i$, we will in fact prove inductively that for all

$i \geq 1$,

$$\mathbf{P}\{F_i\} \leq \frac{\Delta_i}{2^i},$$
$$\mathbf{P}\{B_i\} \leq \frac{\gamma\Delta_i}{2^i}, \text{ and} \tag{2.28}$$
$$\mathbf{P}\{D_{i,j}\} \leq \frac{\gamma\Delta_i}{2^{j+i}}, \text{ for all } j = 0, 1, 2, \ldots$$

Of course, the first of these inequalities establishes our claim. We state at the outset that $\Delta_1 = 2$. As a consequence, the bound on $\mathbf{P}\{F_1\}$ holds trivially.

We make four requirements on the size of $r_0$. First, we let $\epsilon = \gamma/4 = 1/40$ and insist that $r_0$ is large enough that for all $r \geq r_0$ and all $0 \leq s \leq r$, $\mathbf{P}\left\{|S_{T_{r,s}} + s| - r \in (\delta r^{1-\beta}, r^{1-\beta}]\right\} \leq \epsilon$, and for all $j \geq 0$, $\mathbf{P}\left\{|S_{T_{r,s}} + s| - r \in (2^j r^{1-\beta}, 2^{j+1 r^{1-\beta}}]\right\} \leq \epsilon/2^j$; such an $r_0$ exists by Fact 17. In particular, since $u \geq r_0$ this establishes the bound on $\mathbf{P}\{B_1\}$ and the bounds on the $\mathbf{P}\{D_{1,j}\}$, so the base case of our induction holds.

Second, we insist that $r_0$ is large enough that for all $r \geq r_0$ and all $s \leq r - r_0$,

$$\left|\mathbf{P}\left\{S_{T_{r,s}} < 0\right\} - \frac{r-s}{2r}\right| \leq \frac{1}{r^\beta}; \tag{2.29}$$

such a choice exists by Theorem 13. Third, we insist that $r_0 > 2^{3/\beta}$, i.e., that $\beta \log r_0 > 3$. Fourth, we insist that $3\gamma(i + \log r) + 12 \leq i + \log r$ for all integers $i \geq 0$ and for all $r \geq r_0$. (With our choice $\gamma = 1/10$, this inequality is easily seen to hold as long as we choose $r_0 \geq 2^{18}$. This requirement may seem to arise out of thin air – it will be used in bounding a sum at the end of the proof – but we state it here in order that all our bounds on $r_0$ appear in the same place.)

We first argue inductively for the bound on $\mathbf{P}\{B_i\}$. Let $x = S_{T_i} + v_i$ – if $F_i$ occurs then $x$ is the overshoot at the $i$'th doubling. The essence of our argument is that if $x$ is small then for $B_{i+1}$ to occur, the random walk restarted at time $T_i$ must again have a large overshoot, which we know is unlikely. Furthermore, since $x$ is the overshoot of a previous step, we can bound the probability that $x$ is large by induction.

Observe that $B_{i+1} \subset F_{i+1} \subset F_i$, so in particular $\mathbf{P}\{B_{i+1}\} = \mathbf{P}\{B_{i+1}, F_i\}$. If $F_i$ occurs then $x < 0$, and if $-m_{i+1} < x < 0$ then $-v_{i+1} \leq x - v_i = S_{T_i} \leq -v_i$. By restarting the random walk at time $T_i$, and applying the bounds from Fact 17, we thus have

$$\begin{aligned}\mathbf{P}\{B_{i+1} \mid F_i, -m_{i+1} < x < 0\} &\leq \sup_{-m_{i+1} \leq x \leq 0} \mathbf{P}\left\{|S_{T_{m_{i+1},x}} + x| - m_{i+1} \in (\delta m_{i+1}^{1-\beta}, m_{i+1}^{1-\beta}]\right\} \\ &\leq \epsilon \leq \frac{\gamma}{4}.\end{aligned}$$

Next, if $F_i$ occurs and $-(m_{i+1} + m_{i+1}^{1-\beta}) \leq x \leq -m_{i+1}$, then letting $j^* = \lfloor \beta \log m_i \rfloor$, either $D_{i,j^*}$ or $D_{i,j^*+1}$ must have occurred. (This fact is a straightforward consequence of the definitions of the $D_{i,j}$.) It follows that

$$\mathbf{P}\left\{B_{i+1}, F_i, -(m_{i+1} + m_{i+1}^{1-\beta}) \leq x \leq -m_{i+1}\right\} \leq \mathbf{P}\{D_{i,j^*}\} + \mathbf{P}\{D_{i,j^*+1}\}.$$

By our choice of $r_0$, we have $\beta \log m_i \geq \beta \log u \geq \beta \log r_0 \geq 3$, so in particular $j^* \geq 3$. Finally, if $F_i$ occurred and $x < -(m_{i+1} + m_{i+1}^{1-\beta})$ then $B_{i+1}$ can not occur – the overshoot at the $i$'th step was so large we "jumped over" the interval $(v_{i+1} + \delta m_{i+1}^{1-\beta}, v_{i+1} + m_{i+1}^{1-\beta}]$. Thus $\mathbf{P}\left\{B_{i+1}, F_i, m_{i+1} + m_{i+1}^{1-\beta} \leq -x\right\} = 0$.

Combining these bounds using Bayes' formula and applying (2.28) inductively, we have

$$\begin{aligned}
\mathbf{P}\{B_{i+1}\} &= \mathbf{P}\{B_{i+1}, F_i\} \\
&\leq \mathbf{P}\{B_{i+1}|F_i, -m_{i+1} < x < 0\}\,\mathbf{P}\{F_i, -m_{i+1} < x < 0\} \\
&\quad + \mathbf{P}\left\{B_{i+1}, F_i, -(m_{i+1} + m_{i+1}^{1-\beta}) \leq x \leq -m_{i+1}\right\} \\
&\quad + \mathbf{P}\left\{B_{i+1}, F_i, x < -(m_{i+1} + m_{i+1}^{1-\beta})\right\} \\
&\leq \frac{\gamma}{4}\mathbf{P}\{F_i\} + \mathbf{P}\{D_{i,j}\} + \mathbf{P}\{D_{i,j+1}\} + 0 \\
&\leq \frac{\gamma}{4}\frac{\Delta_i}{2^i} + \frac{\gamma\Delta_i}{2^{i+3}} + \frac{\gamma\Delta_i}{2^{i+4}} < \frac{\gamma\Delta_i}{2^{i+1}} \leq \frac{\gamma\Delta_{i+1}}{2^{i+1}}, \qquad (2.30)
\end{aligned}$$

by our assumption that the sequence $\Delta_i$ is increasing. This establishes the inductive step of the bound for $\mathbf{P}\{B_{i+1}\}$ – the inductive argument for bounding the probabilities $\mathbf{P}\{D_{i+1,j}\}$ is essentially the same, and we omit it. We now turn to the inductive step of the bound for $\mathbf{P}\{F_{i+1}\}$.

As above, let $x = S_{T_i} + v_i$. Suppose that $F_i$ occurs, and $-m_{i+1} < x < 0$. In this case $F_{i+1}$ is the event that the first time $j$ after $T_i$ that $|(S_j - S_{T_i}) + x| \geq m_{i+1}$, we have $(S_j - S_{T_i}) + x \leq -m_{i+1}$. Note that if in fact $A_i$ occurs, we also know that $-\delta m_i^{1-\beta} \leq x < 0$. By the strong Markov property and by (2.29), we thus have

$$\begin{aligned}
\mathbf{P}\{F_{i+1}|A_i\} &\leq \sup_{-\delta m_i^{1-\beta} \leq x < 0} \mathbf{P}\left\{S_{T_{m_{i+1},x}} < 0\right\} \leq \sup_{-\delta m_i^{1-\beta} \leq x < 0} \frac{m_{i+1} - x}{2m_{i+1}} + m_{i+1}^{-\beta} \\
&\leq \frac{2^i u + \delta m_{i+1}^{1-\beta}}{2 \cdot 2^i u} + (2^i u)^{-\beta} \\
&< \frac{1}{2} + \frac{\delta (2^i u)^{1-\beta}}{2^i u} + (2^i u)^{-\beta} = \frac{1}{2}\left(1 + \frac{2+\delta}{(2^i u)^\beta}\right). \qquad (2.31)
\end{aligned}$$

Letting $a_i = 1 + (2+\delta)(2^i u)^{-\beta}$, we thus have

$$\mathbf{P}\{F_{i+1}|A_i\} \leq \frac{a_i}{2}, \qquad (2.32)$$

Letting $b_i = 1+3/(2^i u)^\beta$ and mimicking the above calculation leading to (2.31) and to (2.32) yields that

$$\mathbf{P}\{F_{i+1}|B_i\} \;\leq\; \frac{b_i}{2}. \tag{2.33}$$

Similarly, letting $d_{i,j} = 1 + (2+2^j)/(2^i u)^\beta$ for $j = 0, 1, 2, \ldots$, we have

$$\mathbf{P}\{F_{i+1}|D_{i,j}\} \;\leq\; \frac{d_{i,j}}{2}. \tag{2.34}$$

We remark that $a_i \leq b_i \leq d_{i,0} \leq d_{i,1} \leq \ldots$. This is as we expect: if the $i$'th step is negative, then the larger the overshoot in step $i$, the more likely we go negative in step $i+1$. Equation (2.34) may seem a little strange, as once $j$ is large conditioning on $D_{i,j}$ may tell us that the overshoot $x$ is greater than $m_{i+1}$, in which case $F_{i+1}$ occurs with probability 1. However, in this case $d_{i,j} \geq 2$, so (2.34) is still valid. In fact, letting $d^*_{i,j} = \min\{d_{i,j}, 2\}$, we may replace the term $d_{i,j}$ in (2.34) by $d^*_{i,j}$ and the equation remains valid.

We can now bound $\mathbf{P}\{F_{i+1}\}$ using Bayes' formula and (2.32)-(2.34):

$$\begin{aligned}
\mathbf{P}\{F_{i+1}\} &= \mathbf{P}\{F_{i+1}\cap F_i\} = \mathbf{P}\{F_{i+1}\cap A_i\} + \mathbf{P}\{F_{i+1}\cap B_i\} + \sum_{j=0}^{\infty}\mathbf{P}\{F_{i+1}\cap D_{i,j}\} \\
&= \mathbf{P}\{F_{i+1}|A_i\}\mathbf{P}\{A_i\} + \mathbf{P}\{F_{i+1}|B_i\}\mathbf{P}\{B_i\} + \sum_{j=0}^{\infty}\mathbf{P}\{F_{i+1}|D_{i,j}\}\mathbf{P}\{D_{i,j}\} \\
&\leq \frac{a_i}{2}\mathbf{P}\{A_i\} + \frac{b_i}{2}\mathbf{P}\{B_i\} + \sum_{j=0}^{\infty}\frac{d^*_{i,j}}{2}\mathbf{P}\{D_{i,j}\}. \qquad (2.35)
\end{aligned}$$

Since $\mathbf{P}\{A_i\} + \mathbf{P}\{B_i\} + \sum_{j=0}^{\infty}\mathbf{P}\{D_{i,j}\} = \mathbf{P}\{F_i\}$, (2.35) is equivalent to the statement that

$$\mathbf{P}\{F_{i+1}\} \leq \frac{a_i}{2}(\mathbf{P}\{F_i\} - \mathbf{P}\{B_i\} - \sum_{j=0}^{\infty}\mathbf{P}\{D_{i,j}\}) + \frac{b_i}{2}\mathbf{P}\{B_i\} + \sum_{j=0}^{\infty}\frac{d^*_{i,j}}{2}\mathbf{P}\{D_{i,j}\}. \tag{2.36}$$

We bound (2.36) by inductively applying (2.28) to bound $\mathbf{P}\{F_i\}, \mathbf{P}\{B_i\}$, and the $\mathbf{P}\{D_{i,j}\}$.

As $a_i \le b_i \le d^*_{i,0} \le d^*_{i,1} \le \ldots$, (2.36) is weakest if the bounds of (2.28) are tight. We thus have

$$\begin{aligned} \mathbf{P}\{F_{i+1}\} &\le \frac{a_i}{2}\frac{(1-3\gamma)\Delta_i}{2^i} + \frac{b_i}{2}\frac{\gamma\Delta_i}{2^i} + \sum_{j=0}^{\infty}\frac{d^*_{i,j}}{2}\frac{\gamma\Delta_i}{2^{i+j}} \\ &= \frac{\Delta_i}{2^{i+1}}\left((1-3\gamma)a_i + \gamma b_i + \sum_{j=0}^{\infty}\frac{\gamma d^*_{i,j}}{2^j}\right), \end{aligned} \tag{2.37}$$

Finally, we use (2.37) to define $\Delta_{i+1}$, by setting

$$C_i = \left((1-3\gamma)a_i + \gamma b_i + \sum_{j=0}^{\infty}\frac{\gamma d^*_{i,j}}{2^j}\right), \tag{2.38}$$

and letting $\Delta_{i+1} = \Delta_i \cdot \max\{C_i, 1\}$. This definition completes the inductive bound for $\mathbf{P}\{F_{i+1}\}$. The sequence $\Delta_i$ is certainly increasing – to complete the proof it remains to show that this sequence is bounded above by 4. Since $\Delta_1 = 2$, this holds if $\prod_{i=1}^{\infty}\max\{C_i, 1\} \le 2$.

We recall the definitions of $a_i, b_i$, and the $d^*_{i,j}$:

$$a_i = 1 + \frac{2+\delta}{(2^i u)^\beta}, \quad b_i = 1 + \frac{3}{(2^i u)^\beta}, \quad \text{and} \quad d^*_{i,j} = 1 + \min\left\{\frac{2+2^j}{(2^i u)^\beta}, 1\right\}.$$

Collecting terms in (2.38), we have

$$C_i = 1 + \frac{(2+\delta)}{(2^i u)^\beta} - \frac{3\gamma(2+\delta)}{(2^i u)^\beta} + \frac{3\gamma}{(2^i u)^\beta} + \sum_{j=0}^{\infty}\min\left\{\frac{2+2^j}{(2^i u)^\beta}, 1\right\}\cdot\frac{\gamma}{2^j} \tag{2.39}$$

Letting $j' = \lceil \beta(i+1+\log u)\rceil$, it is immediate that $\min\left\{\frac{2+2^j}{(2^i u)^\beta}, 1\right\} = 1$ for all $j > j'$. Since $(2+2^j)/2^j \le 3$ for all integers $j \ge 0$, (2.39) therefore yields

$$\begin{aligned} C_i &\le 1 + \frac{(2+\delta)}{(2^i u)^\beta} + \sum_{j=0}^{j'}\frac{3\gamma}{(2^i u)^\beta} + \sum_{j=j'+1}^{\infty}\frac{3\gamma}{2^{j-j'}(2^i u)^\beta} \\ &\le 1 + \frac{3\gamma(j'+1)+(2+\delta)}{(2^i u)^\beta} \le 1 + \frac{i+\log u}{(2^i u)^\beta}, \end{aligned}$$

since $3\gamma(j'+1)+(2+\delta) \leq 3\gamma(i+\log u)+12$, which is at most $i+\log u$ by our fourth requirement on $r_0$ and since $u \geq r_0$. Letting $c_i = 2^i u$, we thus have

$$\prod_{i=1}^{\infty} \max\{C_i, 1\} \leq \prod_{i=1}^{\infty} \left(1 + \frac{\log c_i}{c_i^{\beta}}\right), \tag{2.40}$$

and since we can make $c_1 = 2u \geq 2r_0$ as large as we like by our choice of $r_0$, it follows that we can ensure that $\prod_{i=1}^{\infty} \max\{C_i, 1\} \leq 2$ by choosing $r_0$ large enough (this can be easily seen by considering the logarithm of the second product in (2.40)). This completes the proof. □

In the above proof, we chose $\beta < \alpha$, then chose the sequence $\Delta_i$ so that $\Delta_{i+1}/\Delta_i \leq 1 + (\log(2^i u)/(2^i u)^{\beta})$. If we wish to prove a similar result when $X \in \mathcal{D}$ but $X$ is not necessarily in $WL^{2+\alpha}$, we may choose the sequence $\Delta_i$ so that $\Delta_{i+1}/\Delta_i = 1 + a$ for some fixed $a > 0$ as small as we wish, and define the events $B_i$ and $D_{i,j}$ in order to split the overshoot into pieces of size $2^j m_i$ instead of $2^j m_i^{1-\beta}$. Having done this, the bounds on the overshoot provided by Theorems 12 and 13 yield that essentially the same proof applies when we only impose that $X \in \mathcal{D}$. In this case, however, we can not bound $\Delta_j$ by a constant when $j$ grows, but have to settle for the bound $\Delta_j = O((1+a)^j)$. If $j$ is at most $\log r$, then by making $a$ small enough we can ensure that $\Delta_j = O(r^b)$ for $b$ as small as we wish. Following this chain of reasoning, we can prove:

**Lemma 18.** *Suppose $X \in \mathcal{D}$. Then for all $\epsilon > 0$ there is $r_0 > 0$ such that for all $r \geq r_0$, for all $0 \leq s \leq r - r_0$,*

$$\mathbf{P}\left\{S_{T_{r,s}} < 0\right\} \leq \frac{(r-s)}{r^{1-\epsilon}}.$$

*and for all $r - r_0 \leq s \leq r$,*

$$\mathbf{P}\left\{S_{T_{r,s}} < 0\right\} \leq \frac{1}{r^{1-\epsilon}}.$$

We omit a formal proof of this lemma as it consists only in mimicking the proof of Lemma 16 along the lines sketched above.

Combining Lemmas 14 and 16, we have proved

**Theorem 19.** *Suppose $X \in WL^{2+\alpha}$ for some $0 < \alpha \leq 1$. Then there are $r_0 > 0$, $c > 0$ such that for all $r \geq r_0$, for all $0 \leq s \leq r - r_0$,*

$$\frac{r-s}{16r} \leq \mathbf{P}\left\{S_{T_{r,s}} < 0\right\} \leq \frac{4(r-s)}{r}.$$

*and for all $r - r_0 < s < r$*

$$\frac{1}{cr} \leq \mathbf{P}\left\{S_{T_{r,s}} < 0\right\} \leq \frac{c}{r}.$$

Similarly, combining Lemmas 15 and 18, we have

**Theorem 20.** *Suppose $X \in \mathcal{D}$. Then for all $\epsilon > 0$ there is $r_0 > 0$ such that for all $r \geq r_0$, for all $0 \leq s \leq r - r_0$,*

$$\frac{r-s}{r^{1+\epsilon}} \leq \mathbf{P}\left\{S_{T_{r,s}} < 0\right\} \leq \frac{r-s}{r^{1-\epsilon}}.$$

*and for all $r - r_0 < s < r$*

$$\frac{1}{r^{1+\epsilon}} \leq \mathbf{P}\left\{S_{T_{r,s}} < 0\right\} \leq \frac{1}{r^{1-\epsilon}}.$$

We now prove that bounds such as those in Lemma 14 hold even if we additionally impose that $T_{r,s}$ is not too large and none of the step sizes are too big. This introduces minor technicalities, but the substance of the proof is the same as that of Lemma 14. (This result is perhaps less "independently interesting" than Theorems 19 and 20, but we will need it in proving our generalized ballot theorems.) We define $T_{r,s}$ as above and let $M_{r,s} = \max\{|X_i| \mid 1 \leq i \leq T_{r,s}\}$.

**Lemma 21.** *Suppose $\mathbf{E}X = 0$, $\mathbf{Var}\{X\} > 0$ and $X \in WL^{2+\alpha}$ for some $0 < \alpha \leq 1$. Then there are $c > 0$, $C > 0$ such that for all $\delta > 0$, there is $r_0 > 0$ such that for all $r \geq r_0$, for* *all* $0 \leq s \leq r - r_0$,

$$\mathbf{P}\left\{S_{T_{r,s}} < 0, T_{r,s} \leq Cr^2, M_{r,s} < \delta r\right\} \geq \frac{(r-s)}{cr},$$

*and for all* $r - r_0 < s < r$,

$$\mathbf{P}\left\{S_{T_{r,s}} < 0, T_{r,s} \leq Cr^2, M_{r,s} < \delta r\right\} \geq \frac{1}{cr}.$$

*Proof.* Just as in Lemma 14, it is straightforward to show that the first claim of the lemma implies the second; we therefore focus our attention on proving the first claim.

In a nutshell, our argument proceeds as follows. We apply the doubling argument just as in the proof of Lemma 14 to bound $\mathbf{P}\left\{S_{T_{r,s}} < 0\right\}$. We then use the independence of disjoint sections of the random walk to *individually* bound the probability that any given doubling "takes too long". Finally, since for a given doubling it is fairly likely that the walk goes negative and that the doubling does not take "too long", we are able to use the fact that $X \in WL^{2+\alpha}$ to bound the probability that in a given doubling, any of the $X_i$ are large, and conclude with a union bound over all the doublings to prove the overall bound. We now turn to the details.

We let $u = r - s$ and $d = r + s$; recall that $u > r_0$ by assumption, that $d > u$ as $s > 0$, and that we are interested in the first time the walk exceeds $u$ or dips below $-d$. For $i \geq 0$ let $v_i = (2^i - 1)u$, let $T_0 = 0$ and for $i \geq 1$ let $T_i$ be the first time $t > T_{i-1}$ that $|S_t - S_{T_i}| \geq v_i$. Finally, let $j^* = \lceil \log(d/u) \rceil + 1$. We proceed to define events $E_i, L_i, M_i$ (for $0 \leq i \leq j^*$), which we will use to control the behavior of the $i$'th doubling. To be more precise, $E_i$ will control the *direction* of the $i$'th doubling, $L_i$ its *duration* (the time it takes to double), and $M_i$ the *maximum step size* during the $i$'th doubling.

We first let $E_0 = L_0 = M_0$ be the event that $S_0 = 0$ (so $\mathbf{P}\left\{E_0 \cap L_0 \cap M_0\right\} = 1$). Next, just as in the proof of Lemma 14, we let $E_i$ be the event that $S_{T_i} < S_{T_{i-1}}$ – we recall that if $\cap_{i=0}^{j^*} E_i$ holds then $T_{r,s} \leq T_{j^*}$ and $S_{T_{r,s}} < 0$.

Before defining $L_i$, we first observe that by Lemma 8 and the strong Markov property, there is $A > 0$ such that for all $i \geq 1$ and $k \geq 1$,

$$\mathbf{P}\left\{T_i - T_{i-1} > Ak(2^i u)^2 | X_1, \ldots, X_{T_{i-1}}\right\} \leq \frac{1}{2^k}.$$

Based on this observation, we let $L_i$ be the event that $T_i - T_{i-1} \leq A(4 + j^* - i)(2^i u)^2$ – so $\mathbf{P}\left\{\bar{L}_i\right\} \leq 2^{-(4+j^*-i)}$ – and let $C = 512A$. With this choice of $C$, if $\bigcap_{i=0}^{j^*} E_i \cap L_i$ holds then

$$\begin{aligned} T_{r,s} &\leq T_{j^*} \leq \sum_{i=1}^{j^*} T_i - T_{i-1} \leq Au^2 \sum_{i=1}^{j^*} (4 + j^* - i) 2^{2i} = Au^2 (2^{2j^*}) \sum_{k=0}^{j^*-1} 2^{-2k}(k+4) \\ &< 2^{2j^*}(8Au^2) < (16d^2/u^2)(8Au^2) = 128Ad^2 = 128A(r+s)^2 < 512Ar^2 = Cr^2. \end{aligned}$$

Finally, we let $M_i$ be the event that $\max\{|X_k| \mid T_{i-1} < k \leq T_i\} < \delta r$. The event $M_i$ controls the maximum step size during the $i$'th doubling – if $\bigcap_{i=0}^{j^*} M_i \cap E_i$ occurs then $M_{r,s} \leq \delta r$. It follows from these definitions and comments that

$$\begin{aligned} \mathbf{P}\left\{S_{T_{r,s}} < 0, T_{r,s} \leq Cr^2, M_{r,s} < \delta r\right\} &\geq \mathbf{P}\left\{\bigcap_{i=0}^{j^*} E_i \cap L_i \cap M_i\right\} \\ &= \prod_{i=1}^{j^*} \mathbf{P}\left\{E_i \cap L_i \cap M_i \,|\, \bigcap_{j=0}^{i-1} E_j \cap L_j \cap M_j\right\} \\ &= \prod_{i=1}^{j^*} \mathbf{P}\left\{E_i \cap L_i \cap M_i\right\}, \end{aligned} \tag{2.41}$$

the last equality holding by the independence of disjoint sections of the random walk. We prove the theorem by bounding the component probabilities of the last product in (2.41); to do so, it is useful to first replace the event $M_i$ by an event that depends on a deterministic number of $X_j$. To this end, we first observe that $M_i$ is contained in the event $M_i^*$ that

$$\max\{|X_k| \mid T_{i-1} < k \leq \min\{T_{i-1} + \lfloor A(3+j^*-i)(2^i u)^2 \rfloor, T_i\}\} < \delta r,$$

and if $L_i$ occurs then $T_i \leq T_{i-1} + \lfloor A(3+j^*-i)(2^i u)^2 \rfloor$, so $M_i \cap L_i = M_i^* \cap L_i$. Furthermore, the event $M_i^*$ contains the event $M_i^{**}$ that

$$\max\{|X_k| \mid T_{i-1} < k \leq T_{i-1} + \lfloor A(3+j^*-i)(2^i u)^2 \rfloor\} < \delta r.$$

Combining these facts, we thus have

$$\begin{aligned} \mathbf{P}\{E_i \cap L_i \cap M_i\} &= \mathbf{P}\{E_i \cap L_i \cap M_i^*\} \\ &\geq \mathbf{P}\{E_i \cap L_i \cap M_i^{**}\} \geq 1 - \mathbf{P}\{\bar{E}_i\} - \mathbf{P}\{\bar{L}_i\} - \mathbf{P}\{\bar{M}_i^{**}\}. \quad (2.42) \end{aligned}$$

We now turn to the bounds on these probabilities. We bound $\mathbf{P}\{E_i\}$ just as in the course of Lemma 14, where we established (2.20); an identical derivation shows that there is a constant $c_0$ such that

$$\mathbf{P}\{E_i\} \geq \frac{1}{2} - \frac{c_0}{(2^i u)^\alpha}. \quad (2.43)$$

Next, as we noted when defining $L_i$, by Lemma 8 and the strong Markov property we have

$$\mathbf{P}\{\bar{L}_i\} \leq \frac{1}{2^{4+j^*-i}}. \quad (2.44)$$

Finally, we bound $\mathbf{P}\{\bar{M}_i^{**}\}$ by a union bound:

$$\begin{aligned} \mathbf{P}\{\bar{M}_i^{**}\} &\leq \sum_{j=T_i}^{\lfloor A(4+j^*-i)(2^i u)^2 \rfloor} \mathbf{P}\{|X_j| > \delta r\} \\ &= A(3+j^*-i)(2^i u)^2 \mathbf{P}\{|X| > \delta r\}. \quad (2.45) \end{aligned}$$

Since $X \in WL^{2+\alpha}$, $\mathbf{P}\{|X| \geq t\} = O(1/t^{2+\alpha})$, so for any $\delta_1 > 0$, by choosing $r_0$ large enough we may in particular ensure that for $r \geq r_0$, $\mathbf{P}\{|X| \geq \delta r\} \leq \delta_1/Ar^2$ (we will choose $\delta_1$ shortly). Since $2^{j^*}u \leq 4r$, it thus follows from (2.45) that for $r \geq r_0$,

$$\begin{aligned} \mathbf{P}\{\bar{M}_i^{**}\} &\leq \frac{\delta_1(4+j^*-i)(2^i u)^2}{r^2} \\ &\leq \frac{16\delta_1(4+j^*-i)}{2^{2(j^*-i)}}. \end{aligned} \tag{2.46}$$

Plugging (2.43), (2.44), and (2.46) into (2.42) yields

$$\mathbf{P}\{E_i \cap L_i \cap M_i\} \geq \frac{1}{2} - \frac{c_0}{(2^i u)^\alpha} - \frac{1}{2^{4+j^*-i}} - \frac{16\delta_1(4+j^*-i)}{2^{2(j^*-i)}}. \tag{2.47}$$

We now choose $\delta_1$ small enough that $16\delta_1(4+j^*-i)/2^{2(j^*-i)} \leq 1/2^{4+j^*-i}$ for all $1 \leq i \leq j^*$, so (2.47) gives

$$\mathbf{P}\{E_i \cap L_i \cap M_i\} \geq \frac{1}{2} - \frac{c_0}{(2^i u)^\alpha} - \frac{1}{2^{3+j^*-i}}, \tag{2.48}$$

and combining (2.48) and (2.41) yields

$$\begin{aligned} \mathbf{P}\left\{S_{T_{r,s}} < 0, T_{r,s} \leq Cr^2, M_{r,s} < \delta r\right\} &\geq \prod_{i=1}^{j^*}\left(\frac{1}{2} - \frac{c_0}{(2^i u)^\alpha} - \frac{1}{2^{3+j^*-i}}\right). \\ &= \prod_{i=1}^{j^*}\frac{1}{2}\left(1 - \frac{2c_0}{(2^i u)^\alpha} - \frac{1}{2^{2+j^*-i}}\right) \\ &\geq \frac{1}{2^{j^*}}\left(1 - \sum_{i=1}^{j^*}\frac{2c_0}{(2^i u)^\alpha} - \sum_{i=1}^{j^*}\frac{1}{2^{2+j^*-i}}\right). \end{aligned} \tag{2.49}$$

The second sum in (2.49) is strictly less than 1/2, and since $u \geq r_0$, we can make the first sum in (2.49) as small as we like by choosing $r_0$ large enough. It in particular follows that as long as $r_0$ is large enough, (2.49) is at least $1/2^{j^*+2}$, say. In other words,

$$\mathbf{P}\left\{S_{T_{r,s}} < 0, T_{r,s} \leq Cr^2, M_{r,s} < \delta r\right\} \geq \frac{1}{2^{j^*+2}} \geq \frac{u}{16d} \geq \frac{r-s}{32r},$$

and we complete the proof by taking $c = 32$. □

We note that by using the bound $\mathbf{P}\{|X| \geq t\} = O(1/t^{2+\alpha})$ more carefully, we could have replaced the event $M_{r,s} \leq \delta r^2$ by the event $M_{r,s} \leq Cr^{2-\alpha}$, perhaps with a changed value of $C$, and derived the same probability bound; we did not bother to do so as the weaker statement is sufficient in our applications of Lemma 21. The following corollary is immediate:

**Corollary 22.** *Under the conditions of Lemma 21, there are $c > 0$, $C > 0$ such that for all $\delta > 0$, there is $r_0 > 0$ such that for all $r \geq r_0$, for all $0 \leq s \leq r - r_0$,*

$$\mathbf{P}\left\{S_{T_{r,-s}} > 0, T_{r,-s} \leq Cr^2, M_{r,-s} < \delta r\right\} \geq \frac{(r-s)}{cr},$$

*and for all $r - r_0 < s < r$,*

$$\mathbf{P}\left\{S_{T_{r,-s}} > 0, T_{r,-s} \leq Cr^2, M_{r,-s} < \delta r\right\} \geq \frac{1}{cr}.$$

Corollary 22 follows by applying Lemma 21 to the random walk $S'$ given by $S'_i = -S_i$. We only bother stating it because in later arguments, it will be convenient to directly apply the corollary rather than first taking the negative of the walk under consideration, then applying Lemma 21. The following analogue of Lemma 21 holds when $X \in \mathcal{D}$, and has a practically identical proof, using the fact that if $S_t/a_t$ tends to the normal distribution then for any $\epsilon > 0$, $\mathbf{P}\{X_1 \geq \epsilon a_t\} = o(1/t)$ (see, e.g., Petrov, 1975, p. 98 for a proof of this fact) in place of the bound $\mathbf{P}\{X \geq t\} = O(1/t^{2+\alpha})$ we used above. We omit the proof.

**Lemma 23.** *Suppose the sequence $\{a_n\}_{n\geq 0}$ is such that $S_n/a_n$ tends to the normal distribution. Then there is $C > 0$ such that for all $\epsilon > 0$, there exists $r_0 > 0$ such that for all $r \geq r_0$ and $0 \leq s \leq r - r_0$,*

$$\mathbf{P}\left\{S_{T_{r,s}} < 0, T_{r,s} \leq Cr^2, M_{r,s} \leq \epsilon a_{\lfloor Cr^2 \rfloor}\right\} \geq \frac{(r-s)}{r^{1+\epsilon}}.$$

*Furthermore, for all $r \geq 0$ if $r - r_0 < s < r$ then*

$$\mathbf{P}\left\{S_{T_{r,s}} < 0, T_{r,s} \leq Cr^2, M_{r,s} \leq C\right\} \geq \frac{1}{r^{1+\epsilon}}.$$

Before continuing with the development of the generalized ballot theorem, we briefly digress to show how Theorem 19 can be used to substantially strengthen the expectation bounds of Lemma 10.

**Lemma 24.** *Suppose $\mathbf{E}\{X\} = 0$ and $X \in WL^{2+\alpha}$ for some $\alpha > 0$. Then there is $C > 0$ such that for $r \geq 0$ and for all $s$ for which $0 \leq |s| \leq r$,*

$$\mathbf{E}T_{r,s} \leq C((r+s)(r-s)+r). \tag{2.50}$$

This lemma is a direct analog of the work of Bertrand and of Rouché on time to ruin, discussed in Section 1.1. In that section we saw that for a symmetric simple random walk, for any integers $r > 0$ and $s$ with $0 \leq |s| \leq r$, the expected time until $S$ first visits $(r-s)$ or $-(r+s)$ is precisely $(r+s)(r-s)$ Grimmett and Stirzaker (1992, Example 3.9.6). Provided that $r - |s| \geq \epsilon > 0$, the bound of Lemma 24 has the same order as that suggested by the symmetric simple random walk.

*Proof of Lemma 24.* We let $v > 0$ and $\epsilon > 0$ be such that $\mathbf{P}\{|X| > v\} > \epsilon$, and choose $C > (8/v)^2(1/\epsilon) + 2$ so that (2.50) holds by Lemma 10 for small $r$, in particular for $r$ such that $(4r/v)^2(1/\epsilon) < 4$, and additionally any time $|s| \leq r/2$. We also choose $C$ large enough that $\epsilon C(v/4)^2 \geq \max\{8, 1/\delta\}$, where $\delta$ is the constant from Theorem 19.

Supposing that (2.50) holds for a given $r$ (for which $(4r/v)^2 > 2$), we proceed to bound $\mathbf{E}T_{2r,2s}$. We suppose $s > 0$, the result following by symmetry if $s < 0$ – as noted, we may assume by our choice of $C$ that $2s > r$. We consider the first time $t$ at which $S_t + 2s < 0$ or

$S_t + 2s \geq 2r$, denoting this time $T$.

If $S_T + 2s > 2r$ then $T = T_{2r,2s}$. If $S_T < 2s$ then we may restart the random walk at time $T$ as in Lemma 10 to see that

$$\mathbf{E}\left\{T_{2r,2s} - T | S_T < s\right\} \leq \sup_{0 \leq |s'| \leq r} \mathbf{E}T_{2r,2s'}.$$

Thus, just as in Lemma 10 we have

$$\begin{aligned} \mathbf{E}T_{2r,2s} &= \mathbf{E}T + \mathbf{P}\left\{S_T < s\right\} \cdot \mathbf{E}\left\{T_{2r,2s} - T | S_T < s\right\} \\ &\leq \mathbf{E}T + \mathbf{P}\left\{S_T < s\right\} \cdot \sup_{0 \leq |s'| \leq r} \mathbf{E}T_{2r,2s'} \end{aligned}$$

By its definition, $T$ is distributed like $T_{r,2s-r}$ and the event $\{S_T < s\}$ corresponds to the event $\{S_{T_{r,2s-r}} < 0\}$ so we have

$$\mathbf{E}T_{2r,2s} \leq \mathbf{E}T_{r,2s-r} + \mathbf{P}\left\{S_{T_{r,2s-r}} < 0\right\} \cdot \sup_{0 \leq |s'| \leq r} \mathbf{E}\left\{T_{2r,2s'}\right\}. \tag{2.51}$$

By Lemma 10 and our assumption on the size of $r$, we have

$$\sup_{0 \leq |s'| \leq r} \mathbf{E}\left\{T_{2r,2s'}\right\} \leq \max\left\{4\left(\frac{r}{v}\right)^2, 2\right\} \cdot \frac{1}{\epsilon} = \left(\frac{4r}{v^2}\right) \cdot \frac{1}{\epsilon},$$

so by (2.51)

$$\mathbf{E}T_{2r,2s} \leq \mathbf{E}T_{r,2s-r} + \mathbf{P}\left\{S_{T_{r,2s-r}} < 0\right\} \cdot \left(\frac{4r}{v}\right)^2 \cdot \frac{1}{\epsilon}. \tag{2.52}$$

By Theorem 19 and our assumption that $\max\{8, 1/\delta\} < \epsilon C(v/4)^2$,

$$\mathbf{P}\left\{S_{T_{r,2s-r}} < 0\right\} = \frac{8(r - (2s - r)) + (1/\delta)}{r} \leq \left(8\left(1 - \frac{s}{r}\right) + \frac{\epsilon C v^2}{32r}\right),$$

which together with (2.52) yields

$$\begin{aligned} \mathbf{E}T_{2r,2s} &\le \mathbf{E}T_{r,2s-r} + \left(8\left(1-\frac{s}{r}\right) + \frac{\epsilon C v^2}{32r}\right)\frac{(4r)^2}{\epsilon v^2} \\ &< \mathbf{E}T_{r,2s-r} + \frac{8\cdot 4^2}{\epsilon v^2} r(r-s) + Cr. \\ &\le \mathbf{E}T_{r,2s-r} + Cr(r-s) + Cr \end{aligned} \tag{2.53}$$

Finally, by induction we have

$$\mathbf{E}T_{r,2s-r} \le C\left((r-(2s-r))(r+(2s-r))+r\right) = 4Cs(r-s)+Cr,$$

so by (2.53),

$$\begin{aligned} \mathbf{E}T_{2r,2s} &\le 4Cs(r-s) + Cr + Cr(r-s) + Cr \\ &\le 4C(r+s)(r-s) + 2Cr \\ &= C\left((2r+2s)(2r-2s)+2r\right). \end{aligned}$$

This completes the proof. □

Intuitively, Theorem 19 is at the heart of our argument; in the language of Section 2.1, it provides precisely the bounds on $\mathbf{P}\{Pos_r\}$ that we sought. Indeed, there is a very simple intuitive argument that something like Theorem 19 should yield a ballot theorem as a corollary. Suppose $S_n$ is a random walk, and we have conditioned on the event $S_n = r$. Then in the *conditioned* random walk $X_1^c, X_2^c, \ldots, X_n^c$, each step has mean $r/n$, so at time $t$, the expected value is $rt/n$. If $rt/n = O(\sqrt{t})$, i.e., $\sqrt{t} = O(n/r)$, then up until time $t$, the "drift" of the random walk is still within its standard deviation; in some sense, the walk still "essentially" has mean zero up to this time. Lemma 10 tells us that by time $t \sim n^2/r^2$, we should have left the interval $[-\sqrt{t}, \sqrt{t}]$; Theorem 19 then suggests the probability we do so without ever leaving $[0, \sqrt{t}]$ is $\Theta(1/\sqrt{t}) = \Theta(r/n)$. (We note that the behavior of the initial steps of a conditioned random walk has been investigated by Zabell (1980), who considered for what functions $h$ and sequences $c_n$, it is the case that $\mathbf{E}\{h(X_1) \mid S_n = c_n\} \to \mathbf{E}h(X_1)$ as $n \to \infty$.)

As Theorem 19 does not apply to conditional sums, however, we can not directly formalize this intuitive argument, and end up having to apply Theorem 19 to the random walk "at both ends", as suggested by the sketch of Section 2.1.

## 2.4 The generalized ballot theorems

We now have all the tools we need to prove our generalized ballot theorems; before stating them we need a final definition. We say a variable $X$ has *period* $d > 0$ if $dX$ is an integer random variable and $d$ is the smallest positive real number for which this holds; in this case $X$ is called a *lattice* random variable, otherwise $X$ is *non-lattice*. We prove the following two theorems:

**Theorem 25.** *Suppose $X$ satisfies $\mathbf{E}X = 0$, $\mathbf{Var}\{X\} > 0$ and $X \in WL^{2+\alpha}$ for some $\alpha > 0$. Then there exists $A > 0$ such that given independent random variables $X_1, \ldots, X_n$ distributed as $X$ with associated partial sums $S_i = \sum_{j=1}^{i} X_i$, for all $0 \le k = O(\sqrt{n})$,*

$$\mathbf{P}\{k \le S_n \le k + A, S_i > 0 \ \forall\ 0 < i < n\} = \Theta\left(\frac{k+1}{n^{3/2}}\right).$$

*Furthermore, if $X$ is a lattice random variable with period $d$, then we may take $A = 1/d$, and if $X$ is non-lattice then we may take $A$ to be any positive real number.*

**Theorem 26.** *Suppose $X$ satisfies $\mathbf{E}X = 0$, $\mathbf{Var}\{X\} > 0$, and $X \in \mathcal{D}$. Then there exists a constant $A$ such that such that given independent random variables $X_1, \ldots, X_n$ distributed* *as $X$ with associated partial sums $S_i = \sum_{j=1}^{i} X_i$ and a sequence $\{a_n\}_{n=0}^{\infty}$ for which $S_n/a_n$ converges to a $\mathcal{N}(0,1)$ random variable, for all $0 \le k = O(a_n)$,*

$$\mathbf{P}\{k \le S_n \le k + A, S_i > 0 \;\forall\; 0 < i < n\} = \frac{k+1}{a_n \cdot n^{1-o(1)}}.$$

*Furthermore, if $X$ is a lattice random variable with period $d$, then we may take $A = 1/d$, and if $X$ is non-lattice then we may take $A$ to be any positive real number.*

From these two theorems, we may derive "true" (conditional) ballot theorems as corollaries, at least in the case that $S_n/a_n$ tends to a normal distribution and $k = O(a_n)$. The following result was proved by Stone (1965b), and is the tip of an iceberg of related results. Let $\Phi$ be the density function a $\mathcal{N}(0,1)$ random variable.

**Theorem 27.** *Suppose $S_n$ is a sum of independent, identically distributed random variables distributed as $X$ with $\mathbf{E}X = 0$, and there is a sequence of constants $a_n$ such that $S_n/a_n$ converges to a $\mathcal{N}(0,1)$ random variable. If $X$ is non-lattice let $B$ be any bounded set; then for any $h \in B$ and $x \in R$*

$$\mathbf{P}\{|S_n - x| \le h/2\} = \frac{h\Phi(x/a_n)}{a_n} + o(a_n^{-1}).$$

*Furthermore, if $X$ is a lattice random variable with period $d$, then for any $x \in \{n/d \mid n \in \mathbb{Z}\}$,*

$$\mathbf{P}\{S_n = x\} = \frac{\Phi(x/a_n)}{a_n} + o(a_n^{-1}).$$

*In both cases, $a_n o(a_n^{-1}) \to 0$ as $n \to \infty$ uniformly over all $x \in \mathbb{R}$ and $h \in B$.*

We discuss this result and its relatives in more detail in Appendix B, for now contenting ourselves with its statement and the observation that if $\mathbf{E}\{X^2\} < \infty$ then we may take $a_n = O(\sqrt{n})$, and Theorem 27 provides a pleasing counterpart to Theorem 9. Together with

Theorems 25 and 26, this immediately yields:

**Corollary 28.** *Under the conditions of Theorem 25,*

$$\mathbf{P}\{S_i > 0 \; \forall \; 0 < i < n | k \leq S_n \leq k + A\} = \Theta\left(\frac{k+1}{n}\right).$$

**Corollary 29.** *Under the conditions of Theorem 26,*

$$\mathbf{P}\{S_i \geq 0 \; \forall \; 1 \leq i \leq n | k \leq S_n \leq k + A\} = \frac{k+1}{n^{1-o(1)}}.$$

We now return to the proofs of Theorems 25 and 26. In fact, we will prove the following, more general results. The first two theorems are upper bounds for the cases $X_1 \in WL^{2+\alpha}$ and $X_1 \in \mathcal{D}$, respectively. We note that the upper bounds do not require $k = O(\sqrt{n})$ or $k = O(a_n)$. The second two theorems are the corresponding lower bounds, for which we require $k = O(\sqrt{n})$ and $k = O(a_n)$, respectively.

**Theorem 30.** *Suppose $X$ satisfies $\mathbf{E}X = 0$, $\mathbf{Var}\{X\} > 0$ and $X \in WL^{2+\alpha}$ for some $\alpha > 0$. Then for any fixed $A > 0$, given independent identically distributed random variables $X_1, \ldots, X_n$ distributed as either $X$ or $-X$, with associated partial sums $S_i = \sum_{j=1}^{i} X_i$, for all $m \geq 0$ and for all $k \geq -m$,*

$$\mathbf{P}\{k \leq S_n \leq k + A, S_i > -m \; \forall \; 0 < i < n\} = O\left(\frac{\min\{k+m+1, \sqrt{n}\} \cdot \min\{m+1, \sqrt{n}\}}{n^{3/2}}\right).$$

**Theorem 31.** *Suppose $X$ satisfies $\mathbf{E}X = 0$, $\mathbf{Var}\{X\} > 0$, and $X \in \mathcal{D}$. Then for any fixed $A > 0$, given independent identically distributed random variables $X_1, \ldots, X_n$ distributed as either $X$ or $-X$, with associated partial sums $S_i = \sum_{j=1}^{i} X_i$, and a sequence $\{a_n\}_{n=0}^{\infty}$ for which $S_n/a_n$ converges to a $\mathcal{N}(0,1)$ random variable, for all $0 < \epsilon < 1/2$, for $n$ large enough,* *for all $m \geq 0$ and for all $k \geq -m$,*

$$\mathbf{P}\left\{k \leq S_n \leq k+A, S_i > -m \ \forall\ 0 < i < n\right\} = O\left(\frac{\min\{k+m+1,\sqrt{n}\} \cdot \min\{m+1,\sqrt{n}\}}{a_n \cdot n^{1-\epsilon}}\right).$$

Clearly proving this theorem just for step sizes $X$ would be sufficient as $-X$ satisfies the same conditions as $X$. We have stated the conclusion of the theorem for both $X$ and $-X$ to satisfy the requirements of our inductive proof. We remark that if $S_n/a_n$ converges to a normal random variable than necessarily $\mathbf{E}X_1 = 0$. The lower bounds are:

**Theorem 32.** *Suppose $X$ satisfies $\mathbf{E}X = 0$, $\mathbf{Var}\{X\} > 0$ and $X \in WL^{2+\alpha}$ for some $\alpha > 0$. Then there exists $A > 0$ such that given independent random variables $X_1, \ldots, X_n$ distributed as $X$ with associated partial sums $S_i = \sum_{j=1}^{i} X_i$, for all $0 \leq m = O(\sqrt{n})$ and for all $k$ for which $-m \leq k = O(\sqrt{n})$,*

$$\mathbf{P}\left\{k \leq S_n \leq k+A, S_i > -m \forall\ 0 < i < n\right\} = \Omega\left(\frac{\min\{k+m+1,\sqrt{n}\} \cdot \min\{m+1,\sqrt{n}\}}{n^{3/2}}\right).$$

*Furthermore, if $X$ is a lattice random variable with period $d$, then we may take $A = 1/d$, and if $X$ is non-lattice then we may take $A$ to be any positive real number.*

**Theorem 33.** *Suppose $X$ satisfies $\mathbf{E}X = 0$, $\mathbf{Var}\{X\} > 0$, and $X \in \mathcal{D}$. Then there exists a constant $A$ such that such that given independent random variables $X_1, \ldots, X_n$ distributed as $X$ with associated partial sums $S_i = \sum_{j=1}^{i} X_i$ and a sequence $\{a_n\}_{n=0}^{\infty}$ for which $S_n/a_n$ converges to a $\mathcal{N}(0,1)$ random variable, for all $0 < \epsilon < 1/2$, for $n$ large enough, for all $0 \leq m = O(a_n)$ and for all $k$ for which $-m \leq k = O(a_n)$,*

$$\mathbf{P}\left\{k \leq S_n \leq k+A, S_i > -m \forall\ 0 < i < n\right\} = \Omega\left(\frac{\min\{k+m+1,\sqrt{n}\} \cdot \min\{m+1,\sqrt{n}\}}{a_n \cdot n^{1+\epsilon}}\right).$$

*Furthermore, if $X$ is a lattice random variable with period $d$, then we may take $A = 1/d$, and if $X$ is non-lattice then we may take $A$ to be any positive real number.*

We note that we may combine the upper and lower bounds of these stronger theorems using Theorem 27 to obtain conditional corollaries that are exact analogs of Corollaries 28 and 29. We refrain from stating these corollaries explicitly; they contain no surprises.

Both the lower bounds and the upper bounds are proved by splitting the random walk $S_1, \ldots, S_n$ into three parts: a beginning, a middle, and an end. In each part we consider what behavior must occur (when proving upper bounds), or what behavior suffices (when proving lower bounds) in order that $k \leq S_n \leq k + A$ and $S_i > -m$ for all $0 < i < n$. In the first and last parts it is necessary and sufficient that the random walk "go positive" in the roughly the sense of the events *Pos* and *Pos*$'$ of Section 2.1. In the middle, it is necessary that the random walk sum to the right value, and sufficient that additionally, the random walk does not "go too negative" during this time. We first prove the upper bounds.

*Proof of Theorem 30.* We first fix any $A$ satisfying the conditions of Theorem 30. We will prove (2.54) by induction on $n$. Essentially, we use the argument sketched in Section 2.1. Because we allow $k = \Omega(\sqrt{n})$ (so $k$ is not "small", in the terminology of that section), however, we can not define event such as $Pos_k$ and $Pos'$ in terms of a stopping time as we did there. Instead, we stop the walk *deterministically* at time $t_1 = \lfloor n/4 \rfloor$ and consider its behavior up to that point. If there is $i \leq t_1$ for which $S_i \geq \delta\sqrt{n}$ (for some carefully chosen $\delta$), and $S$ has stayed above $-m$ until this time, then we essentially argue as in Section 2.1. Otherwise, we inductively apply the ballot theorem.

In the course of the proof, we will apply induction to $S$ as well as to the *negative reversed* walk $S^r$ defined by $S^r_0 = 0$, and for $0 \leq i < n$, $S^r_{i+1} = S^r_i - X_{n-i}$. We note that if $S$ has step size $X$ then $S^r$ has step size $-X$, and vice-versa. For the purposes of our induction, it is useful to replace the order notation in the desired upper bound by an explicit constant: we will show that there is $C > 0$ such that for all $n$, we have

$$\mathbf{P}\left\{k \leq S_n \leq k+A, S_i > -m \forall\, 0 < i < n\right\} \leq \frac{C \min\{k+m+1, \sqrt{n}\} \cdot \min\{m+1, \sqrt{n}\}}{n^{3/2}}, \tag{2.54}$$

whether $S$ has step size $X$ or $-X$. We note that by Theorem 9, there is a constant $a$ (and we can and will assume $a > 1$) such that $\mathbf{P}\left\{k \leq S_n \leq k+A\right\} \leq aA/\sqrt{n}$, whether $S$ has step size $X$ or $-X$. Therefore, for $n \leq C/aA$, (2.54) follows immediately from Theorem 9.

We now make our choice of $C$ more explicit, stating bounds that we will use at various points in the proof. We have gathered these together so that it is easy to see that our assumptions on the size of $C$ are well-defined and are not contradictory. We first require that $1/8C^{2/5} > 1/C^{4/7}$, and, letting $a$ be the constant from Theorem 9, that $C \geq 8aA$, that $C^{1/5} \geq 128a(A+1)$, that $C^{1/5} \geq 256aA^3$, and that $C^{1/5} \geq 144aA$. (Letting $\delta = 1/C^{2/5}$ and $\epsilon = \delta/8 = 1/8C^{2/5}$, we will use the following inequalities, which are immediate from our bounds on $C$: $\epsilon > 1/C^{4/7}$; $C \geq 2a(A+1)/\epsilon^2$; $(\epsilon/2)\sqrt{C/aA} \geq A$; and $2\epsilon\sqrt{C/aA} < \delta\sqrt{C/4aA} - 3$.)

We additionally fix a constant $r_0$ – eventually, we will apply Theorem 19 to bound events of the form $\{S_{T_{r,s}} < 0\}$, for $r \geq r_0$, and we choose $r_0$ large enough that Theorem 19 indeed applies for such $r$, whether $S$ has step size $X$ or $-X$. We presume $C$ has been chosen large enough that $C^{1/10}/16\sqrt{aA} \geq r_0$. Most of our restrictions on the size of $C$ appear in this and the previous paragraph, but a few more (which are much more naturally stated "in context") will arise as we go along.

As noted, (2.54) holds for all $n \leq C/aA$ by Theorem 9. Fix some $n > C/aA$ and suppose that (2.54) holds for all $n_0 \leq n$, for all $m \geq 0$ and all $k \geq -m$, whether $S$ has step size $X$

or $-X$. We will prove that for all $k \geq -A$,

$$\mathbf{P}\left\{k \leq S_n \leq k+A, S_i > -m \forall\, 0 < i < n\right\} \leq \frac{C\min\{k+m+1,\sqrt{n}\}\cdot\min\{m+1,\sqrt{n}\}}{(A+1)n^{3/2}}, \tag{2.55}$$

whether $S$ has step size $X$ or $-X$. Suppose for a moment that (2.55) holds for all $k \geq -A$, and choose some $k < -A$. If $k \leq S_n \leq k+A$ and $S_i > -m$ for all $0 < i < n$ are to occur, then it must be the case that $-(k+A) \leq S_n^r \leq -k$, and $S_i^r \geq -(k+m+A)$ for all $0 < i < n$. Letting $k' = -(k+A)$ and $m' = (k+m+A)$, we have $k' > 0 > -A$ and $k'+m' = m$; (2.55) thus yields that

$$\begin{aligned}
\mathbf{P}\left\{k \leq S_n \leq k+A, S_i > -m \forall 0 < i < n\right\} &\leq \mathbf{P}\{k' \leq S_n^r \leq k'+A, S_n^r \geq -m' \,\forall\, 0 < i < n\} \\
&\leq \frac{C\min\{k'+m'+1,\sqrt{n}\}\cdot\min\{m'+1,\sqrt{n}\}}{(A+1)n^{3/2}} \\
&= \frac{C\min\{m+1,\sqrt{n}\}\min\{k+m+A+1,\sqrt{n}\}}{(A+1)n^{3/2}} \\
&\leq \frac{C\min\{m+1,\sqrt{n}\}(\min\{k+m+1,\sqrt{n}\}+A)}{(A+1)n^{3/2}} \\
&\leq \frac{C\min\{m+1,\sqrt{n}\}\min\{k+m+1,\sqrt{n}\}}{n^{3/2}},
\end{aligned}$$

the last inequality holding since $\min\{k+m+1,\sqrt{n}\} \geq 1$, which establishes (2.54) for such a choice of $k$ and of $m$. Therefore, to prove that (2.54) holds for this value of $n$ it suffices to show that (2.55) holds for all $k \geq -A$; this is the subject of the remainder of the proof.

We first prove (2.55) in the case that $S$ has step size $X$, and begin by fixing $k \geq -A$. We note that by our assumptions on $n$ and on $C$, $\epsilon\sqrt{n} - A \geq \epsilon\sqrt{n} - (\epsilon/2)\sqrt{C/aA} \geq \epsilon\sqrt{n}/2$, so if $m$ is greater than $\epsilon\sqrt{n}$ then since $k \geq -A$, we have $k+m > \epsilon\sqrt{n} - A \geq \epsilon\sqrt{n}/2$. Therefore, $\min\{k+m+1,\sqrt{n}\}\min\{m+1,\sqrt{n}\}/n^{3/2}$ is at least $\epsilon^2/2\sqrt{n}$, in which case (2.55) follows immediately from Theorem 9 and our assumption that $C \geq 2a(A+1)/\epsilon^2$. We thus need only consider the case that $m \leq \epsilon\sqrt{n}$, and hereafter presume that this is indeed the case.

Let $E$ be the event that $k \leq S_n \leq k + A$ and $S_i \geq -m$ for all $0 < i < n$ – so we seek a bound on $\mathbf{P}\{E\}$. Recall that $t_1 = \lfloor n/4 \rfloor$. We consider the value of the walk at time $S_{t_1}$. If $E$ is to occur then one of the following events must occur:

- either $S$ exceeds $\lfloor \delta\sqrt{t_1} \rfloor$ before the first time $S$ drops below $-m$, and additionally before time $t_1$ (we denote this event $E_1$);
- or $-m + Aj \leq S_{t_1} < -m + A(j+1)$ (for some $j = 0, 1, \ldots, \lfloor (m + \delta\sqrt{t_1})/A \rfloor$) and $S_i \geq -m$ for all $1 \leq i \leq t_1$ (we denote these events $E_{2,j}$ for $j$ as above).

Let $r = (\lfloor \delta\sqrt{t_1} \rfloor + m)/2$, let $s = (\lfloor \delta\sqrt{t_1} \rfloor - m)/2$, for $i \geq 0$ let $S_i^- = -S_i$, and let $T_{r,s}$ be the first time $t$ that $|S_i^- + s| \geq r$. Then in the terminology of Theorem 19, $E_1$ is contained in the event $\{S^-_{T_{r,s}} < 0\}$. Furthermore, $r \geq \delta\sqrt{t_1}/2 = \delta\sqrt{\lfloor n/4 \rfloor}/2 \geq \sqrt{n}/16C^{2/5}$ since $n \geq C/Aa \geq 8$ and $\delta = 1/C^{2/5}$. Since $n > C/aA$, it follows that $r \geq C^{1/10}/16\sqrt{aA} \geq r_0$ by our choice of $C$. By Theorem 19, therefore, there is a constant $c = c(X) > 1$ such that

$$\mathbf{P}\{E_1\} \leq \frac{4(r-s)+c}{r} = \frac{8m+2c}{\lfloor \delta\sqrt{t_1} \rfloor + m} \leq \frac{8(m+c)}{\lfloor \delta\sqrt{t_1} \rfloor} \leq \frac{24(m+c)}{\delta\sqrt{n}}. \tag{2.56}$$

Since $c$ is constant, we may certainly choose $C$ large enough that $c \leq (1-\epsilon)\sqrt{C/aA}$; since $n \geq C/aA$ and we have assumed that $m \leq \epsilon n$, it follows that $m + c \leq \sqrt{n}$, so $24(m+c) = 24\min\{m+c, \sqrt{n}\}$. From (2.56), we thus have

$$\mathbf{P}\{E_1\} \leq \frac{24\min\{m+c, \sqrt{n}\}}{\delta\sqrt{n}}. \tag{2.57}$$

We bound the probability of the events $E_{2,j}$ by induction; for such events applying (2.54) to the random walk $S_1, \ldots, S_{t_1}$ yields

$$\begin{aligned}\mathbf{P}\{E_{2,j}\} &\leq \frac{C(\min\{Aj+1, \sqrt{t_1}\}\min\{m+1, \sqrt{t_1}\})}{(t_1)^{3/2}} \\ &\leq \frac{16C(Aj+1)\min\{m+1, \sqrt{t_1}\}}{n^{3/2}}\end{aligned}$$

Thus, letting $E_2 = \bigcup_{j=0}^{\lfloor (m+\delta\sqrt{t_1})/A \rfloor} E_{2,j}$, we have

$$\begin{aligned}
\mathbf{P}\{E_2\} &\le \sum_{j=0}^{\lfloor (m+\delta\sqrt{t_1})/A \rfloor} \mathbf{P}\{E_{2,j}\} \\
&\le \frac{16C\min\{m+1,\sqrt{t_1}\}}{n^{3/2}} \sum_{j=0}^{\lfloor (m+\delta\sqrt{t_1})/A \rfloor} (Aj+1). \\
&\le \frac{16C\min\{m+1,\sqrt{t_1}\}}{n^{3/2}} \left(\frac{m+\delta\sqrt{t_1}}{A}+1\right)(m+\delta\sqrt{t_1}+1). \\
&\le \frac{16C\min\{m+1,\sqrt{t_1}\}}{n^{3/2}} \max\left\{\frac{1}{A},1\right\}\left(\delta\sqrt{t_1}+m+1\right)^2. \qquad (2.58)
\end{aligned}$$

Since $2\epsilon\sqrt{C/aA} \le \delta\sqrt{C/4aA}-3$ by assumption, and $n \ge C/aA$, we have $2\epsilon\sqrt{n} \le \delta\sqrt{n/4}-3$. Together with the facts that $t_1 = \lfloor n/4 \rfloor$ and that $\delta < 1$, this immediately yields the bound

$$\frac{\delta\sqrt{t_1}}{2} - 1 > \frac{\delta\sqrt{n/4}}{2} - \frac{3}{2} \ge \epsilon\sqrt{n} = m.$$

It follows that $(\lfloor \delta\sqrt{t_1} \rfloor + m + 1)^2 \le (3\delta t_1/2)^2 < \delta^2 n$, and we thus have from (2.58) that

$$\mathbf{P}\{E_2\} \le \frac{16C\delta^2\max\{1/A,1\}\min\{m+1,\sqrt{n}\}}{\sqrt{n}}.$$

Combining this with (2.57) yields that

$$\mathbf{P}\{E_1 \cup E_2\} \le \frac{24\min\{m+c,\sqrt{n}\}}{\delta\sqrt{n}} + \frac{16C\delta^2\max\{1/A,1\}\min\{m,\sqrt{n}\}}{\sqrt{n}}.$$

Since $\delta = 1/C^{2/5}$, we thus have

$$\mathbf{P}\{E_1 \cup E_2\} \le \frac{24C^{2/5}\min\{m+c,\sqrt{n}\}}{\sqrt{n}} + \frac{16C^{1/5}\max\{1/A,1\}\min\{m+1,\sqrt{n}\}}{\sqrt{n}}. \qquad (2.59)$$

We assume $C$ is chosen large enough that $24cC^{2/5} + 16C^{1/5}\max\{1/A,1\} < C^{3/7}/2a(A+1)$.

Since $c > 1$, this implies that also $24C^{2/5} + 16C^{1/5}\max\{1/A, 1\} < C^{3/7}/2a(A+1)$, so

$$24C^{2/5}(m+c) + 16C^{1/5}\max\{1/A, 1\}(m+1) < C^{3/7}(m+1)/2a(A+1),$$

which combined with (2.59) yields

$$\mathbf{P}\{E_1 \cup E_2\} \leq \frac{C^{3/7}\min\{m+1, \sqrt{n}\}}{2a(A+1)\sqrt{n}}. \tag{2.60}$$

If $k + m + 1 \geq \epsilon\sqrt{n}$, we apply Theorem 9 and the strong Markov property to the random walk $S_{t_1}, \ldots, S_n$ to conclude that

$$\mathbf{P}\{k \leq S_n \leq k + A | E_1 \cup E_2\} \leq \frac{a}{\sqrt{3n/4}} < \frac{2a}{\sqrt{n}}.$$

Therefore, in this case,

$$\begin{aligned} \mathbf{P}\{E\} &\leq \mathbf{P}\{(E_1 \cup E_2) \cap k \leq S_n \leq k + A\} \\ &= \mathbf{P}\{E_1 \cup E_2\}\mathbf{P}\{k \leq S_n \leq k + A | E_1 \cup E_2\} \\ &\leq \frac{C^{3/7}\min\{m+1, \sqrt{n}\}}{2a(A+1)\sqrt{n}} \cdot \frac{2a}{\sqrt{n}} \\ &\leq \frac{C^{3/7}\min\{m+1, \sqrt{n}\}}{(A+1)n}. \end{aligned} \tag{2.61}$$

Since $\min\{k+m+1, \sqrt{n}\} \geq \epsilon\sqrt{n}$, and $\epsilon > 1/C^{4/7}$ by assumption, (2.61) implies that

$$\begin{aligned} \mathbf{P}\{E\} &\leq \frac{C^{3/7}\min\{m+1, \sqrt{n}\}\min\{k+m+1, \sqrt{n}\}}{\epsilon(A+1)n^{3/2}} \\ &< \frac{C\min\{m+1, \sqrt{n}\}\min\{k+m+1, \sqrt{n}\}}{(A+1)n^{3/2}}. \end{aligned}$$

This establishes (2.55) in the case that $S$ has step size $X$ and $k + m + 1 \geq \epsilon\sqrt{n}$. If $S$ has step size $X$ but $k + m + 1 < \epsilon\sqrt{n}$, then we must additionally use our inductive bounds on $S^r$. If we are to have $k \leq S_n \leq k + 1$ and $S_i > -m\ \forall 0 < i < n$, then in particular, it must be the case that $S_i^r \geq -\max(k+m+1, \sqrt{n})$ for all $1 \leq i \leq t_1$. We define events $E_1^r$ and $E_{2,j}^r$ (for $j = 0, 1, \ldots, \lfloor m+k+\delta\sqrt{t_1} \rfloor$), $E_2^r$ just as we defined $E_1, E_{2,j}$, and $E_2$, but this time with respect to $S_r$. If we are to have $S_i^r \geq -\max(k+m+1, \sqrt{n})$ for all $1 \leq i \leq t_1$ then either $E_1^r$ or $E_2^r$ must occur. An identical argument to that leading to 2.60 (but applying (2.54) inductively to $S_1^r, \ldots, S_{t_1}^r$ instead of to $S_1, \ldots, S_{t_1}$) then yields the bound

$$\mathbf{P}\{E_1^r \cup E_2^r\} \leq \frac{C^{3/7} \min\{k+m+1, \sqrt{n}\}}{2a(A+1)\sqrt{n}}. \tag{2.62}$$

Finally, since $S_{t_1}^r = -(X_{n-t_1+1} + \ldots + X_n)$, in order that $k-1 \leq S_n \leq k$ hold we must also have that

$$k - S_{t_1} + S_{t_1}^r \leq S_{n-t_1} - S_{t_1} \leq k+1 - S_{t_1} + S_{t_1}^r;$$

we denote this event $E_3$. By the strong Markov property, $S_{n-t_1} - S_{t_1}$ is independent of $E_1 \cup E_2$ and of $E_1^r \cup E_2^r$, so by Theorem 9, $\mathbf{P}\{E_3 | E_1 \cup E_2, E_1^r \cup E_2^r\} \leq a/\sqrt{n/2} < 2a/\sqrt{n}$, where $a > 1$ is the same constant as above. Since $E_1 \cup E_2$ and $E_1^r \cup E_2^r$ are likewise independent, and, as we have seen, all of $E_1 \cup E_2$, $E_1^r \cup E_2^r$, and $E_3$ must occur in order for $E$ to occur, we therefore have

$$\begin{aligned}
\mathbf{P}\{E\} &\leq \mathbf{P}\{E_3\}\mathbf{P}\{E_1 \cup E_2\}\mathbf{P}\{E_1^r \cup E_2^r\} \\
&\leq \frac{2a}{\sqrt{n}}\left(\frac{C^{3/7}\min\{m+1, \sqrt{n}\}}{2a(A+1)\sqrt{n}}\right)\left(\frac{C^{3/7}\min\{k+m+1, \sqrt{n}\}}{2a(A+1)\sqrt{n}}\right) \\
&< \frac{C\min(k+1, \sqrt{n})\min(k+m+1, \sqrt{n})}{(A+1)n^{3/2}},
\end{aligned}$$

as $C > C^{6/7}$ and $(A+1)^2 > A+1$. We have therefore shown that (2.55) holds when $S$ has step size $X$; an identical argument shows that (2.55) holds when $S$ has step size $-X$. This completes the proof. □

When $X \in \mathcal{D}$, an identical argument using Theorem 20 in place of Theorem 19 proves Theorem 31; the proof is omitted. We now turn our attention to the lower bounds. Theorem

32 is a fairly straightforward consequence of Corollary 22 and of the following lemma.

**Lemma 34.** *Suppose $X$ satisfies $\mathbf{E}X = 0$, $\mathbf{Var}\{X\} > 0$ and $X \in WL^{2+\alpha}$ for some $\alpha > 0$. Then there exists $A > 0$ such that given independent random variables $X_1, \ldots, X_n$ distributed as $X$ with associated partial sums $S_i = \sum_{j=1}^{i} X_i$, for any $K > 0$ and $\epsilon > 0$, there is $c^* > 0$ such that for $n$ large enough, for all $0 < a \leq K$, $0 < a' \leq K$, and all $r$ for which $-a'\sqrt{n} < r < a\sqrt{n}$,*

$$\mathbf{P}\left\{r \leq S_n \leq r + A, S_i \geq -(a' + \epsilon)\sqrt{n}\ \forall\ 1 \leq i \leq n\right\} \ \geq\ \frac{c^*}{\sqrt{n}}. \tag{2.63}$$

*Furthermore, if $X$ is a lattice random variable with period $d$, then we may take $A = 1/d$, and if $X$ is non-lattice then we may take $A$ to be any positive real number.*

We prove Theorem 32 presuming this lemma holds, then return to its proof.

*Proof of Theorem 32.* Fix any $A$ satisfying the conditions of Theorem 32, and fix a constant $a_1 > 1$ such that $|k| \leq a_1\sqrt{n}$, $|m| \leq a_1\sqrt{n}$ (such a constant exists by our assumption that both $k$ and $m$ are $O(\sqrt{n})$). We first demonstrate that for any fixed constant $c_0$, it suffices to prove the theorem for pairs $k, m$ for which $|k| \leq a_0\sqrt{n}$, $|m| \leq a_0\sqrt{n}$ (for some constant $a_0 > a_1$ which may depend on $c_0$) and for which additionally $m \geq c_0$ and $k + m \geq c_0$. To see this, suppose for a moment that the theorem holds for all such pairs $k, m$. As $\mathbf{E}X = 0$ and $\mathbf{Var}\{X\} > 0$, and by our choice of $A$, it is easy to see that there exist constants $c_1 \geq c_0$ and $t_1 > 0$ such that with probability $\Omega(1)$, $X_1, \ldots, X_{t_1}$ are all positive and $c_1 \leq S_{t_1} \leq c_1 + A$. Similarly, there are $c_2 \geq c_0 + A$ and $t_2 > 0$ such that with probability $\Omega(1)$, $X_{n-t_2+1}, \ldots, X_n$ are all negative and, letting $S^r_{t_2} = -\sum_{i=n-t_2+1}^{n} X_i$, we have $c_2 \leq S^r_{t_2} \leq c_2 + A$. Furthermore, for all $n > t_1 + t_2$, for $\{k \leq S_n \leq k + A\}$ and $\{S_i \geq -m$ for all $0 < i < n\}$ to occur it suffices that

(I) $X_1, \ldots, X_{t_1}$ are all positive and $c_1 \leq S_{t_1} \leq c_1 + A$,

(II) $X_{n-t_2+1}, \ldots, X_n$ are all negative and $c_2 \le S^r_{t_2} \le c_2 + A$, and

(III) considering the walk restarted at time $t_1$ given by $S'_i = S_{t+i} - S_i$ and letting $k' = k - S_{t_1} + S^r_{t_2}$, $m' = m + c_1$, $n' = n - t_1 - t_2$, we have $k' \le S'_{n'} \le k' + A$ and $S'_i \ge -m'$ for all $0 < i < n'$.

The events (I) and (II) both occur with probability $\Omega(1)$. Furthermore, $m' \ge c_1 \ge c_0$, and since $k \ge -m$, we also have

$$m' + k' = m + c_1 + k - S_{t_1} + S^r_{t_2} \ge m + k + c_1 - (c_1 + A) + c_2 \ge c_2 - A \ge c_0.$$

Finally, $n' \ge n/t_1 t_2$, so letting $a_0 = a_1 t_1 t_2$, we have $|m'| \le a_0\sqrt{n'}$ and $|k'| \le a_0\sqrt{n'}$. Therefore, the probability of (III) is $\Omega(\min\{k' + m' + 1, \sqrt{n'}\}\min\{m' + 1, \sqrt{n'}\}/(n')^{3/2})$ by assumption. Since $m' = m + O(1)$, $k' = k + O(1)$, and $n' = n + O(1)$, combining our bounds on (I), (II), and (III) then yields the bound we desire for $\mathbf{P}\{k \le S_n \le k + A, S_i \ge -m \ \forall\ 0 < i < n\}$.

We will shortly apply Corollary 22 with the choice $\delta = 1/8$; for the remainder of the proof we let $r_0 = r_0(\delta) = r_0(1/8)$ be as in the statement of Corollary 22. Based on the comments at the start of the proof, from this point on we can and will presume that $m \ge r_0$ and that $m + k \ge r_0$. Since $r_0$ is constant, by the above comments we can and will also presume that $|m| \le a_0\sqrt{n}$ and $|k| \le a_0\sqrt{n}$, where $a_0$ is a constant possibly depending on $r_0$ but not on $n$. Finally, fix $\gamma = 1/8C$, where $C$ is the constant from Corollary 22, and let $m^* = \min\{m+1, (\gamma/2)\sqrt{n}\}$, $k^* = \min\{k + m + 1, (\gamma/2)\sqrt{n}\}$. We presume $n$ is large enough that $(\gamma/2)\sqrt{n} \ge r_0$, so $m^* \ge r_0$ and $k^* \ge r_0$.

We consider the first time $t > 0$ that $S_t \ge \gamma\sqrt{n}$ or $S_t \le -m^*$, denoting this time $T$. We likewise consider the negative reversed walk $S^r$ with $S^r_0 = 0$, for $i \ge 0$ $S^r_{i+1} = S^r_i - X_{n-i}$, and let $T^*$ be the first time $t$ that $S^r_t \ge \gamma\sqrt{n}$ or $S_t \le -k^*$. In order that $k \le S_n \le k + A$, and $S_i \ge -m$ for all $0 < i < n$, it suffices that the following three events occur (these events control the behavior of the beginning, end, and middle of the random walk, respectively):

$E_1$: $\gamma\sqrt{n} \le S_T < (5/4)\gamma\sqrt{n}$ and $T < n/4$,

$E_2$: $\gamma\sqrt{n} \le S^r_{T^*} < (5/4)\gamma\sqrt{n}$ and $T^* < n/4$,

$E_3$: letting $\Delta = S^r_{T^*} - S_T$, we have $k + \Delta \le S_{n-T^*} - S_T \le k + \Delta + A$ and $S_i \ge -m - \gamma\sqrt{n}$ for all $T < i < n - T^*$.

In order for $E_1$ to occur, it suffices that

(1) $S_T \ge \gamma\sqrt{n}$,

(2) $T \le n/4$, and

(3) letting $M = \max_{1 \le i \le T} X_i$, we have $M < (\gamma/4)\sqrt{n}$.

We use Corollary 22 to bound the probability of $E_1$. In the notation of that corollary, $T$ is a stopping time $T_{r,-s}$ with $r = (\gamma\sqrt{n} + m^*)/2$ and $s = (\gamma\sqrt{n} - m^*)/2$, and $M$ is at most the corresponding maximum $M_{r,-s} = \max_{1 \le i \le T} |X_i|$. Corollary 22, applied with $\delta = 1/8$ and $r_0 = r_0(\delta)$ chosen as above, then states that there are $c > 0$, $C > 0$ such that as long as $r - s = m^* \ge r_0$, with probability at least $(r-s)/cr$, it is the case that $S_T \ge \gamma\sqrt{n}$, $T \le Cr^2$, and $M \le \delta r = r/8$. We have

$$\frac{r}{8} = \frac{1}{8}(\gamma\sqrt{n} + m^*) \le \frac{1}{8}\left(\frac{3\gamma\sqrt{n}}{2}\right) < \frac{\gamma\sqrt{n}}{4},$$

and note that since $\gamma = 1/8C$, we additionally have

$$Cr^2 = \frac{C(\gamma\sqrt{n} + m^*)^2}{2} \le \frac{C(\gamma\sqrt{n} + (\gamma/2)\sqrt{n})^2}{2} < 2C\gamma n \le \frac{n}{4}.$$

Finally, applying Corollary 22, it follows that

$$\mathbf{P}\{E_1\} \geq \mathbf{P}\{S_T \geq \gamma n, T \leq Cr^2, M \leq \delta r\} \geq \frac{r-s}{cr} = \frac{2m^*}{c(\gamma\sqrt{n}+m^*)} \geq \frac{m^*}{c\gamma\sqrt{n}}.$$

Similarly, applying Corollary 22 to the random walk $S^r$ to bound $\mathbf{P}\{E_2\}$ yields that

$$\mathbf{P}\{E_2\} \geq \frac{k^*}{c\gamma\sqrt{n}}.$$

Lastly, we wish to apply Lemma 34 to bound $\mathbf{P}\{E_3|E_1,E_2\}$. In order to apply Lemma 34, we show that if $E_1$ and $E_2$ occur then for $E_3$ to occur, it suffices that an event of the form $\{r' \leq S'_{n'} \leq r'+A, S'_i \geq -(a'+\epsilon)\sqrt{n'}\ \forall\ 1 \leq i \leq n'\}$ occur, for a suitable walk $S'$ and a suitable choice of $n'$, of $K>0$ and $\epsilon>0$, of $0<a<K$, $0<a'<K$, and of $r'$ for which $-a'\sqrt{n'} < r' < a\sqrt{n'}$ (we emphasize that though $a, a'$ and $r'$ may depend on $n'$, $K$ and $\epsilon$ will not depend on $n'$).

We let $S'$ be the random walk $S$ restarted at time $T$, i.e., $S'_i = S_{T+i} - S_T$, and set $n' = n-T-T^*$. We let $r' = (k+\Delta)$ – given that $E_1$ and $E_2$ occur, for $\{k+\Delta \leq S_{n-T^*} - S_T \leq k+\Delta+A\}$ to occur it suffices that $r' \leq S'_{n'} \leq r'+A$. Furthermore, $-a_0\sqrt{n} \leq -m \leq k \leq a_0\sqrt{n}$, $|\Delta| < (\gamma/2)\sqrt{n}$, and $n' \geq n/2$, so we have

$$-\sqrt{2}(a_0+\gamma/2)\sqrt{n'} \leq -m - \frac{\gamma\sqrt{n'}}{\sqrt{2}} \leq -m - \frac{\gamma\sqrt{n}}{2} < r' < a_0\sqrt{n} + \frac{\gamma\sqrt{n}}{2} \leq \sqrt{2}(a_0+\gamma/2)\sqrt{n'}. \tag{2.64}$$

We may thus let $K = \sqrt{2}(a_0+\gamma/2)$, and let $a = K$ and $a' = m/\sqrt{n'} + \gamma/\sqrt{2}$; (2.64) then yields that $-a'\sqrt{n'} < r' < a\sqrt{n'}$ as required. For $\{S_{T+i} \geq -\gamma\sqrt{n}\ \forall\ T \leq i \leq n-T^*\}$ to occur given $E_1$ and $E_2$, it suffices that $S'_i \geq -(m+\gamma\sqrt{n})$ for all $1 \leq i \leq n'$. Since

$$-\frac{m+\gamma\sqrt{n}}{\sqrt{n'}} < -\frac{m}{\sqrt{n'}} - \gamma < -a' - \gamma\left(1 - \frac{1}{\sqrt{2}}\right),$$

letting $\epsilon = \gamma(1-1/\sqrt{2})$ we have that $-(m+\gamma\sqrt{n}) \leq -(a'+\epsilon)\sqrt{n'}$, so

$$\begin{aligned}
\mathbf{P}\{E_3|E_1,E_2\} &\geq \mathbf{P}\{r' \leq S'_{n'} \leq r'+A, S'_i \geq -(m+\gamma\sqrt{n})\ \forall\ 1 \leq i < n'\} \\
&\geq \mathbf{P}\left\{r' \leq S'_{n'} \leq r'+A, S'_i \geq -(a'+\epsilon)\sqrt{n'}\ \forall\ 1 \leq i < n'\right\}. \quad (2.65)
\end{aligned}$$

It follows from (2.65) and by applying Lemma 34 to $S'$ with these choices of $n'$, $a, a', r'$ and $\epsilon$ that there is $c^*$ such that

$$\mathbf{P}\{E_3|E_1,E_2\} \geq \frac{c^*}{\sqrt{n'}} \geq \frac{c^*}{\sqrt{n}}.$$

Combining this bound with our bounds on $\mathbf{P}\{E_1\}$ and $\mathbf{P}\{E_2\}$, and using the independence of $S$ on disjoint sections of the random walk, we thus have

$$\begin{aligned}
\mathbf{P}\{k \leq S_n \leq k+A, S_i \geq -m\ \forall\ 0<i<n\} &\geq \mathbf{P}\{E_1\}\mathbf{P}\{E_2\}\mathbf{P}\{E_3|E_1\cap E_2\}, \\
&\geq \frac{m^*k^*c^*}{c^2\gamma^2 n^{3/2}}, \\
&= \Omega\left(\frac{\min\{m+1,\sqrt{n}\}\min\{k+m+1,\sqrt{n}\}}{n^{3/2}}\right),
\end{aligned}$$

proving the theorem. □

*Proof of Lemma 34.* Fix any $A > 0$ satisfying the conditions of Lemma 34. Fix $K > 0$, $\epsilon > 0$, and choose $a, a'$, and $r$ as in the statement of the lemma. In brief, our argument is the following. We split the random walk up into a large but constant number of deterministic "slices" (subsections of the walk), so that the walk takes much fewer than $\epsilon n$ steps in each slice. In each slice we bound the probability that the random walk "behaves", which, roughly speaking, means that it does not dip below $-(a'+\epsilon)\sqrt{n}$ and, at the end of the slice, the value of the walk is not far from where it "should be" if we are to have $r \leq S_n \leq r+A$ (so if some slice ends at time $k$, for example, then $S_k$ is not far from $rk/n$). We show that in each slice except the last, the random walk has at least some fixed positive probability of "behaving" given that it behaved in all the previous slices. In the last step, we use the fact that the slices are extremely narrow to show that *given* that we have behaved in all previous slices, the probability we hit our desired target *and in addition dip below* $-(a'+\epsilon)\sqrt{n}$ is much smaller than the probability we hit our target. This is the picture the reader should keep in mind when working through the details below.

We let $\sigma = \sqrt{\mathbf{Var}\{X\}}$, choose some large integer $t$ and let $\delta = 1/t$. We require that $\delta$ is much smaller than $\epsilon$ and in particular that $\sigma\sqrt{\delta} < (\epsilon/2)$. We additionally require that $\sqrt{\delta}|r|$ is much smaller than $\sigma\sqrt{n}$; we will make our upper bounds on $\delta$ (which are equivalently lower bounds on $t$) more precise in the course of the proof, but emphasize that $\delta$ depends only on $X$, $K$, and $\epsilon$, and not on $n$.

Let $D_0$ be the event that $S_0 = 0$ (so $\mathbf{P}\{D_0\} = 1$), let $n_0 = 0$, and let $m = \delta n$. For $1 \le i \le t$, let $n_i = \lfloor im \rfloor$ – the $n_i$ are the boundaries of the "slices". Note that $n_t = n$ and for all $1 \le i \le t$, $m - 1 < n_i - n_{i-1} < m + 1$. For $1 \le i < t$ define the following events:

- $B_i$ is the event that $i\delta r - \sigma\sqrt{m} \le S_{n_i} \le i\delta r + \sigma\sqrt{m}$.
- $C_i$ is the event that $S_j \ge -(a'+\epsilon)\sqrt{n}$ for all $n_{i-1} < j \le n_i$, and
- $D_i$ is the event $B_i \cap C_i$. ($D_i$ is the event that the $i$'th slice "behaves".)

Note that for all $1 \le i < t$, $i\delta r > -a'\sqrt{n}$, so if $B_i$ occurs then $S_{n_i} > -a'n - \sigma\sqrt{m} = -a'n - \sigma\sqrt{\delta n} > -(a'+\epsilon)\sqrt{n}$, the last inequality holding by our choice of $\delta$. It follows that if $\cap_{i=1}^{t-1} D_i$ occurs then in particular $S_j > -(a'+\epsilon)\sqrt{n}$ for all $1 \le j \le n_{t-1}$. We claim that there is $c_0 > 0$ such that for $n$ large enough,

$$\mathbf{P}\{D_0, D_1, \ldots, D_{t-1}\} \ge c_0, \tag{2.66}$$

and that there is $\delta_0 > 0$ such that

$$\mathbf{P}\left\{r \le S_n \le r + A, S_i \ge -(a' + \epsilon)\sqrt{n} \;\forall\; n_{t-1} < i \le n \mid \cap_{i=0}^{t-1} D_i\right\} \;\ge\; \frac{\delta_0}{\sqrt{n}} \tag{2.67}$$

Combining (2.66) and (2.67) using Bayes' formula gives that

$$\mathbf{P}\left\{r \le S_n \le r + A, S_i \ge -(a' + \epsilon)\sqrt{n} \;\forall\; 1 \le i \le n\right\} \;\ge\; \frac{c_0\delta_0}{\sqrt{n}}, \tag{2.68}$$

which establishes (2.63) by letting $c^* = c_0\delta_0$. It remains to prove (2.66) and (2.67).

To prove (2.67), we write

$$\mathbf{P}\{D_0, D_1, \ldots, D_{t-1}\} \;=\; \prod_{i=1}^{t-1} \mathbf{P}\{D_i | D_{i-1}\}, \tag{2.69}$$

and bound the probabilities $\mathbf{P}\{D_i|D_{i-1}\} = \mathbf{P}\{B_i, C_i|D_{i-1}\}$ for $1 \le i \le t-1$. We do so by bounding $\mathbf{P}\{B_i|D_{i-1}\}$ from below and bounding $\mathbf{P}\left\{B_i, \bar{C}_i|D_{i-1}\right\}$ from above; we now turn to the first of these bounds.

We split the event $D_{i-1}$ into two events, depending on whether $S_{n_{i-1}}$ is in the "upper half" or the "lower half" of the interval $[(i-1)\delta r - \sigma\sqrt{m}, (i-1)\delta r + \sigma\sqrt{m}]$. If $D_{i-1}$ occurs then either $(i-1)\delta r \le S_{n_{i-1}} \le (i-1)\delta r + \sigma\sqrt{m}$, which we denote by $H$, or $(i-1)\delta r - \sigma\sqrt{m} \le S_{n_{i-1}} < (i-1)\delta r$, which we denote by $L$. If $L$ occurs then for $B_i$ to occur it suffices that $\delta r \le S_{n_i} - S_{n_{i-1}} \le \delta r + \sigma\sqrt{m}$. By the strong Markov property it follows that

$$\begin{aligned}
\mathbf{P}\{B_i|D_{i-1}, L\} &\ge \mathbf{P}\left\{\delta r \le S_{n_i - n_{i-1}} \le \delta r + \sigma\sqrt{m}\right\} \\
&= \mathbf{P}\left\{0 \le \frac{S_{n_i - n_{i-1}}}{\sigma\sqrt{m}} - \frac{\delta r}{\sigma\sqrt{m}} \le 1\right\} \\
&= \mathbf{P}\left\{0 \le \frac{S_{n_i - n_{i-1}}}{\sigma\sqrt{m}} - \frac{\sqrt{\delta} r}{\sigma\sqrt{n}} \le 1\right\}
\end{aligned} \tag{2.70}$$

We mentioned when defining $\delta$ that we required $\sqrt{\delta}|r|$ to be much smaller than $\sigma\sqrt{n}$; since $|r| \leq \max\{a, a'\}\sqrt{n} \leq K\sqrt{n}$, for any $\alpha > 0$, by choosing $\delta$ small enough we may in particular ensure that $|\sqrt{\delta}r/\sigma\sqrt{n}| < \alpha$. For such a choice of $\delta$ it follows from (2.70) that

$$\mathbf{P}\{B_i|D_{i-1}, L\} \;\geq\; \mathbf{P}\left\{\alpha \leq \frac{S_{n_i-n_{i-1}}}{\sigma\sqrt{m}} \leq 1-\alpha\right\}. \tag{2.71}$$

Since $|n_i - n_{i-1} - m| \leq 1$, $S_{n_i-n_{i-1}}/\sigma\sqrt{m}$ tends to a $\mathcal{N}(0,1)$ random variable as $n$ tends to $\infty$, and the latter probability in (2.71) tends to $\Phi(1-\alpha) - \Phi(\alpha)$, where $\Phi$ is the distribution function of a $\mathcal{N}(0,1)$ random variable; we may ensure $\Phi(1-\alpha) - \Phi(\alpha) > 1/3$ by choosing $\alpha$ small enough. For such an $\alpha$, (2.71) yields that for $n$ large enough, $\mathbf{P}\{B_i|D_i, L\} > 1/3$. An identical argument yields the same bound for $\mathbf{P}\{B_i|D_{i-1}, H\}$; it follows that for $n$ large enough, for all $1 \leq i < t$,

$$\begin{aligned}\mathbf{P}\{B_i|D_{i-1}\} &= \mathbf{P}\{B_i|D_{i-1}, L\}\,\mathbf{P}\{L|D_{i-1}\} - \mathbf{P}\{B_i|D_{i-1}, H\}\,\mathbf{P}\{H|D_{i-1}\} \\ &> \frac{1}{3}\left(\mathbf{P}\{L|D_{i-1}\} + \mathbf{P}\{R|D_{i-1}\}\right) = \frac{1}{3}.\end{aligned} \tag{2.72}$$

We next bound $\mathbf{P}\{B_i, \bar{C}_i|D_{i-1}\}$; to this end, suppose that $D_{i-1} = B_{i-1} \cap C_{i-1}$ occurs, and $B_i$ occurs but $C_i$ does not occur. Since $B_{i-1}$ occurs and $\sigma\sqrt{m} = \sigma\sqrt{\delta n} < (\epsilon/2)\sqrt{n}$ and $(i-1)\delta r > -a'\sqrt{n}$, we have

$$S_{n_{i-1}} \geq (i-1)\delta r - \sigma\sqrt{m} > -(i-1)\delta r - (\epsilon/2)\sqrt{n} > -(a' + \epsilon/2)\sqrt{n}, \tag{2.73}$$

Since $C_i$ does not occur, it follows that there must be some $n_{i-1} < j < n_i$ such that $S_j < -(a'+\epsilon)\sqrt{n}$. Since $B_i$ *does* occur, a derivation just as that of (2.73) shows that $S_{n_i} \geq -(a'+\epsilon/2)\sqrt{n}$, so $S_{n_i} - S_j > (\epsilon/2)\sqrt{n}$. Let $J$ be the *first* time $j > n_{i-1}$ that $S_j < -(a'+\epsilon)\sqrt{n}$ – then from the above comments and the strong Markov property it follows that

$$\begin{aligned}
\mathbf{P}\left\{B_i, \bar{C}_i | D_{i-1}\right\} &\leq \mathbf{P}\left\{J < n_i, S_{n_i} - S_J > \frac{\epsilon\sqrt{n}}{2}\right\} \\
&\leq \sum_{j=n_{i-1}+1}^{n_i-1} \mathbf{P}\left\{J = j\right\} \mathbf{P}\left\{S_{n_i} - S_j \geq \frac{\epsilon\sqrt{n}}{2}\right\} \\
&\leq \max_{n_{i-1}<j<n_i} \mathbf{P}\left\{S_{n_i} - S_j \geq \frac{\epsilon\sqrt{n}}{2}\right\} \\
&\leq \max_{1\leq k\leq n_i-n_{i-1}} \mathbf{P}\left\{S_k \geq \frac{\epsilon\sqrt{n}}{2}\right\}, \qquad (2.74)
\end{aligned}$$

Since $X_1 \in \mathcal{D}$, there is $\beta > 0$ such that $\mathbf{P}\left\{S_j \geq 0\right\} > \beta$ for all $j > 0$. (This follows easily from convergence to the normal distribution, and is proved in (Griffin and McConnell, 1992), for example.) It follows by the strong Markov property that for all $1 \leq k \leq n_i - n_{i-1}$,

$$\begin{aligned}
\mathbf{P}\left\{S_{n_i-n_{i-1}} \geq \frac{\epsilon\sqrt{n}}{2}\right\} &\geq \mathbf{P}\left\{S_{n_i-n_{i-1}} \geq \frac{\epsilon\sqrt{n}}{2} \middle| S_k \geq \frac{\epsilon\sqrt{n}}{2}\right\} \mathbf{P}\left\{S_k \geq \frac{\epsilon\sqrt{n}}{2}\right\} \\
&\geq \mathbf{P}\left\{S_{n_i-n_{i-1}-k} \geq 0\right\} \mathbf{P}\left\{S_k \geq \frac{\epsilon\sqrt{n}}{2}\right\} > \beta \mathbf{P}\left\{S_k \geq \frac{\epsilon\sqrt{n}}{2}\right\},
\end{aligned}$$

which combined with (2.74) gives

$$\begin{aligned}
\mathbf{P}\left\{B_i, \bar{C}_i | D_{i-1}\right\} &\leq \beta^{-1} \mathbf{P}\left\{S_{n_i-n_{i-1}} > \frac{\epsilon\sqrt{n}}{2}\right\} \\
&= \beta^{-1} \mathbf{P}\left\{\frac{S_{n_i-n_{i-1}}}{\sigma\sqrt{m}} \geq \frac{\epsilon}{2\sigma\sqrt{\delta}}\right\}. \qquad (2.75)
\end{aligned}$$

As $S_{n_i-n_{i-1}}/\sigma\sqrt{m}$ converges to a $\mathcal{N}(0,1)$ random variable, for $n$ large we may make the last probability in (2.75) as small as we like by choosing $\delta$ small enough; in particular we may therefore ensure that for $n$ large, $\mathbf{P}\left\{B_i, \bar{C}_i | D_{i-1}\right\} \leq 1/6$. It follows by this bound and by (2.72) that for $n$ large,

$$\mathbf{P}\left\{D_i | D_{i-1}\right\} = \mathbf{P}\left\{B_i | D_{i-1}\right\} - \mathbf{P}\left\{B_i, \bar{C}_i | D_{i-1}\right\} \geq \frac{1}{3} - \frac{1}{6} = \frac{1}{6}.$$

Finally, combining this bound with (2.69), we have

$$\mathbf{P}\{D_0, D_1, \ldots, D_{t-1}\} = \prod_{i=1}^{t-1} \mathbf{P}\{D_i | D_{i-1}\} \geq \frac{1}{6^{t-1}},$$

which establishes (2.66) with $c^* = 6^{-(t-1)} > 0$.

To prove (2.67), first let $n' = n - n_{t-1}$, and note that $m \leq n' < m+1$. Let $B$ be the event that $r \leq S_n \leq r + A$ and let $C$ be the event that $S_i \geq -(a' + \epsilon)\sqrt{n}\ \forall\ n_{t-1} \leq i \leq n$ – we aim to show that $\mathbf{P}\left\{B, C | \cap_{i=0}^{t-1} D_i\right\} \geq \delta_0/\sqrt{n}$. By the definitions of $B$ and $C$, establishing this bound will prove (2.67) and complete the proof of Lemma 34. We note that by the strong Markov property, $\mathbf{P}\left\{B, C | \cap_{i=0}^{t-1} D_i\right\} = \mathbf{P}\{B, C | D_{t-1}\}$. Much as in proving (2.66), we will establish our lower bound for $\mathbf{P}\{B, C | D_{t-1}\}$ by bounding $\mathbf{P}\{B | D_{t-1}\}$ from below and bounding $\mathbf{P}\left\{B, \bar{C} | D_{t-1}\right\}$ from above.

We wish to apply Theorem 27 to bound $\mathbf{P}\left\{x \leq S_n - S_{n_{t-1}} \leq x + A\right\}$; since $\sigma^2 = \mathbf{E}\{X^2\} < \infty$, and $n - n_{t-1} = n'$, by the central limit theorem we may take $a_n = \sigma\sqrt{n'}$ when applying Theorem 27. Since $n' \geq m$, by Theorem 27 there is $a^* > 0$ such that for $n$ large enough, for all $0 \leq |x| \leq 2\sigma\sqrt{m} = 2\sigma\sqrt{\delta n}$, we have $\mathbf{P}\left\{x \leq S_n - S_{n_{t-1}} \leq x + A\right\} \geq a^*/\sqrt{n'}$. We recall that we chose $\delta$ small enough that $\sqrt{\delta}|r| < \sigma\sqrt{n}$, and note that if $D_{t-1}$ holds then since also $(t-1)\delta r = r - \delta r$,

$$|S_{n_{t-1}} - r| \leq |S_{n_{t-1}} - (t-1)\delta r| + |(t-1)\delta r - r| \leq \sigma\sqrt{m} + \delta|r| < \sigma\sqrt{m} + \sqrt{\delta}(\sigma\sqrt{n}) = 2\sigma\sqrt{m}.$$

By Theorem 27 and the strong Markov property, we therefore have

$$\mathbf{P}\left\{B | D_{t_{i-1}}\right\} = \mathbf{P}\left\{r - S_{n_{t-1}} \leq S_n - S_{n_{t-1}} \leq r + A - S_{n_{t-1}}\right\} \geq \frac{a^*}{\sqrt{n'}}. \qquad (2.76)$$

We bound $\mathbf{P}\left\{B, \bar{C} | D_{t-1}\right\}$ from above in much the same fashion as we bounded $\mathbf{P}\left\{B_i, \bar{C}_i | D_{i-1}\right\}$. The intuition of the bound is that if $D_{t-1}$ occurs then for $B$ and $\bar{C}$ to occur the walk must first dip far below its mean and then end up in a *specific interval* of length $A$ at time $n$. The fact that the walk must end up in a specific small interval allows us to use Theorem 9 to obtain upper bounds on $\mathbf{P}\left\{B,\bar{C}|D_{t-1}\right\}$ that are a factor of $\sqrt{n}$ stronger than our bounds on $\mathbf{P}\left\{B_i,\bar{C}_i|D_{i-1}\right\}$. To apply Theorem 9, however, we end up having to split our bound into two parts, separately bounding the events that the walk $S_{n_{t-1}+1},\ldots,S_n$ dips below $-(a'+\epsilon)\sqrt{n}$ in the first half or in the second half of its steps. We now formalize this sketch.

Let $S'$ be the random walk given by $S'_i = S_{n_{t-1}+i} - S_{n_{t-1}}$, for $0 \leq i \leq n'$, and let $S^r$ be the reversed walk given by $S^r_0 = 0$ and for $i \geq 0$, $S^r_{i+1} = S^r_i - X_{n-i}$, also for $0 \leq i \leq n'$. Given that $D_{t-1}$ and $B$ both occur, $S_{n_{t-1}} > -(a'+\epsilon/2)\sqrt{n}$ and $S_n > -a'\sqrt{n}$. If additionally $C$ does *not* occur, then there is $1 \leq j < n'$ such that $S_{n_{t-1}+j} \leq -(a'+\epsilon)\sqrt{n}$. Given $D_{t-1}$ and $B$, for $C$ to occur one of the following two events must therefore occur:

- either there is $1 \leq j \leq \lfloor n'/2 \rfloor$ such that $S'_j \leq -(\epsilon/2)\sqrt{n}$, and

$$r - S_{n_{t-1}+\lfloor n'/2 \rfloor} \leq S'_{n'} - S'_{\lfloor n'/2 \rfloor} \leq r + A - S_{n_{t-1}+\lfloor n'/2 \rfloor},$$

- or there is $1 \leq j \leq \lceil n'/2 \rceil$ such that $S^r_j \leq -\epsilon\sqrt{n}$, and

$$r - S_{n_{t-1}} - (S_n - S_{n-\lfloor n'/2 \rfloor - 1}) \leq S^r_{n'} - S^r_{\lfloor n'/2 \rfloor} \leq r + A - S_{n_{t-1}} - (S_n - S_{n-\lfloor n'/2 \rfloor - 1}).$$

We denote these two events $A'$ and $A^r$, respectively. We remark that the rather complicated two-sided inequalities in the definitions of $A'$ and $A^r$ are both equivalent, after rearrangement, to the condition that $r \leq S_n \leq r + A$. We have written them as we did in order to highlight that we use this as a condition on the difference $S'_{n'} - S'_{\lfloor n'/2 \rfloor}$ (when defining $A'$) and on the difference $S^r_{n'} - S^r_{\lfloor n'/2 \rfloor}$ (when defining $A^r$). With these definitions, we thus have $\mathbf{P}\left\{B,\bar{C}|D_{t-1}\right\} \leq 2\max\{\mathbf{P}\left\{A'|D_{t-1}\right\}, \mathbf{P}\left\{A^r|D_{t-1}\right\}\} = 2\max\{\mathbf{P}\left\{A'\right\}, \mathbf{P}\left\{A^r\right\}\}$, the preceding equality holding by the strong Markov property. We next prove that $\max\{\mathbf{P}\left\{A'\right\}, \mathbf{P}\left\{A^r\right\}\}$

is at most $a^*/4\sqrt{n'}$, from which it follows that $\mathbf{P}\left\{B, \bar{C}|D_{t-1}\right\} \leq a^*/2\sqrt{n'}$. Combining this bound with (2.76) yields that $\mathbf{P}\left\{B, C|D_{t-1}\right\} \geq a^*/2\sqrt{n'} > a^*/2\sqrt{n}$, which proves (2.67) with $\delta_0 = a^*/2$ and thus completes the proof. It remains to prove our bounds on $\mathbf{P}\left\{A^l\right\}$ and on $\mathbf{P}\left\{A^r\right\}$.

For $1 \leq j \leq n'/2$, let $A_j$ be the event that $j$ is the smallest integer for which $S'_j \leq -(\epsilon/2)\sqrt{n}$. The $A_j$ are disjoint events so $\sum_{j=1}^{\lfloor n'/2 \rfloor} \mathbf{P}\left\{A_j\right\} \leq 1$. Furthermore, if $A_j$ occurs then either $S'_{\lfloor n'/2 \rfloor} \leq -\epsilon\sqrt{n}/4$, or else $S'_{\lfloor n'/2 \rfloor} - S'_j \geq \epsilon\sqrt{n}/4$. Letting $n^* = \lfloor n'/2 \rfloor$, we thus have

$$
\begin{aligned}
\mathbf{P}\left\{\exists 1 \leq j \leq n^* \text{ s.t. } S'_j \leq \frac{-\epsilon\sqrt{n}}{2}\right\} &\leq \mathbf{P}\left\{S'_{n^*} \leq -\frac{\epsilon\sqrt{n}}{4}\right\} \\
&\quad + \sum_{i=1}^{n^*-1} \mathbf{P}\left\{A_j\right\} \mathbf{P}\left\{S'_{n^*} - S'_j \geq \frac{\epsilon\sqrt{n}}{4}\right\} \\
&\leq \mathbf{P}\left\{|S'_{n^*}| \geq \frac{\epsilon\sqrt{n}}{4}\right\} + \max_{1 \leq j \leq n^*-1} \mathbf{P}\left\{S'_{n^*-j} \geq \frac{\epsilon\sqrt{n}}{4}\right\},
\end{aligned}
$$

the second inequality holding by the strong Markov property and as $\sum_{j=1}^{\lfloor n'/2 \rfloor} \mathbf{P}\left\{A_j\right\} \leq 1$. Using the fact that there is $\beta > 0$ such that $\mathbf{P}\left\{S'_i \geq 0\right\} \geq \beta$ for all $i$ and a simple conditioning argument, just as we did in proving (2.75), then yields that

$$
\begin{aligned}
\mathbf{P}\left\{\exists 1 \leq j \leq n^* \text{ s.t. } S'_j \leq \frac{-\epsilon\sqrt{n}}{2}\right\} &\leq (1+\beta^{-1})\mathbf{P}\left\{|S'_{n^*}| \geq \frac{\epsilon\sqrt{n}}{4}\right\}, \\
&= (1+\beta^{-1})\mathbf{P}\left\{\frac{|S'_{n^*}|}{\sigma\sqrt{n'/2}} \geq \frac{\epsilon\sqrt{n}}{4\sigma\sqrt{n'/2}}\right\} \\
&\leq (1+\beta^{-1})\mathbf{P}\left\{\frac{|S'_{n^*}|}{\sigma\sqrt{n'/2}} \geq \frac{\epsilon}{4\sigma\sqrt{\delta}}\right\}. \qquad (2.77)
\end{aligned}
$$

Since $n^* = \lfloor n'/2 \rfloor$, $S'_{n^*}/\sigma\sqrt{n'/2}$ converges to a $\mathcal{N}(0,1)$, by choosing $\delta$ small we may make the last probability in (2.77) as small as we wish, for all $n$. In particular, fixing any $\delta_1 > 0$ we may presume we have chosen $\delta$ small enough that (2.77) yields

$$
\mathbf{P}\left\{\exists 1 \leq j \leq \lfloor n'/2 \rfloor \text{ s.t. } S'_j \leq \frac{-\epsilon\sqrt{n}}{2}\right\} \leq \delta_1. \qquad (2.78)
$$

By Theorem 9 there is a constant $a^{**}$ depending only on $X$ and $A$ such that for all $x$ and all $k$, $\mathbf{P}\{x \le S_k \le x + A\} \le a^{**}/\sqrt{k}$. In particular, we thus have

$$\mathbf{P}\left\{r - S_{n_{t-1}+\lfloor n'/2\rfloor} \le S'_{n'} - S'_{\lfloor n'/2\rfloor} \le r + A - S_{n_{t-1}+\lfloor n'/2\rfloor}\right\} \le \frac{a^{**}}{\sqrt{\lfloor n'/2\rfloor}} \le \frac{2a^{**}}{\sqrt{n'}} \qquad (2.79)$$

for $n$ large enough. The events in whose probabilities are bounded in (2.78) and (2.79) are determined on disjoint sections of the random walk $S'$. Since both must occur in order that $A'$ occur, we thus have

$$\mathbf{P}\{A'\} \le \frac{2\delta_1 a^{**}}{\sqrt{n'}}.$$

Choosing $\delta_1 = a^*/8a^{**}$, we thus have that $\mathbf{P}\{A'\} \le a^*/4\sqrt{n'}$ for $n$ large enough. An identical argument to bound $\mathbf{P}\{A^r\}$ yields that $\mathbf{P}\{A^r\} \le a^*/4\sqrt{n'}$ for $n'$ large enough, so $\max\{\mathbf{P}\{A'\}, \mathbf{P}\{A^r\}\} \le a_1/4\sqrt{n'}$, as claimed. This completes the proof. □

We note that Lemma 34 has an exact analogue in the case that $X_1 \in \mathcal{D}$ and $S_n/a_n$ tends to a normal distribution; in this case, by an identical proof, we end up with a lower bound of $c^*/a_n$ instead of $c^*/\sqrt{n}$ in (2.63). We can then establish Theorem 33 in exactly the same fashion as we did Theorem 32; once again, we omit the details. This completes our work on general ballot theorems for the normal case. We now turn our attention to a ballot-style result that is is interesting when $S_n$ is much farther from its mean.

## 2.5 Ballot theorems for landslide elections

In this section we consider what meaningful ballot-style statements can be made when $S_n - \mathbf{E}\{S_n\} = \Theta(n)$. The central limit theorem suggests that if $X$ is in the range of attraction of the normal, then $S_n - \mathbf{E}\{S_n\}$ is likely $\Theta(\sqrt{n})$, and it is for such random variables and such deviations that the results of Chapter 2 are essentially optimal. The results of this section apply to a different kind of random variable; we will restrict our attention to sums of iid random variables with positive variance whose mean is 0 (as usual) and whose maximum value is 1. (Of course, any non-negative random variable that is bounded from above can be renormalized to satisfy this constraint.) We place no restriction on the lower tail behavior of the random variables – in particular, it is possible that $\mathbf{E}\{X^2\} = \infty$

This is a setup we have already seen in Chapter 1. For example, if $X$ is integer valued we saw that $\mathbf{P}\{S_i > 0 \ \forall\ 1 \le i \le n | S_n = k\} = k/n$. In this chapter we are interested in a related though distinct question: given that $S_n = k > 0$, what is the probability that $S_i < k \cdot (i/n)$ for all $0 < i < n$? More generally, for some positive $a$ possibly depending on $n$ and under the same conditioning, what is the probability that $S_i < k \cdot (i/n) + a$ for all $0 < i < n$? The answer to the first question follows from the rotation argument we have already used several times. Let $\hat{X}_i = X_i - k/n$ and set $\hat{S}_i = \hat{X}_1 + \ldots + \hat{X}_i = S_i - ik/n$ – note that in particular $\hat{S}_n = 0$. The variables $\hat{X}_1, \ldots, \hat{X}_n$ are interchangeable, and $S_i < S_n \cdot (i/n) = ki/n$ for all $0 < i < n$ precisely if $\hat{S}_i < 0$ for all $0 < i < n$. Since $\hat{S}_n = 0$, the probability of the latter event is $\Theta(1/n)$, as we saw in Section 2.1.

The results and techniques we have seen so far in this chapter suggest it might be reasonable to believe that for suitable $a$, the probability that $S_i < k \cdot (i/n) + a$ for all $0 < i < n$, given that $S_n = k$, is $O(a^2/n)$. We are not able to prove this, and have to settle for a somewhat weaker result (which, roughly stated, consists in replacing $O(a^2/n)$ by $O(a^5/n)$). In a nutshell, the weakness in our argument is due to the fact that in addition to the rotation argument, we end up relying on asymptotic estimates of the probability that $S_n > k$ for $k = \Omega(n)$ (i.e., *asymptotic estimates for large deviations*). These estimates introduce a sort of "impurity" into the ballot-style argument, which interferes with our ability to apply the techniques we used above. Our intuition is that this reliance on asymptotic estimates for large deviations should be unnecessary and that its removal would lead to a strengthening of our results. We remark that Reed (2003) has proved a similar result for random walks in which the steps are exponential mean 1 random variables, and we adopt his approach in this section.

At a high level, our argument is no different from that sketched in Section 2.1 (for the moment, we assume we are dealing with integer random variables that are never 0). Let us say that $S_n$ *stays below* $a$ (and denote this event $S\ bel\ a$) if, for all $0 < i < n$, $S_i < S_n(i/n)+a$. We are then interested in comparing $\mathbf{P}\{S_n = \epsilon n \cap S_n\ bel\ a\}$ with $\mathbf{P}\{S_n = \epsilon n\}$ for $0 < \epsilon < 1$ and for "suitable" $a$. We observe that if $S_{a^2} = \epsilon a^2 - a$, say, and $S_n - S_{n-a^2} = \epsilon a^2 + a$, then for $S_n = \epsilon n$ and $S_n\ bel\ 0$ to occur, it suffices that in addition, the following events occur:

(a) $S_i < \epsilon i$ for $0 < i < a^2$,

(b) $S_n - S_{n-i} > \epsilon i$ for $0 < i < a^2$,

(c) $S_{n-a^2} - S_{a^2} = \epsilon(n - 2a^2)$, and $S_{a^2+i} - S_{a^2} < \epsilon i + a$ for $0 < i < n - 2a^2$.

The third of these events is just the event $\{S^*_{n'} = \epsilon n' \cap S^*\ bel\ a\}$, where $S^*$ is just $S$ restarted at time $a^2$, and $n' = n - 2a^2$. By the standard rotation argument, the probability of (a) given that $S_{a^2} < \epsilon a^2$ is at least $1/a^2$. Similarly, the conditional probability of (b) is at least $1/a^2$. If the events $S_{a^2} = \epsilon a^2 - a$ and $S_n - S_{n-a^2} = \epsilon a^2 + a$ both had probability $\Omega(1)$, then the independence of disjoint sections of the random walk would yield:

$$\mathbf{P}\{S_n = \epsilon n \cap S\ bel\ 0\} = \mathbf{P}\left\{S^*_{n-2a^2} = \epsilon(n-2a^2) \cap S\ bel\ a\right\} \cdot \Omega\left(\frac{1}{a^4}\right).$$

Since we also know that the left-hand-side of the above equation is $\Theta(\mathbf{P}\{S_n = \epsilon n\}/n)$, it would follow from the above equation that

$$\mathbf{P}\left\{S_{n-2a^2} = \epsilon(n-2a^2) \cap S_{n-2a^2}\ bel\ a\right\} = O\left(\frac{a^4}{n}\right)\mathbf{P}\{S_n = \epsilon n\}. \qquad (2.80)$$

Finally, if additionally $\mathbf{P}\{S_{n-2a^2} = \epsilon(n-2a^2)\}$ and $\mathbf{P}\{S_n = \epsilon n\}$ were of the same order, then (2.80) would yield the sort of statement we are aiming to prove. The key way in which our sketch is incorrect is that $\mathbf{P}\{S_{n-2a^2} = \epsilon(n-2a^2)\}$ and $\mathbf{P}\{S_n = \epsilon n\}$ are significantly different – as we shall see later, they differ by a factor of roughly $e^{c \cdot 2a^2}$ for some $c = c(\epsilon) > 0$ – so (2.80) would not yield a statement of the form $\mathbf{P}\{S_n \ bel \ a | S_n = \epsilon n\} = O(a^4/n)$, but instead, a bound which is incorrect. Similarly, the events $S_{a^2} = \epsilon a^2 - a$ and $S_n - S_{n-a^2} = \epsilon a^2 + a$ do not have probability $\Omega(1)$.

Fortunately, bounds on the probabilities of these events follow from large deviations asymptotics. Though the necessity of these estimates complicates the proofs that follow, the complication is purely technical, and turning the above sketch into a proof is by-and-large an exercise in using them at the appropriate points of the argument. The requisite results on large deviations appear in Appendix A; we combine the results from that section which we will need into the following lemma.

**Lemma 35.** *Let $S$ be a random walk with iid steps distributed as $X$ with $\mathbf{E}X = 0$, $\mathbf{Var}\{X\} > 0$, and $\sup\{x \mid \mathbf{P}\{X > x\} > 0\} = 1$. Then for any $0 < \epsilon < 1$ there are positive constants $b, c$, and $d$, and $f$ such that*

$$\mathbf{P}\{S_n > \epsilon n\} = (1+o(1))\frac{b}{\sqrt{n}}e^{-cn},$$

*and for $a, a' = o(\sqrt{n})$ (with $a'$ an integer),*

$$\mathbf{P}\{S_{n-a'} > \epsilon n + a\} = (1+o(1))e^{-ad-a'f}\mathbf{P}\{S_n > \epsilon n\},$$

*where for any $g(n)$ tending to infinity with $n$, the convergence of the term $o(1)$ is uniform over $|a|, |a'| < \sqrt{n}/g(n)$.*

A caveat: this lemma applies only when $X$ is *non-lattice*, i.e., there is no $r \neq 0$ for which $rX$ is an integer random variable. As discussed in Appendix A, an equivalent lemma exists when $X$ is lattice. Though we prove the following result using Lemma 35, we state it more generally and omit the proof for lattice random variables to avoid a near-verbatim repetition of the non-lattice proof. We remark that the condition $\sup\{x \mid \mathbf{P}\{X > x\} > 0\} = 1$ can be replaced by the condition "$X$ is bounded from above" by renormalizing.

**Theorem 36.** *Let $X_1, \ldots, X_n$ be iid random variables with mean 0, positive variance and maximum value 1, and let $S_i = X_1 + \ldots + X_i$ for $1 \le i \le n$. Fix constants $\alpha > 0$, $0 < \epsilon < 1$, and choose real numbers $a, r$ with $1 \le a = O(n^{1/5})$ and $r = o(\sqrt{n})$. If $X_1$ is non-lattice, then*

$$\mathbf{P}\{S_n \ bel\ a \mid \epsilon n + r \le S_n \le \epsilon n + r + \alpha\} = O\left(\frac{a^5}{n}\right), \tag{2.81}$$

*any $g(n)$ tending to infinity with $n$, the convergence of the term $o(1)$ is uniform over $|r| < \sqrt{n}/g(n)$. Furthermore, if $X_1$ is lattice then for any $\alpha$ for which $\alpha X_1$ is integer, the same result holds.*

*Proof.* As discussed, we restrict our attention to the case that $X_1$ is non-lattice. We prove the lemma assuming $\alpha = 1$ and $r = 0$; the proof is identical for general $\alpha$ and $r$. Fix $n$ and let $n^* = n + 2a^2$; as $a = O(n^{1/5})$ we have $n = \Theta(n^*)$. We let $S'$ be the walk with $S'_i = S_{a^2+i} - S_{a^2}$. As in the above sketch, we proceed by comparing the following two probabilities:

$$\mathbf{P}\{\epsilon n^* \le S_{n*} \le \epsilon n^* + 3 \cap S_{n^*} \ bel\ 0\} \qquad \text{and} \qquad \mathbf{P}\{\epsilon n \le S'_n \le \epsilon n + 1 \cap S'_n \ bel\ a\}.$$

Let $E$ (resp. $E^*$) be the event $\{\epsilon n \le S'_n \le \epsilon n + 1\}$ (resp. $\{\epsilon n^* \le S_{n^*} \le \epsilon n^* + 3\}$). For any $0 \le i \le a - 1$, for $E^* \cap \{S_{n^*}\ bel\ 0\}$ to occur it suffices that the following events occur:

$E_{1,i}$ is the event that $-(a+i) \le S_{a^2} - \epsilon a^2 \le -(a+i-1)$.

$E_{2,i}$ is the event that $a + i \le S_{n^*} - S_{n^*-a^2} - \epsilon a^2 \le a + i + 1$.

$E_3$ is the event that $\epsilon n \le S'_n \le \epsilon n + 1$.

Bel is the event that $S_{a^2}\ bel\ 0$ and, letting $S^*_j = S_{n^*-a^2+j} - S_{n^*-a^2}$, that $S^*_{a^2}\ bel\ 0$.

$\mathrm{Bel}_a$ is the event that $S'_n$ *bel* $a$.

We will show that

$$\begin{aligned}\mathbf{P}\{\mathrm{Bel}_a \mid E_3\} &= O\left(\frac{a^5}{n}\right), \text{i.e.,}\\ \mathbf{P}\{S'_n \textit{ bel } a | \epsilon n \le S'_n \le \epsilon n + 1\} &= O\left(\frac{a^5}{n}\right),\end{aligned}$$

which proves (2.81) with $\alpha = 1$ as the walk $S'$ is distributed identically to the walk $S$; as noted above, the proof for general $\alpha$ is identical.

Since for any $i \le 3a$, $a + i/3$ is $O(a)$, by Lemma 35 there is $d > 0$ such that

$$\mathbf{P}\{E_{1,i}\} = \Theta(e^{(a+i)d})\mathbf{P}\left\{\epsilon a^2 \le S_{a^2} \le \epsilon a^2 + 1\right\},$$

and

$$\mathbf{P}\{E_{2,i}\} = \Theta(e^{-(a+i)d})\mathbf{P}\left\{\epsilon a^2 \le S_{a^2} \le \epsilon a^2 + 1\right\}.$$

As $E_{1,i}$ and $E_{2,i}$ are independent and $\mathbf{P}\{\epsilon a^2 \le S_{a^2} \le \epsilon a^2 + 1\}$ is $\Theta(\mathbf{P}\{S_{a^2} \ge \epsilon a^2\})$ by Lemma 35, it follows that

$$\mathbf{P}\{E_{1,i} \cap E_{2,i}\} = \Theta(\mathbf{P}\{S_{a^2 \ge \epsilon a^2}\}^2) = \Theta\left(\frac{b^2}{a^2}e^{-2ca^2}\right), \tag{2.82}$$

for some constants $b$ and $c$, again by Lemma 35. Also, the standard rotation argument and the strong Markov property yield that $\mathbf{P}\{Bel | E_{1,i}, E_{2,i}\} = \Theta(1/a^4)$. Combining this fact and (2.82) and again using the strong Markov property we have

$$\mathbf{P}\{E_{1,i} \cap E_{2,i} \cap E_3 \cap \mathrm{Bel} \cap \mathrm{Bel}_a\} = \Theta\left(\frac{b^2}{a^6}e^{-2ca^2}\right) \cdot \mathbf{P}\{E_3 \cap \mathrm{Bel}_a\}. \tag{2.83}$$

Since $E^* \cap \{S_{n^*} \textit{ bel } 0\}$ occurs if the above conjunction of events occurs for *any* $1 \le i \le a-1$, and for $i \neq j$ the events $E_{1,i}$ and $E_{1,j}$ are disjoint, by summing over $i$ in (2.83) it follows that

$$\mathbf{P}\{E^* \cap \{S_{n^*} \text{ bel } 0\}\} = \Omega\left(\frac{b^2}{a^5}e^{-2ca^2}\right) \cdot \mathbf{P}\{E_3 \cap \text{Bel}_a\} \tag{2.84}$$

The standard rotation argument and Lemma 35 yield that

$$\mathbf{P}\{E^* \cap \{S_{n^*} \text{ bel } 0\}\} = \Theta\left(\frac{1}{n^*}\right)\mathbf{P}\{E^*\} = \Theta\left(\frac{b}{(n^*)^{3/2}}e^{-cn^*}\right) = \Theta\left(\frac{b}{n^{3/2}}e^{-c(n+2a^2)}\right),$$

which, combined with (2.84), yields that

$$\mathbf{P}\{E_3 \cap \text{Bel}_a\} = O\left(\frac{a^5 e^{-cn}}{bn^{3/2}}\right). \tag{2.85}$$

A final application of Lemma 35 to bound $\mathbf{P}\{E_3\}$ gives

$$\mathbf{P}\{E_3\} = \Theta\left(\frac{b}{\sqrt{n}}e^{-cn}\right). \tag{2.86}$$

As $b$ is a constant, (2.85) and (2.86) together imply $\mathbf{P}\{\text{Bel}_a \mid E_3\} = O(a^5/n)$. □

## 2.6 Conclusion

In this this chapter and the last, I aimed to demonstrate that the theory of ballots is not only rich and beautiful, in-and-of itself, but is also very much alive. The results of this chapter are far from conclusive in terms of when ballot-style behavior can be expected of sums of independent random variables, and more generally of permutations of sets of real numbers. In the next few paragraphs, I highlight some of the questions that remain unanswered.

The results of Section 2.4 are unsatisfactory in that they only yield "true" (conditional) ballot theorems when $S_n = O(\sqrt{n})$. Ideally, we would like results such Corollaries 28 and 29 to hold *whatever* the range of $S_n$. The weakness of our approach is that it relies on estimates for $\mathbf{P}\{x \leq S_n \leq x+c\}$ that are based on the central limit theorem, and these estimates are not good enough when $S_n$ is not $O(\sqrt{n})$. There is also room to improve the results of Section 2.4 for $S_n = O(\sqrt{n})$. When we only assume $X \in \mathcal{D}$, the term $n^{1-o(1)}$ in our result is surely not optimal, and should be replaced by $\Theta(n)$.

The restriction to variables $X \in \mathcal{D}$ is not obviously necessary in order for ballot-style behavior to occur. On the other hand, the proof technique used above can not be generalized very much. Kesten and Maller (1994) and, independently, Griffin and McConnell (1994), have derived necessary and sufficient conditions in order that $\mathbf{P}\left\{S_{T_{r,0}}\right\} \to 1/2$ as $r \to \infty$; in particular they show that for any $\alpha < 2$, there are distributions with $\mathbf{E}\{X^\alpha\} = \infty$ for which $\mathbf{P}\left\{S_{T_{r,0}}\right\} \to 1$. Therefore, we can not expect to use a result such as Theorem 19 in this case, which seriously undermines our approach. The results of Section 2.5 do hold for variables with infinite variance; however, they give bounds with respect to the conditioned mean. It would be nice to have a ballot theorem, in this case, that bounded the probability that the *actual* expected value stayed above its mean.

In terms of the sort of conditional expected value ballot theorem we saw in Section 2.5, there are also unanswered questions. As we noted in Section 2.5, Theorem 36 should hold for $a = O(n^{1/2})$ (instead of $a = O(n^{1/5})$), with the bound $O(a^5/n)$ replaced by $O(a^2/n)$. With a little more work, we could perhaps have forced our approach to yield a bound of $O(a^4/n)$; however, the correct upper bound seems beyond the reach of our approach. It is also unclear just when the approch of Section 2.5 can be made to work at all if the one-sided boundedness condition is relaxed. (We could have generalized our results somewhat, but only by imposing analytic conditions on the random variable $X$ that have nothing to do with the ballot theorem on any intuitive level.)

As we touched upon at various points in the chapter, aspects of our technique seem as though they should work for analyzing more general random permutations of sets of real numbers. Since Andersen observed the connection between conditioned random walks and random permutations (Andersen, 1953, 1954), and Spitzer (1956) pointed out the full generality of Andersen's observations, just about every result on conditioned random walks has been approached from the permutation-theoretic perspective sooner or later. There is no reason the results of this chapter should not benefit from such an approach.

# Chapter 3

# The size and structure of the incipient giant component of $G_{n,p}$

## 3.1 Introduction

Random graphs lie at the heart of probabilistic combinatorics, but were first introduced to answer a completely deterministic problem. Perhaps the most famous classical theorem in combinatorics is the theorem of Ramsey (1929), a special form of which states that for any integer $k$ there is $n$ such that given any graph $G$ with $n$ vertices, either $G$ contains a clique of size $k$ or an independent set of size $k$. Ramsey in particular proved that, viewed as a function of $n$, we may take $k > (\log n)/2$. In other words, any graph with $n$ vertices contains either a clique or an independent set of size at least $(\log n)/2$.

Erdös (1947) used random graphs to give a simple proof that this lower bound on $k$ is within a factor of 4 of the best possible bound, by showing that there exist graphs on $n$ vertices containing neither a clique nor an independent set of size greater than $2 \log n$; we present his proof in a slightly altered form. Let $G_{n,1/2}$ be a labeled graph with $n$ vertices chosen uniformly at random from among all such graphs. We may think of constructing $G_{n,1/2}$ starting from $n$ isolated vertices by flipping a fair coin for each edge and adding the edge if the coin comes up heads. In other words, in $G_{n,1/2}$ each edge is present independently and with probability exactly $1/2$.

We may easily bound the probability that $G_{n,1/2}$ contains a clique of size $k$. The probability a *given* set of $k$ vertices in $G_{n,1/2}$ is a clique is $2^{-\binom{k}{2}}$, as each of the pairs within the $k$ vertices must be joined by an edge. By a union bound, the probability *any* $k$ vertices form a clique is at most $\binom{n}{k}2^{-\binom{k}{2}} \leq n^k 2^{-k(k-1)/2}/k!$. If $n > 2$ and $k > 2\log n$, this is strictly less than $1/2$. By considering the complement of $G_{n,1/2}$ (the graph whose edges are the non-edges of $G_{n,1/2}$), an identical bound follows for the size of the largest independent set. Thus the probability that $G_{n,1/2}$ contains a clique of size $k > 2\log n$ or an independent set of size $k$ is strictly less than 1, so there is some graph with $n$ vertices that contains neither of these subgraphs. (We remark that this may also have been the first use of the probabilistic method.)

The notation $G_{n,1/2}$ highlights the fact that we choose each edge independently with probability $1/2$; the more general model $G_{n,p}$ (for $0 \leq p \leq 1$) chooses each edge independently with probability $p$. In the closely related model $G_{n,m}$, we select uniformly at random from among all graphs with $m$ edges. One may pose the question: *how does the structure of* $G_{n,m}$ (resp. $G_{n,p}$) *change as* $m$ *ranges from* 0 *to* $\binom{n}{2}$ (resp. $p$ ranges from 0 to 1)?

The seminal papers addressing this issue are due to Erdős and Rényi (1959, 1960, 1961). These papers analyzed the random graph model $G_{n,m}$, establishing the remarkable fact that many commonly studied graph properties have a *threshold function* in $G_{n,m}$. To explain precisely what we mean, it helps to fix a specific example; we consider the question of whether $G_{n,m}$ is connected. Erdős and Rényi (1959) showed that, letting $m = m(n) = (n \log n)/2$, for all $\epsilon = \epsilon(n) > 0$ for which $\epsilon(n) = \omega(1/\log n)$, $G_{n,(1+\epsilon)m}$ is connected with probability tending to 1 as $n$ tends to infinity (*asymptotically almost surely*, or a.a.s.), and a.a.s, $G_{n,(1-\epsilon)m}$ is not connected. This paper also found asymptotic bounds for the probability that $G_{n,(n\log n+cn)/2}$ is connected, for fixed $c$, in particular bounding this probability away from both 0 and 1. Adopting a piece of terminology from statistical physics, we say that $m = (n\log n + cn)/2$ is in the *critical window* for connectivity.

The subjects addressed in (Erdős and Rényi, 1960) are close to the heart of this chapter. It was established in that paper that for any fixed $\epsilon > 0$, if $m = m(n) < (1-\epsilon)n/2$ then a.a.s the largest component of $G_{n,m}$ has size $O(\log n)$. If $m > (1+\epsilon)n/2$ then a.a.s.:

($\star$) The largest component $H_{n,m}$ of $G_{n,m}$ has size $\Omega(n)$ and all other components have size $O(\log n)$.

(In choosing $H_{n,m}$, we break ties by choosing the largest component so as to maximize the largest vertex index it contains – thus this component is unique.) Finally, they showed that if $m = n/2$ then the largest component of $G_{n,m}$ has size $\Theta(n^{2/3})$ and there may be many components of this order.

It is possible to consider the of random graphs $\{G_{n,m}\}_{m=0}^{\binom{n}{2}}$ as a *graph process*: we label the edges of the graph with the integers $L = \{1, \ldots, \binom{n}{2}\}$, chose a uniformly random permutation $\sigma$ of $L$, and sequentially add the edges in the order given by $\sigma$. The in this process, for each $0 \le m \le \binom{n}{2}$, the graph with edges $\sigma(1), \ldots, \sigma(m)$ is distributed as $G_{n,m}$. This process has the property that for $m < m'$, $G_{n,m}$ is a subgraph of $G_{n,m'}$, which is often useful for analysis.

In fact, (Erdős and Rényi, 1960) established a stronger result than that stated above: they showed that in the random graph process for $G_{n,m}$, a.a.s. ($\star$) holds for *all* $m > (1+\epsilon)n/2$. This implies that a.a.s., for all $m' > m > (1+\epsilon)n/2$, $H_{n,m} \subseteq H_{n,m'}$. Restating this in colourful but imprecise language, which component is the largest never changes, the largest component simply expands by gobbling up smaller ones. When $m > (1+\epsilon)n/2$, $H_{n,m}$ is called the *giant component.*

We remark that in the random graph model $G_{n,p}$, the number of edges is distributed as $\text{Bin}(\binom{n}{2}, p)$, and is thus with high probability about $\binom{n}{2}p + O(\sqrt{n^2p})$; this fact allows us to translate results for one model into results for the other with relative ease. In particular, it follows from bounds on the tails of the binomial distribution that given a function $M(n)$ and a graph property $P$, if for all $m = O(\sqrt{M(n)})$, a.a.s. $G_{n,M(n)+m}$ has $P$, then a.a.s. $G_{n,p}$ has $P$, and is not difficult to see that the converse also holds. It turns out that $G_{n,p}$ is often easier to study than $G_{n,m}$ because of the independence between edges and, as we shall see, because for certain values of $p$ the behavior of $G_{n,p}$ can be analyzed via a branching process. For the remainder of the chapter, we phrase our discussion in terms of $G_{n,p}$, even when stating results originally proved for $G_{n,m}$.

The aim of this chapter is to understand aspects of the structure of $G_{n,p}$ when $p$ is in the critical window for the existence of a giant component, i.e., when $p - 1/n = o(1/n)$. (For the remainder of this chapter, we will refer to this as "the" critical window, as we will consider no others). More precisely, we aim to establish explicit (i.e. not asymptotic) bounds on the size and number of edges of the largest component $H_{n,p}$ of $G_{n,p}$ when $p - 1/n = o(1/n)$. We return to this question after providing a brief history of the study of the giant component.

The results of Erdős and Rényi (1960) were rediscovered by Stepanov (1970a,b), who worked directly with the $G_{n,p}$ model. (Stepanov was the first person to refer to the threshold phenomena appearing in $G_{n,p}$ as "phase transitions", another term borrowed from statistical physics). In the course of his work, Stepanov also introduced the *continuous time* random graph process, in which each edge $e$ of the complete graph $K_n$ is assigned an independent copy $X_e$ of some continuous, non-negative random variable $X$, and the graph $G_n(t)$ consists of all edges $e$ with $X_e \leq t$. We note that if $X$ is a $[0,1]$ random variable, then for $0 \leq p \leq 1$, $G_n(p)$ is distributed as $G_{n,p}$. (This process allows us to think of the graph $G_{n,p}$ as "evolving" over time, and will be important both in this chapter and in Chapter 4.) Ivchenko (1973a,b) built on the work of Stepanov; his work was primarily focussed on the critical window for connectivity and not for the existence of a giant.

The first major progress on the behavior of $G_{n,p}$ in the critical window was due to Bollobás (1984). He showed that for any function $h(n) > 0$ which is $\omega((\log n)^{1/2} n^{-4/3})$, a.a.s. for all $p > 1/n + h(n)$,

(a) $|H_{n,p}| = (4 + o(1))n^2 h(n)$, and all other components have size $o(n^{2/3})$, and

(b) for all $p' > p$, $H_{n,p} \subseteq H_{n,p'}$.

It follows that for such $p$ it already makes sense to refer to $H_{n,p}$ as "the" giant component. Łuczak (1990, 1991, 1993, 1998) published a sequence of papers analyzing the behavior of $G_{n,p}$ for $p$ in the critical window, and in particular strengthening Bollobás's results. He showed showing (a) and (b), above, hold as long as $h(n) = \omega(n^{-4/3})$ (so it makes sense to speak of "the" giant component for all $p = 1/n + \omega(n^{-4/3})$).

Łuczak also proved a strong "symmetry principle" for the lower and upper parts of the critical window. Roughly stated, his result is that for $h(n) = \omega(n^{-4/3})$, if $p = 1/n + h(n)$ and $p^* = 1/n - h(n)$, then the structure of $G_{n,p^*}$ is "the same as" the structure of $G_{n,p}$ with the giant component $H_{n,p}$ removed (we denote this graph $G_{n,p} - H_{n,p}$), in that if any graph property $P$ holds a.a.s. for for $G_{n,p} - H_{n,p}$, then it holds a.a.s. for $G_{n,p^*}$. For $p^* < 1/n - \omega(n^{-4/3})$ as above, he also proved explicit upper tail bounds on the size of the largest components and and on the greatest distance between any two vertices in the same connected component of $G_{n,p^*}$ (i.e. on the *diameter* of $G_{n,p^*}$).

Another topic of considerable interest to researchers has been the *excess* of the giant component, defined as the difference between the number of edges and the number of vertices.

A connected graph with excess $-1$, for example, is a tree; a connected graph with excess 0 contains exactly one cycle (and is called a *unicyclic* graph); graphs with excess at least one are called "complex". Janson et al. (1993) studied the size and excess of $G_{n,p}$ in the critical window in detail, in particular deriving asymptotic bounds on the excess of the giant and the probability that there is ever a complex component aside from the giant; similar results for random multigraphs are also derived. Related work also appears in Łuczak et al. (1994).

Aldous (1997) has defined a stochastic process related to a procedure for *growing* $G_{n,p}$ and shown that this process has a Brownian motion as a weak limit. Based on this fact, he is able to derive the asymptotic joint distribution of the sizes and excesses of the $k$ largest components of $G_{n,p}$ for $p$ in the critical window. (We will discuss this in a little more detail later, when we use a related process in our own analysis.) Further results on the size of the $H_{n,p}$ during the critical phase appear in the recent paper of Pittel (2001), which proves a central limit theorem for the random variable measuring the size of $H_{n,p}$ and asymptotics for the tails of this random variable, together with an interesting account of known results.

We study $H_{n,p}$ by analyzing a branching process for growing $G_{n,p}$ and a random walk which can be associated to this branching process. Using this approach, we are able to prove fairly strong upper and lower tail bounds on the size and excess of $H_{n,p}$. (Independently, Kim (2006) has recently demonstrated an elegant use of a graph process with Poisson-distributed vertex degrees to arrive at similar bounds to ours for the size of $H_{n,p}$; his approach can also be used to treat the excess of $H_{n,p}$, though he does not do so explicitly.) We now proceed to the details of our argument.

## 3.2 Understanding $G_{n,p}$ through breadth-first search

We analyze the component structure of $G_{n,p}$ using a process similar to breadth-first search (BFS) (Cormen et al., 2001) and to a process used by Aldous (1997) to study random graphs in the critical window from a weak limit point of view. We highlight that $G_{n,p}$ is a labeled random graph model with vertex set $\{v_1, v_2, \ldots, v_n\}$. For $i \geq 0$, we define the set $\mathcal{O}_i$ of open vertices at time $i$, and the set $A_i$ of the vertices that have already been explored at time $i$. We set $\mathcal{O}_0 = v_1$, $A_0 = \emptyset$, and construct $G_{n,p}$ as follows:

Step $i$ $(0 \leq i \leq n-1)$: Let $v$ be an arbitrary vertex of $\mathcal{O}_i$ and let $N_i$ be the random set of neighbours of $v$. Set $\mathcal{O}_{i+1} = \mathcal{O}_i \cup N_i - \{v\}$ and $A_{i+1} = A_i \cup \{v\}$. If $\mathcal{O}_{i+1} = \emptyset$, then reset $\mathcal{O}_{i+1} = \{u\}$, where $u$ is the element of $\{v_1, v_2, \ldots, v_n\} - A_i$ with the smallest index.

Each time $\mathcal{O}_{i+1} = \emptyset$ during some Step $i$, then a component of $G_{n,p}$ has been created. To get a handle on this process, we now further examine what may happen during Step $i$. The number of neighbours of $v$ not in $A_i \cup \mathcal{O}_i$ is distributed as a binomial random variable $\mathrm{Bin}(n - i - |\mathcal{O}_i|, p)$. By the properties of $G_{n,p}$, the distribution of edges from $v$ to $V - A_i$ is independent of what happens in the previous steps of the process. Furthermore, if $\mathcal{O}_{i+1} = \emptyset$ does not occur during Step $i$, then $w \in \mathcal{O}_{i+1} - \mathcal{O}_i$ precisely if $w \notin A_i \cup \mathcal{O}_i$ and we expose an edge from $v$ to $w$ during this step. It follows that $|\mathcal{O}_{i+1}|$ is distributed as $\max(|\mathcal{O}_i| + \mathrm{Bin}(n - i - |\mathcal{O}_i|, p) - 1, 1)$. An advantage of this method of construction is that if $\mathcal{O}_{t+1} = \emptyset$ during Step $t$, instead of thinking of the process continuing to construct $G_{n,p}$ we may think of *restarting* the process to construct $G_{n-t,p}$.

We can thus analyze the growth of the components of $G_{n,p}$, created by the above BFS-based process, by coupling the process to the following random walk. Let $S_0 = 1$. For $i \geq 0$, let

$X_{i+1} = \text{Bin}(n - i - S_i, p) - 1$, and let

$$S_{i+1} = \max(S_i + X_{i+1}, 1).$$

With this definition, for all $i < n - 1$, $S_i$ is precisely $|\mathcal{O}_i|$, and any time $S_{i-1} + X_i = 0$, a component of $G_{n,p}$ has been created. We will sometimes refer to such an event as $\{S_i = 0\}$ or say that "$S$ visits zero at time $i$".

An analysis of the height of the random walk $S$ and its concentration around its expected value will form a crucial part of our derivation of bounds for the size of $H_{n,p}$. We will prove matching upper and lower bounds that more-or-less tie down the behavior of the random variable $S_i$ for $i$ in a certain key range, and thereby imply bounds on the sizes of the components of $G_{n,p}$. In analyzing this random walk, we find it convenient to use the following related, but simpler processes:

- $S'$ is the walk with $S'_0 = 1$ and $S'_{i+1} = S'_i + X'_{i+1}$, where $X'_{i+1} = \text{Bin}(n - i - |\mathcal{O}_i|, p) - 1$, for $i \geq 0$. This walk behaves like $S_i$ but is allowed to take non-positive values.
- $S^*$ is the walk with $S^*_0 = 1$ and $S^*_{i+1} = S^*_i + \text{Bin}(n, p) - 1$.
- $S^{ind}$ is the walk with $S^{ind}_0 = 1$ and $S^{ind}_{i+1} = S^{ind}_i + \text{Bin}(n - (i + 1), p) - 1$, for $i \geq 0$.
- $S^h$ is the walk with $S^h_0 = 1$ and $S^h_{i+1} = S^h_i + \text{Bin}(n - (i + 1) - h, p) - 1$, for $i \geq 0$.

Note that *all* of these walks are allowed to go negative. *We couple all the above walks to $S$.* $S'$ is coupled to $S$ by its definition. We couple $S^{ind}$ to $S$ by letting, for each $i$, $S^{ind}_{i+1} = S^{ind}_i + X_{i+1} + \text{Bin}(S_i, p)$, and couple $S^*$ to $S$ in a similar fashion. We couple $S^h$ to $S$ by for each $i$ considering the random variable $\text{Bin}(n - (i + 1) - h, p)$ as a sum of either a subset or a superset of the random variables comprising the sum $X_i = \text{Bin}(n - i - S_i, p)$, according to whether $S_i \leq h + 1$ or $S_i > h + 1$, respectively. We emphasize that until the first visit of $S$ to 0, $S'$ agrees with $S$ while $S^{ind}$ and $S^*$ strictly dominate it. Finally, $S$ dominates $S^h$ until the first time that $S$ exceeds $h+1$. We will rely on the properties of these simpler walks when analyzing $S$.

As a preliminary exercise, consider what happens to the random walk $S$ which corresponds to the random graph process $G_{n,1/n}$. It is known (Erdős and Rényi, 1960) that at this point, the largest component has size $\Theta(n^{2/3})$. We provide a simple proof of this fact using the random walks defined above and the ballot theorem approach of the previous chapter. Denote by $\mathcal{C}(v)$ the component of $G_{n,p}$ containing $v$ – we focus on the size of the component $\mathcal{C}(v_1)$ containing $v_1$.

Note that $|\mathcal{C}(v_1)| > t$ precisely if $S_i \neq 0$ for all $i \leq t$. The probability of this event is bounded above by the probability that the random walk $S^*$ has not returned to zero by time $t$. The steps of $S^*$ are iid random variables distributed as $\text{Bin}(n, 1/n) - 1$. Furthermore, the size of the component containing $v_1$ is deterministically at most $n$. It follows that $\mathbf{P}\{|\mathcal{C}(v_1)| \geq t\}$ is at most the probability that the first return to 0 of $S^*$ is between times $t$ and $n$.

By Theorem 9 (or in this case, since $S_k^* + k$ is a binomial, by explicit computation), the probability that $S_k^* = 0$ is $O(1/\sqrt{k})$. By the standard rotation argument, it follows that the probability that $S^*$ *first* returns to zero at time $k$ is $O(1/k^{3/2})$. Summing over $t \leq k \leq n$, it follows that the probability $S^*$ first returns to zero between times $t$ and $n$ is $O(1/t^{1/2})$. By the comments of the previous paragraph, it follows that the probability $v_1$ is in a component of size at least $t$ is $O(1/t^{1/2})$. Let $N_t$ be the number of vertices in components of size at least $t$. By linearity of expectation, it follows from the above fact that $\mathbf{E}N_t = O(n/t^{1/2})$. If there are *any* vertices in components of size $t$, there are at least $t$ such vertices, so $\mathbf{E}\{N_t \mid N_t \geq 1\} \geq t$. Since

$$\mathbf{E}N_t = \mathbf{E}\{N_t \mid N_t \geq 1\}\mathbf{P}\{N_t \geq 1\} + 0 \geq t\mathbf{P}\{N_t \geq 1\},$$

our bound on $\mathbf{E}N_t$ yields that there is $c$ such that $\mathbf{P}\{N_t \geq 1\} \leq t/\mathbf{E}N_t \leq cn/t^{3/2}$, which immediately yields that probability there is a component of size much larger than $n^{2/3}$ is small. Letting $L$ be the size of the largest component, we thus have

$$\mathbf{E}L \leq \sum_{i=0}^{n} \mathbf{P}\{L \geq i\} \leq n^{2/3} + \sum_{t=n^{2/3}}^{n} \mathbf{P}\{L \geq t\} \leq n^{2/3} + \sum_{t=n^{2/3}}^{n} \frac{cn}{t^{3/2}} \leq n^{2/3} + \frac{2cn}{n^{1/3}} = (2c+1)n^{2/3},$$

i.e., the expected size of the largest component is $O(n^{2/3})$.

This technique does not yield a corresponding lower bound due to the fact that the expected change in $S$ in a step becomes increasingly negative as the walk continues. As an aside, however, we note that we may use an essentially identical approach to study a critical (mean one) branching process. We can then easily derive the precise probability that such a process has size exactly $t$ in terms of the probability that $S_t = 0$. This probability was first calculated by Dwass (1969); in the same paper he pointed out the link with a random walk such as ours. Though his proof did not proceed via the ballot theorem, he used his result to *prove* a ballot theorem for such random walks.

We can use a similar approach to bound the size of the largest component for $p = 1/n - f/n^{4/3}$ and $f > 0$, by additionally using Chernoff's tail bounds for the binomial distribution:

**Theorem 37** (Chernoff, 1952). *If $Y = \mathrm{Bin}(m, q)$, then denoting $\mathbf{E}Y$ (which is $mq$) by $\lambda$, we have:*

$$\mathbf{P}\{Y - \mathbf{E}Y > r\} \leq e^{-r^2/2(\lambda + r/3)}, \tag{3.1}$$

*and*

$$\mathbf{P}\{Y - \mathbf{E}Y < -r\} \leq e^{-r^2/2\lambda}. \tag{3.2}$$

When $p = 1/n - f/n^{4/3}$ and $0 < f = o(n^{1/3})$, the probability that $|\mathcal{C}(v_1)| \geq t$ is again at most the probability that $S^*$ first returns to zero between times $t$ and $n$. Since $\mathbf{E}S_t = -fk/n^{1/3}$, by Chernoff's bounds $\mathbf{P}\{S_t \geq 0\} \leq e^{-3fk/8n^{2/3}}$. In fact, it is an easy computation from the properties of the binomial distribution to show that $\mathbf{P}\{S_k = 0\} = O(\mathbf{P}\{S_k \geq 0\}/\sqrt{k}) = O(e^{-3fk/8n^{2/3}}/\sqrt{k})$. By the standard rotation argument, the probability that $S^*$ *first* returns to zero at time $k$ is at most $2\mathbf{P}\{S_t = 0\}/k = O(e^{-3fk/8n^{2/3}}/k^{3/2})$.

Summing over $t \leq k \leq n$ yields that $\mathbf{P}\{|\mathcal{C}(v_1)| \geq t\} = O(e^{-3ft/8n^{2/3}}/\sqrt{t})$, so by linearity of expectation $\mathbf{E}|N_t| = O(ne^{-3ft/8n^{2/3}}/\sqrt{t})$. From here we can argue just as in the case $p = 1/n$ to obtain that the expected size of the largest component of $G_{n,1/n-f/n^{4/3}}$ is at most $n^{2/3}/f$.

We note that Nachmias and Peres (2005) have used an analysis similar to, but more powerful than that sketched above to derive stronger bounds than those shown above for the size of the largest component of $G_{n,1/n}$. Their proof uses essentially the same BFS-based exploration process for $G_{n,1/n}$, but takes advantage of the fact that the random walk $S^*$ is in fact a martingale. By applying the optional stopping theorem for martingales to the event that $S^*$ first returns to zero at time $t$, they obtain the bound $\mathbf{P}\{|\mathcal{C}(v_1)| \geq t\} \leq 1/t$ where we obtained $\mathbf{P}\{|\mathcal{C}(v_1)| \geq t\} = O(t^{-1/2})$. This strengthening allows them to prove that the probability that the largest component has size at least $Cn^{2/3}$ is at most $3/C^2$. Additionally, their martingale analysis allows them to prove explicit bounds on the probability that this component is much *smaller* than $n^{2/3}$; such bounds do not follow straightforwardly using the above approach. (We note that in the course of their proof, Nachmias and Peres prove analogs of Theorem 19 and Lemma 24 for the martingale $S^*$. For the case they consider, their results are stronger than ours as they are able to take advantage of the fact that the steps of $S^*$ are binomial.) Furthermore, aspects of their analysis can be used to bound the sizes of components of $G_{n,p}$ for $p$ in the critical window (Peres, 2006).

In the upper part of the critical window, the above approach fails. If $p = 1/n + f/n^{4/3}$ and $f > 0$ then at the outset the walk $S$ has positive drift (the steps have positive mean), so we expect the walk to move into the positive values. As the walk continues, at some point the steps have roughly zero mean, then negative mean, and finally we reach a time $t$ where we expect the total drift to be zero (a back-of-the-envelope approximation pretending that $S$ behaves exactly as $S^{ind}$ suggests that this $t$ should be about $2fn^{2/3}$). With a reasonably high probability, the first time $t$ we return to zero is indeed close to $2fn^{2/3}$. But the probability that we stay above 0 before time $t$ given that $S$ visits zero at time $t$ is *not* $O(1/t)$ – the the nature of the drift makes it much more likely that such an event occurs than it would be were the steps identically distributed. For this reason, we can not straightforwardly apply the ballot theorem argument to derive an upper bound on the size of the giant component in the upper portion of the critical window.

Instead, we use one of the tools that was crucial in the *proofs* of the ballot theorems of Section 2.4 to develop tools for showing that with high probability, the random walk $S$ is *never* far from its expected value. We introduce a degree of independence to the problem via the random walks $S^{ind}$ and $S^h$. Intuitively speaking, these walks provide "upper bounds" and "lower bounds" on the value of $S$, and have the advantage that we can control them directly using Chernoff bounds due to the independence of their steps. This will allow us to show that with high probability, a large component of size about $2fn^{2/3}$ appears in the first $(2+\epsilon)fn^{2/3}$ steps of the random walk.

Once this large component has appeared, in the portion of the random walk that remains the drift *is* negative; we can therefore use the ballot theorem-based technique above to bound the sizes of the remaining components of $G_{n,p}$ and thereby show that with high probability, this large component is indeed the giant component. (The behavior of the remainder of the random walk can be seen as confirmation of the fact mentioned in the introduction that the structure of $G_{n,p} - H_{n,p}$ is much like that of a subcritical random graph. As a matter of fact, using existing knowledge about the behavior of subcritical random graphs will turn out to yield better probability bounds than using the ballot theorem.)

The preceding argument and remarks lend credence to two claims: (1) that the largest component of $G_{n,1/n+f/n^{2/3}}$ has size $O(fn^{2/3})$, and (2) that any component of this size must arise early in the branching process. The main goal of the rest of this chapter is to state and prove precise versions of these claims To do so, we need to tie down the behavior of $S$. First, however, we analyze $S^{ind}$,$S^h$ and $S'$, as they buck a little less wildly.

## 3.3 The height of the tamer walks

We can handle $S^{ind}$ for $p = 1/n + \delta$ using the analysis discussed above, which consists of little more than standard results for the binomial distribution. Specifically, we have that for $t \geq 1$, $S_t^{ind} + (t-1)$ is distributed like $\text{Bin}(nt - \binom{t+1}{2}, p)$, so by linearity of expectation, we have:

**Fact 38.** *For $p = 1/n + \delta$ with $\delta < 1/n$,*

$$\mathbf{E}S_t^{ind} = \delta nt - \frac{t(t+1)}{2n} - \frac{t(t+1)\delta}{2} + 1 \leq t + 1.$$

Using the fact that the variance of a $\text{Bin}(m,p)$ random variable is $m(p-p^2)$, we have:

**Fact 39.** *For $p = 1/n + \delta$ with $\delta = o(1/n)$ and $t = o(n)$,*

$$\mathbf{Var}\left\{S_t^{ind}\right\} = \mathbf{Var}\left\{S_t^{ind} + (t-1)\right\} = (1+o(1))t$$

Intuitively, $S_t^{ind}$ has a good chance of being negative if the variance exceeds the square of the expectation and a tiny chance of being negative if the expectation is positive and dwarfs the square root of the variance. Indeed, we can formalize this intuition using the Chernoff (1952) bounding method.

We are interested in the critical range, $p = 1/n + \delta$ for $\delta = o(1/n)$. For such $\delta$, $t(t+1)\delta/2$ is $o(t(t+1)/2n)$, so we see that $\mathbf{E}S_t^{ind}$ goes negative when $\delta nt \simeq t(t+1)/2n$, i.e., when $t \simeq 2\delta n^2$. Furthermore, for any $\alpha \in (0,1)$, there exist $a_1 = a_1(\alpha) > 0$ and $a_2 = a_2(\alpha) > 0$ such that $\mathbf{E}S_t^{ind}$ is sandwiched between $a_1\delta nt$ and $a_2\delta nt$, for $\alpha\delta n^2 \le t \le (2-\alpha)\delta n^2$. As a consequence, $(\mathbf{E}S_t^{ind})^2 = \Theta(\delta^2 n^2 t^2) = \Theta(\delta^3 n^4 t)$ for such $p$ and $t$.

Also, Fact 39 states that $\mathbf{Var}\left\{S_t^{ind}\right\} = (1+o(1))t$, so the square of the expectation dwarfs the variance in this range provided $\delta^3 n^4$ is much greater than 1, i.e., provided $\delta$ is much greater than $1/n^{4/3}$.

Writing $\delta = f/n^{4/3} = f(n)/n^{4/3}$, we will focus on the case where $f > 1$ and $f = o(n^{1/3})$. We assume for the remainder of Section 3.2, and in particular as a hypothesis in all lemmas and theorems of this section, that $p = 1/n + f/n^{4/3}$ and that $f$ satisfies these constraints. In the lemma that follows we use Chernoff bounds to show that $S_t^{ind}$ is close to its expected value for all such $f$.

**Lemma 40.** *For all* $1 \le t \le n-1$ *and* $0 \le x \le t$,

$$\mathbf{P}\left\{\left|S_t^{ind} - \mathbf{E}S_t^{ind}\right| > x\right\} \le 2e^{-x^2/5t}.$$

*Furthermore, for any* $1 \le i < j \le t$,

$$\mathbf{P}\left\{\left|(S_j^{ind} - S_i^{ind}) - \mathbf{E}\left\{S_j^{ind} - S_i^{ind}\right\}\right| > x\right\} \le 2e^{-x^2/5t}.$$

*Proof.* The tail bound on $S_j^{ind} - S_i^{ind}$ follows by applying Theorem 37 to $(S_j^{ind} - S_i^{ind}) + (j - i)$, which is a binomial random variable. Before applying it, we observe that by Fact 38, $\mathbf{E}\left\{S_j^{ind} - S_i^{ind} + (j-i)\right\} \le 2j \le 2t$. Thus

$$\mathbf{P}\left\{\left|(S_j^{ind} - S_i^{ind}) - \mathbf{E}\left\{S_j^{ind} - S_i^{ind}\right\}\right| > x\right\} \le 2e^{-x^2/(4t+2x/3)} \le 2e^{-x^2/5t},$$

which establishes the latter claim. The former is obtained by applying an identical argument to $S_t + (t-1)$, which is also a binomial random variable with mean at most $2t$. $\square$

We turn now to $S^h$, which is also easier to handle than $S$.

**Lemma 41.** *For all* $1 \leq t \leq n-1$,

$$\mathbf{E}S_t^h = \frac{tf}{n^{1/3}} - \frac{t(t+1+2h)}{2n} - \frac{t(t+1+2h)f}{2n^{4/3}} + 1.$$

*Furthermore, for all integers* $0 \leq i < j \leq t$ *and for all* $0 \leq x \leq t$,

$$\mathbf{P}\left\{\left|(S_j^h - S_i^h) - \mathbf{E}\left\{S_j^h - S_i^h\right\}\right| > x\right\} \leq 2e^{-x^2/5t}.$$

We omit the proof of this lemma as it is established just as Fact 38 and Lemma 40. The above lemmas yield tail bounds on the value of some of the random walks associated with $S$ at some *specific* time $t$. These bounds rather straightforwardly yield bounds on the probability that $S$ is far from its expected value at *any* time up to some fixed time $t$:

**Lemma 42.** *Fix* $1 \leq t \leq n-1$ *and* $1 \leq x \leq t$. *Then*

$$\mathbf{P}\left\{|S_i^{ind} - \mathbf{E}S_i^{ind}| \geq x \text{ for some } 1 \leq i \leq t\right\} \leq 4e^{-x^2/20t}.$$

*Furthermore, an identical bound holds for* $S^h$, *for any* $h$ *for which* $t + h \leq n$.

*Proof.* Let $A$ be the event that there is $i \leq t$ for which $|S_i^{ind} - \mathbf{E}S_i^{ind}| \geq x$ - we aim to show bounds on $\mathbf{P}\{A\}$. We consider the first time $i^*$ at which $|S_{i^*}^{ind} - \mathbf{E}S_{i^*}^{ind}| \geq x$ (or $i^* = t+1$ if this never occurs).

For $i \leq t$, let $A_i$ be the event that $i^* = i$ and let $B_i$ be the event that $A_i$ occurs and $|S_t^{ind} - \mathbf{E}S_t^{ind}| \leq x/2$. Finally, let $B$ be the event that $|S_t^{ind} - \mathbf{E}S_t^{ind}| > x/2$. If $A$ occurs then either one of the events $B_i$ occurs or $B$ occurs. Furthermore, if $A_i$ occurs then for $B_i$ to occur it must be the case that $|(S_t^{ind} - S_i^{ind}) - (\mathbf{E}S_t^{ind} - \mathbf{E}S_i^{ind})| \geq x/2$. As $i^*$ is a stopping time, it follows by the strong Markov property that for any $i \leq t$,

$$\mathbf{P}\{B_i|A_i\} \leq \mathbf{P}\left\{|(S_t^{ind} - S_i^{ind}) - (\mathbf{E}S_t^{ind} - \mathbf{E}S_i^{ind})| \geq x/2\right\}.$$

Furthermore, the $A_i$ are disjoint so the $\mathbf{P}\{A_i\}$ sum to at most 1. It follows that

$$\begin{aligned}
\mathbf{P}\{A\} &\leq \mathbf{P}\{B\} + \sum_{i=1}^{t} \mathbf{P}\{B_i\} = \mathbf{P}\{B\} + \sum_{i=1}^{t} \mathbf{P}\{B_i|A_i\}\mathbf{P}\{A_i\} \\
&\leq \mathbf{P}\{B\} + \max_{1\leq i\leq t} \mathbf{P}\left\{\left|(S_t^{ind} - S_i^{ind}) - \mathbf{E}\left\{S_t^{ind} - S_i^{ind}\right\}\right| \geq x/2\right\} \\
&\leq 2\max_{1\leq i\leq t} \mathbf{P}\left\{\left|(S_t^{ind} - S_i^{ind}) - \mathbf{E}\left\{S_t^{ind} - S_i^{ind}\right\}\right| \geq x/2\right\} \\
&\leq 4e^{-x^2/20t},
\end{aligned}$$

by applying Lemma 40. An identical bound holds for $S^h$ by mimicking the above argument but applying Lemma 41 at the last step. □

## 3.4 The height of $S$

We now turn to the walk we are really interested in. For all $i$ it is deterministically the case that $S_i \leq S_i^{ind} + i$, so we may use Lemma 42 to bound $S_i$ (equivalently, $|\mathcal{O}_i|$) for $1 \leq i \leq t \leq n-1$. Letting $x = \epsilon t$ in Lemma 42 yields:

**Corollary 43.** *Fix $1 \leq t \leq n-1$ and $0 < \epsilon \leq 1$. Then*

$$\mathbf{P}\{|\mathcal{O}_i| \geq (1+\epsilon)t \text{ for some } 1 \leq i \leq t\} \leq 4e^{-\epsilon^2 t/20}.$$

The above crude bound on $|\mathcal{O}_i|$ is a result of bounds on $S^{ind}$ and the fact that $S_i - S_i^{ind} \leq i$.

To get more precise information on the height of $S_i$, we need to improve our bound on $S_i - S_i^{ind}$. To this end, we note that letting $Z_t$ be the number of times that $S_i$ hits zero up to time $t$, we have $S_t = S_t' + Z_t$. Since $S_t'$ hits a new minimum each time $S_t$ hits zero, $Z_t = -\min\{S_i' - 1 | 1 \leq i \leq t\}$. Since $S^{ind}$ strictly dominates $S'$, we thus have $S_i \leq S_i^{ind} + Z_i$ for all $1 \leq i \leq n-1$, which will turn out to yield considerably better bounds than $S_i \leq S_i^{ind} + i$ once we have obtained bounds on $Z_i$. Such bounds follow from the following lemma:

**Lemma 44.** *For all* $1 \leq t \leq n-1$

$$\mathbf{P}\left\{S_i' \leq \frac{if}{n^{1/3}} - \frac{2t^2}{n} \textit{ for some } 1 \leq i \leq t\right\} \leq 8e^{-t^3/400n^2}.$$

*Proof of Lemma 44.* By Corollary 43, the probability $|\mathcal{O}_i| \geq (5/4)t$ for some $1 \leq i \leq t$ is at most $4e^{-t/320}$. On the other hand, as long as $|\mathcal{O}_i| \leq 5t/4$ for all $1 \leq i \leq t$, $S_{i+1}' - S_i' \geq \text{Bin}(n-i-5t/4, p) - 1$, so $S_i' \geq S_i^{5t/4}$ for all $1 \leq i \leq t$. Furthermore, it follows from Lemma 41 and the fact that $f = o(n^{1/3})$ that for any $\epsilon > 0$, for $n$ large enough, for all $1 \leq i \leq t$,

$$\mathbf{E}S_i^{5t/4} \geq \frac{if}{n^{1/3}} - \left(1 + \frac{f}{n^{1/3}}\right)\frac{i(i+1+5t/2)}{2n} \geq \frac{if}{n^{1/3}} - \frac{(7/4+\epsilon)t^2}{n}.$$

Thus, if $S_i' \leq if/n^{1/3} - 2t^2/n$ for some $1 \leq i \leq t$, then either

(a) $|\mathcal{O}_j| \geq 5t/4$ for some $1 \leq j \leq t$, or

(b) $\mathbf{E}S_i^{5t/4} - (1/4 - \epsilon)t^2/n \geq S_i' \geq S_i^{5t/4}$.

We have already seen that (a) occurs with probability at most $4e^{-t/320} \leq 4e^{-t^3/400n^2}$. By choosing $\epsilon = 1/40$, say, and applying Lemma 42 with $x = (1/4 - \epsilon)t^2/n = 9t^2/40n$, it follows that for $n$ large enough (b) occurs for some $1 \leq i \leq t$ with probability at most $4e^{-(9/40)^2t^3/20n^2} \leq 4e^{-t^3/400n^2}$. The lemma follows. □

It is immediate that

**Corollary 45.** *For $n$ large enough, the probability that $Z_t > 2t^2/n$ is at most $8e^{-t^3/100n^2}$.*

We are now able to derive much stronger upper tail bounds on $S_i$:

**Theorem 46.** *For $n$ large enough, the probability that $S_i > 20f^2n^{1/3}$ for some $1 \leq i \leq 3fn^{2/3}$ is at most $12e^{-f^3/60}$.*

*Proof.* Let $t = 3fn^{2/3}$. If $S_i \geq 20f^2n^{1/3}$ for some $1 \leq i \leq t$ then either $Z_t \geq Z_i \geq 18f^2n^{1/3}$ or $S_i^{ind} \geq 2f^2n^{1/3}$. Corollary 45 yields that the former event has probability at most $8e^{-f^3/15}$. Furthermore, using Fact 38 it is straightforward to see that $\mathbf{E}S_i^{ind} \leq f^2n^{1/3}$ for all $i$, so applying Lemma 42 with $x = f^2n^{1/3}$ yields that the probability $S_i^{ind}$ is more than $2f^2n^{1/3}$ for some $1 \leq i \leq t$ is at most $4e^{-x^2/20t} = 4e^{-f^3/60}$. Combining these bounds yields the result. □

## 3.5 Growing a giant component

Using these bounds on the height of $S$, we are able to determine the structure of the giant component of $G_{n,p}$ for $p$ in the range we are focussing on. Recall that the excess of a connected graph $H$ is equal to $|E(H)| - |V(H)|$. In this section we prove:

**Theorem 47.** *For all $0 < c < 1$ there is $F > 1$ such that for $f > F$ and $n$ large enough, with probability at least $1 - 85e^{-c^4f^3/2^6 \cdot 100}$, the random graph $G_{n,p}$ contains a component $H$ of size between $(2-c)fn^{2/3}$ and $(2+c)fn^{2/3}$, and excess between $f^3/20$ and $150f^3$.*

We note that though this theorem is stated for $c < 1$, we could derive a version that held for larger $c$ by straightforwardly extending Theorem 46 to apply to values of $t$ other than $3fn^{2/3}$. We prove this theorem in two steps. We first bound the probability that we obtain a component of the desired size, and then bound the excess of this component.

### 3.5.1 The size of the giant component

To begin, we strengthen the argument used in Lemma 44 by using the stronger bound on the height of $S$ given by Theorem 46, to show:

**Theorem 48.** *Fix $0 < \alpha \leq 1$. Then for $n$ large enough, the probability that $S_i = 0$ for some $\alpha f n^{2/3} \leq i \leq (2-\alpha) f n^{2/3}$ is at most $13e^{-\alpha^4 f^3/200}$.*

*Proof.* As $S_i \geq S_i'$ for all $i$, it suffices to prove that the probability $S_i' \leq 0$ for some such $i$ is at most $13e^{-\alpha^4 f^3/200}$. Letting $h = 20 f^2 n^{1/3}$, we have that $S'$ is at least $S^h$ until the first time $i$ that $S_i \geq h$. From Lemma 41,

$$\mathbf{E}S_i^h \;=\; \mathbf{E}S_i^{ind} - \left(1 + \frac{f}{n^{1/3}}\right)\frac{hi}{n} \geq \mathbf{E}S_i^{ind} - \frac{40 f^2 i}{n^{2/3}},$$

for $n$ large enough. For $\alpha f n^{2/3} \leq i \leq (2-\alpha) f n^{2/3}$ and $n$ large enough, it follows from Fact 4 that $\mathbf{E}S_i^{ind} \geq (\alpha^2/2) f^2 n^{1/3}$, so

$$\mathbf{E}S_i^h \geq \frac{\alpha^2 f^2 n^{1/3}}{2} - \frac{40 f^2 i}{n^{2/3}}.$$

Furthermore, since $f = o(n^{1/3})$, for $n$ large enough and $i \leq (2-\alpha) f n^{2/3}$ we have $i/n^{2/3} \leq (2-\alpha) f \leq \alpha^2 n^{1/3}/800$, so

$$\mathbf{E}S_i^h \;\geq\; \frac{\alpha^2 f^2 n^{1/3}}{2} - \frac{40\alpha^2 f^2 n^{1/3}}{800} = \frac{9\alpha^2 f^2 n^{1/3}}{20}.$$

Therefore, if $S_i' \leq 0$ for some $\alpha f n^{2/3} \leq i \leq (2-\alpha) f n^{2/3}$, either $S_j \geq h$ for some $j \leq i$ or $S_i^h \leq \mathbf{E}S_i^h - 9\alpha^2 f^2 n^{1/3}/20$. By Theorem 46, the former event has probability at most $12e^{-f^3/60} < 12e^{-\alpha^4 f^3/200}$. Letting $t = (2-\alpha) f n^{2/3}$ and $x = 9\alpha^2 f^2 n^{1/3}/20$ and applying

Lemma 42 yields that the latter event has probability at most

$$4e^{-x^2/20t} \leq 4e^{-81\alpha^4 f^4 n^{2/3}/8000(2-\alpha)fn^{2/3}} \leq e^{-\alpha^4 f^3/200}.$$

This completes the proof. □

**Corollary 49.** *For $n$ large enough, the probability that $S_i \leq f^2 n^{1/3}/10$ for some $fn^{2/3}/2 \leq i \leq 3fn^{2/3}/2$ is at most $16e^{-f^3/3000}$.*

*Proof.* We let $\alpha = 1/2$ and follow an identical chain of reasoning from the fact that $S_i \geq S_i'$ for all $i$. If $S_i' \leq f^2 n^{1/3}/10$ for some $fn^{2/3}/2 \leq i \leq 3fn^{2/3}/2$, then for some $j \leq i$, either $S_j \geq h = 20f^2 n^{1/3}$ or $S_i^h \leq \mathbf{E}S_i^h - f^2 n^{1/3}/10$. By Theorem 46, the former event has probability at most $12e^{-f^3/605}$. By Lemma 42 the latter event has probability at most $4e^{-f^3/3000}$. □

Theorem 48 tells us that with high probability, $S$ does not visit zero between times $\alpha fn^{2/3}$ and $(2-\alpha)fn^{2/3}$. Furthermore, $S_i \leq S_i^{ind} + Z_{\alpha fn^{2/3}}$ until the first time after $\alpha fn^{2/3}$ that $S$ visits zero. Combining this fact with our tail bounds on $S^{ind}$ and $Z_{\alpha fn^{2/3}}$, we can show that $S$ very likely *does* visit zero around time $2fn^{2/3}$:

**Theorem 50.** *Fix $0 < \alpha \leq 1$. Then for $n$ large enough, the probability that $S$ does not visit zero between time $(2-\alpha)fn^{2/3}$ and $(2+2\alpha)fn^{2/3}$ is at most $23e^{-\alpha^4 f^3/100}$.*

*Proof.* For simplicity, let $\underline{t} = (2-\alpha)fn^{2/3}$, $\bar{t} = (2+2\alpha)fn^{2/3}$ and let $N$ be the event that $S$ does not visit zero between time $\underline{t}$ and $\bar{t}$. If $S_{\bar{t}}^{ind} < -Z_{\underline{t}}$ then $S_{\bar{t}}' < -Z_{\underline{t}}$, so $S$ has visited 0 between times $\underline{t}$ and $\bar{t}$. Therefore,

$$\mathbf{P}\{N\} \leq \mathbf{P}\left\{S_{\bar{t}}^{ind} \geq -Z_{\underline{t}}\right\}.$$

We bound the right hand side of this equation by writing

$$\mathbf{P}\left\{S_{\bar{t}}^{ind} \geq -Z_{\underline{t}}\right\} \leq \mathbf{P}\left\{S_{\bar{t}}^{ind} \geq -r\right\} + \mathbf{P}\left\{Z_{\underline{t}} > r\right\}, \tag{3.3}$$

and deriving bounds on the two terms of the right-hand-side of (3.3) for suitably chosen $r$. For any $r$,

$$\mathbf{P}\left\{Z_{\underline{t}} > r\right\} \leq \mathbf{P}\left\{Z_{\alpha f n^{2/3}} > r\right\} + \mathbf{P}\left\{Z_{\underline{t}} > Z_{\alpha f n^{2/3}}\right\}. \tag{3.4}$$

Since $Z_{\underline{t}} > Z_{\alpha f n^{2/3}}$ occurs precisely if $S$ visits zero between times $\alpha f n^{2/3}$ and $\underline{t}$, Theorem 48 yields that

$$\mathbf{P}\left\{Z_{\underline{t}} > Z_{\alpha f n^{2/3}}\right\} \leq 13e^{-\alpha^4 f^3/200}. \tag{3.5}$$

By its definition, $Z_{\alpha f n^{2/3}} > 2\alpha^2 f^2 n^{1/3}$ precisely if $S_i' \leq -2\alpha^2 f^2 n^{1/3}$ for some $i \leq \alpha f n^{2/3}$. Applying Lemma 44 with $t = \alpha f n^{2/3}$ thus yields that

$$\begin{aligned}\mathbf{P}\left\{Z_{\alpha f n^{2/3}} > 2\alpha^2 f^2 n^{1/3}\right\} &\leq \mathbf{P}\left\{S_i' \leq \frac{if}{n^{1/3}} - \frac{2t^2}{n} \text{ for some } 1 \leq i \leq t\right\} \\ &\leq 8e^{-t^3/400n^2} \\ &\leq 8e^{-\alpha^3 f^3/400}. \end{aligned} \tag{3.6}$$

Letting $r = 2\alpha^2 f^2 n^{1/3}$, (3.4), (3.5), and (3.6) yield

$$\begin{aligned}\mathbf{P}\left\{Z_{\underline{t}} > 2\alpha^2 f^2 n^{1/3}\right\} &\leq \mathbf{P}\left\{Z_{\underline{t}} > Z_{\alpha f n^{2/3}}\right\} + \mathbf{P}\left\{Z_{\alpha f n^{2/3}} > 2\alpha^2 f^2 n^{1/3}\right\} \\ &\leq 13e^{-\alpha^2 f^3/200} + 8e^{-\alpha^3 f^3/400} < 21e^{-\alpha^4 f^3/400}. \end{aligned} \tag{3.7}$$

Furthermore,

$$\mathbf{E}S_{\bar{t}}^{ind} \leq \frac{\bar{t}f}{n^{1/3}} - \frac{\bar{t}^2}{2n} \leq -(2\alpha + 2\alpha^2) f^2 n^{1/3},$$

so $-2\alpha^2 f^2 n^{1/3} \geq \mathbf{E}S_{\bar{t}}^{ind} + 2\alpha f^2 n^{1/3}$. By applying Lemma 40 with $x = 2\alpha f^2 n^{1/3}$, it follows that

$$\mathbf{P}\left\{S_{\bar{t}}^{ind} \geq -2\alpha^2 f^2 n^{1/3}\right\} < 2e^{-x^2/5\bar{t}} = 2e^{-4\alpha^2 f^4 n^{2/3}/5\bar{t}} \leq 2e^{-\alpha^2 f^3/5}. \tag{3.8}$$

Combining (3.3), (3.7) and (3.8) yields that

$$\mathbf{P}\left\{S_{\bar{t}}^{ind} \geq -Z_{\underline{t}}\right\} \leq 23e^{-\alpha^4 f^3/400}.$$

This completes the proof. □

### 3.5.2 The excess of the giant

Theorems 48 and 50 tell us about the size of the giant component of $G_{n,p}$. We now turn to its excess. Letting $H_p$ be the component of $G_{n,p}$ alive at time $fn^{2/3}$, we will prove

**Lemma 51.** *Let Exc be the event that $H_p$ has excess at least $f^3/20$ and at most $150f^3$. Then for $n$ large enough,*

$$\mathbf{P}\left\{\overline{Exc}\right\} \leq 49e^{-f^3/(2^6\cdot 100)}. \tag{3.9}$$

*Proof.* For simplicity in coming calculations, we define the *net excess* of a connected graph $H$ to be equal to the excess of $H$, plus 1. The net excess of components of $G_{n,p}$ can be analyzed much as we have just analyzed their size. In the process defined at the beginning of Section 3.2, each element of the random set $N_i$ of neighbours of $v_i$ that is in the set $\mathcal{O}_i$ contributes exactly 1 to the net excess of the component alive at time $i$. Thus, if a component is created between times $t_1$ and $t_2$ of the process (precisely, if $S_{t_1} - 1 = 0$ and the first time greater than $t_1 - 1$ at which $S$ visits 0 is $t_2$), then the net excess of this component is precisely $\sum_{i=t_1}^{t_2-1} \mathrm{Bin}(|\mathcal{O}_i| - 1, p) = \mathrm{Bin}(\sum_{i=t_1}^{t_2-1} S_i - 1, p)$. Our upper bound on $S$ in Theorem 46 can be thus used to prove upper bounds on the net excess of $H_p$. We split the event *Exc* into the events $Exc_1$ that $H_p$ has excess greater than $150f^3$ and $Exc_2$ that $H_p$ has excess less than $f^3/20$.

Let *Big* be the event that $H_p$ has size more than $3fn^{2/3}$, let *High* be the event that $S_i \geq 20f^2n^{1/3}$ for some $i \leq 3fn^{2/3}$. If *Big* occurs then $S$ does not return to zero between time $(3/2)fn^{2/3}$ and time $3fn^{2/3}$, so by applying Theorem 50 with $\alpha = 1/2$ yields that $\mathbf{P}\{Big\} \leq 23e^{-\alpha^4 f^3/400} = 23e^{-f^3/(2^6 \cdot 100)}$. By Theorem 46, $\mathbf{P}\{High\} \leq 8e^{-f^3/60}$. If neither *Big* nor *High* occurs, then the net excess of $H_p$ is at most $\text{Bin}(M, p)$, where $M = \sum_{i=1}^{3fn^{2/3}}(S_i - 1) \leq 60f^3n$. For *any* $m \leq 60f^3n$, $\mathbf{E}\text{Bin}(m,p) \leq 120f^3$, so by Theorem 37, it follows that $\mathbf{P}\left\{\text{Bin}(m,p) \geq 150f^3 | \overline{Big}, \overline{High}\right\} \leq e^{-f^3}$. Combining these bounds yields

$$\mathbf{P}\left\{\overline{Exc_1}\right\} \leq \mathbf{P}\{Big\} + \mathbf{P}\{High\} + e^{-f^3} \leq 32e^{-f^3/(2^6 \cdot 100)}.$$

Next let *Small* be the event that $S_i < f^2n^{1/3}/10$ for some $fn^{2/3}/2 \leq i \leq (3fn^{2/3}/2$. By Corollary 49, $\mathbf{P}\{Small\} \leq 16e^{-f^3/3000}$ If *Small* does not occur then the net excess of $H_p$ is at least $\text{Bin}(M,p)$, where $M = \sum_{i=fn^{2/3}/2}^{3fn^{2/3}/2}(S_i - 1) \geq f^3n/10$. By for any $m \geq f^3n/10$, Theorem 37 yields that $\mathbf{P}\{\text{Bin}(m,p) \leq f^3/20\} \leq e^{-f^3/1200}$. Combining these bounds yields $\mathbf{P}\{Exc_2\} \leq 17e^{-f^3/3000}$. Finally, combining our bounds on $\mathbf{P}\{Exc_1\}$ and $\mathbf{P}\{Exc_2\}$ proves the theorem. □

### 3.5.3 The proof of Theorem 47

Theorems 48 and 50, applied with $\alpha = c/2$, yield that $H_p$ has size between $(3/2)fn^{2/3}$ and $(5/2)fn^{2/3}$ with probability at least $1 - 36e^{-c^4f^3/(2^4 \cdot 100)}$. Lemma 51 shows that the excess of $H_p$ is between $f^3/20$ and $150f^3$ with probability at least $1 - 49e^{-f^3/(2^4 \cdot 100)}$. Thus, the probability both hold is at least $1 - 85e^{-c^4f^3/(2^4 \cdot 100)}$, as claimed. (We note that this establishes something slightly stronger than Theorem 47; namely, we have shown that *the component* $H_p$ has such size and excess with the desired probability.)

## 3.6 The Giant Towers Over the Others

As discussed in the introduction, the probability of growing a large component which starts in iteration $t$ of the process decreases as $t$ increases. This is what allows us to show that very likely there is a unique giant component and all the other components are much smaller.

To be precise, let $T_1$ be the first time that $S$ visits zero after time $(2-\alpha)fn^{2/3}$. Then the remainder of $G_{n,p}$ has $n' = n - T_1$ vertices and each pair of vertices is joined independently with probability $p$. If $\alpha = 1/4$, say, then

$$\begin{aligned} p &= \frac{1}{n} + \frac{(2-\alpha)f}{n^{4/3}} \leq \frac{1}{n'}\left(1 - \frac{(2-\alpha)f}{n^{1/3}}\right) + \frac{f}{n^{4/3}} \\ &< \frac{1}{n'} - \frac{(f/2)}{(n')^{4/3}}. \end{aligned} \tag{3.10}$$

Thus the final stages of the process look like a subcritical process on $n'$ vertices. We could analyze how this procedure behaves by looking at the behaviour of our random walks as in the last three subsections but instead we find it convenient to quote results of Łuczak who did obtain tail bounds for the subcritical process.

The following theorem is a reformulation of Łuczak (1990, Lemma 1) and Łuczak (1998, Theorem 11). Those results are stated for the case $f = f(n) \to \infty$, but in both cases the proof is easily adapted to our formulation; the details are omitted.

**Theorem 52.** *For all fixed $K > 1$, there exists $F > 1$ such that for all $f > F$, $n$ large enough and $p = 1/n - f/n^{4/3}$, for all $k > K$ the probability that $G_{n,p}$ contains a component of size larger than $(k + \log(f^3))n^{2/3}/f^2$ or a complex component of size larger than $2k$ is at most $3e^{-k}$. Furthermore, the probability there is a tree or unicyclic component of $G_{n,p}$ with size at most $n^{2/3}/f$ and longest path at least $12n^{1/3}\log f/\sqrt{f}$ is at most $e^{-\sqrt{f}}$.*

Using this result we easily obtain the following two theorems:

**Theorem 53.** *There is $F > 1$ such that for $f > F$ and $n$ large enough, with probability at least $1-2e^{-f^2/2}$, the component alive at time $fn^{2/3}$ is the largest component, i.e., $H_p = H_{n,p}$.*

**Theorem 54.** *For any $\epsilon > 0$ There is $F = F(\epsilon) > 1$ so that for all $f > F$ and $p = 1/n + f/n^{4/3}$, the expected number of components of $G_{n,p}$ of size exceeding $(3/2)fn^{2/3}$ is at most $1 + 2e^{-f^2/2}$.*

Theorems 53 and 54 are simple consequences of Theorem 52 and of the above BFS-based process. Let $T_1$ be the first time after $(7/4)fn^{2/3}$ that $S$ visits 0. By applying Theorem 48 with $\alpha = 1/4$, the probability that the random walk returns to zero between times $fn^{2/3}/4$ and $7fn^{2/3}/4$ is at most $13e^{-f^3/2^9 \cdot 100}$, which is at most $e^{-f^2/2}$ for $f$ large enough. If the random walk does not return to zero between these times, then $H_p$ is the largest component grown up to time $T_1$ and $|H_p| \geq 3fn^{2/3}/2$. We restart the branching process to grow the graph $G_{n-T_1,p}$. Theorem 52 guarantees that the probability a component of size exceeding $n^{2/3}$ ever occurs after time $T_1$ is at most $e^{-f^2/2}$. Combining these two bounds proves Theorem 53.

If a component of size exceeding $n^{2/3}$ does occur after time $T_1$, then once it dies we again restart the branching process to grow the remainder of the graph; again, and independently, the probability a component of size exceeding $n^{2/3}$ ever occurs is at most $e^{-f^2/2}$. Continuing in this manner yields the geometric upper bound $(e^{-f^2/2})^i$ on the probability there are precisely $i$ large components grown after time $T_1$; by making $F$ large enough we may ensure that $\sum_{i=1}^{\infty}(e^{-f^2/2})^i < 2$, which proves Theorem 54.

## 3.7 Conclusion

The results of this chapter find their place in an already rich body of theory; we pose just one question about how it might be extended. Our bounds address the "large deviations" of the size of the giant component, in the sense that we allow both the size and the excess a range of the same order as their expected value before our bounds kick in. We expect that similar bounds should be achievable for smaller deviations, and in particular conjecture that for $p = 1/n + f(n)/n^{4/3}$, $0 < f(n) \to \infty$ and $f(n) = o(n^{1/3})$, $\mathbf{P}\left\{|H_{n,p} - \mathbf{E}H_{n,p}| \geq n^{2/3}\right\} = e^{-O(f)}$. We expect that this result is achievable by being more careful with the methods of this section.

In this chapter, we connected the growth of the components of $G_{n,p}$ for $p$ in the critical window to the behavior of a random walk. In particular, we showed that the size of the giant component $H_{n,p}$ can be found by studying the the first return to zero of this random walk after a certain key time. In the next chapter, we show how the information we have derived can be used to bound the diameter of a random tree closely linked to the random graph process $G_{n,p}$.

# Chapter 4

# Minimum Weight Spanning Trees

*"A fool sees not the same tree that a wise man sees."*

William Blake, 1790

## 4.1 Introduction

Given a connected graph $G = (V, E)$, $E = \{e_1, \ldots, e_{|E|}\}$, together with edge weights $W = \{w(e) | e \in E\}$, a minimum weight spanning tree of $G$ is a spanning tree $T = (V, E')$ that minimizes

$$\sum_{e \in E'} w(e).$$

As we show below, if the edge weights are distinct then this tree is unique; in this case we denote it by $MWST(G, W)$ or simply $MWST(G)$ when $W$ is clear.

Minimum spanning trees are at the heart of many combinatorial optimization problems. In particular, they are easy to compute (Borůvka, 1926; Jarník, 1930; Kruskal, 1956; Prim, 1957), and may be used to approximate hard problems such as the minimum weight traveling salesman tour (Vazirani, 2001). (A complete account on the history of the minimum spanning tree problem may be found in the surveys of Graham and Hell (1985), and Nešetřil (1997).) As a consequence, much attention has been given to studying their structure, especially in random settings and under various models of randomness. For instance, Frieze (1985) determined the weight of a the *MWST* of a complete graph whose edges have been weighted by independent and identically distributed (i.i.d.) $[0,1]$-random variables. This result has been reproved and generalized by Frieze and McDiarmid (1989) and Aldous (1990). Under the same model, Aldous (1990) derived the degree distribution of the *MWST*. Both these results rely on local properties of minimum spanning trees. We are interested in their global structure.

The *distance* between vertices $x$ and $y$ in a graph $H$ is the length of the shortest path from $x$ to $y$. The *diameter* $diam(H)$ of a connected graph $H$ is the greatest distance between any two vertices in $H$. We are interested in the diameters of the minimum weight spanning trees of a clique $K_n$ on $n$ vertices whose edges have been assigned i.i.d. real weights. We use $w(e)$ to denote the weight of $e$. In this paper we prove the following theorem, answering a question of Frieze and McDiarmid (1997, Research Problem 23):

**Theorem 55.** *Let $K_n = (V, E)$ be the complete graph on $n$ vertices, and let $\{X_e | e \in E\}$ be independent identically distributed edge-weights. Then conditional upon the event that for all $e \neq f$, $X_e \neq X_f$, it is the case that the expected value of the diameter of* $\mathrm{MWST}(K_n)$ *is* $\Theta(n^{1/3})$.

We start with some general properties of minimum spanning trees. Let $T$ be some minimum weight spanning tree of $G$. If $e$ is not in $T$ then the path between its endpoints in $T$ consists only of edges with weight at most $w(e)$. If $e = xy$ is in $T$ then every edge $f$ between the component of $T - e$ containing $x$ and the component of $T - e$ containing $y$ has weight at least $w(e)$, since $T - e + f$ is also a spanning tree. Thus, if the edge weights are distinct, $e$ is in $T$ precisely if its endpoints are in different components of the subgraph of $G$ with edge set $\{f|w(f) < w(e)\}$. It follows that if the edge weights are distinct, $T = MWST(G)$ is unique and the following greedy algorithm (Kruskal, 1956) generates $MWST(G)$:

(1) Order $E$ as $\{e_1, \ldots, e_m\}$ so that $w(e_i) < w(e_{i+1})$ for $i = 1, 2, ..., m-1$.

(2) Let $E_T = \emptyset$, and for $i$ increasing from 1 to $m$, add edge $e_i$ to $E_T$ unless doing so would create a cycle in the graph $(V, E_T)$. The resulting graph $(V, E_T)$ is the unique $MWST$ of $G$.

Kruskal's algorithm above lies at the heart of the proof of Theorem 55. It provides a way to grow the minimum spanning tree that is perfectly suited to keeping track of the evolution of the diameter of $E_T$ as the edges are processed. We now turn our attention to this forest growing process and review its useful properties.

Observe first that, if the weights $w(e)$ are distinct, one does not need to know $\{w(e), e \in E\}$ to determine $MWST(G)$, but merely the ordering of $E$ in (1) above. If the $w(e)$ are i.i.d. random variables, then conditioning on the weights being distinct, this ordering is a random permutation. Thus, for any i.i.d. random edge weights, conditional upon all edge weights being distinct, the distribution of $MWST(G)$ is the same as that obtained by weighting $E$ according to a uniformly random permutation of $\{1, \ldots, m\}$.

This provides a natural link between Kruskal's algorithm and the $G_{n,m}$ random graph evolution process of Erdős and Rényi (1960). This well-known process, discussed in Chapter 3, consists of an increasing sequence of $|E| = \binom{n}{2}$ random subgraphs of $K_n$ defined as follows. Choose a uniformly random permutation $e_1, \ldots, e_{|E|}$ of the edges, and set $G_{n,m}$ to be the subgraph of $K_n$ with edge set $\{e_1, \ldots, e_m\}$. If we let $e_i$ have weight $i$, $1 \leq i \leq \binom{n}{2}$, then

$e_m \in MWST(K_n)$ precisely if $e_m$ is a cutedge of $G_{n,m}$. We may view Kruskal's algorithm, above, as a *restricted* random graph process – we run the usual random graph process but rather than adding *all* edges in order, we only add those that do not create cycles. (This and other restricted random graph processes have been studied by (Aldous, 1990; Ruciński and Wormald, 1992; Erdös et al., 1995; Ruciński and Wormald, 1997; Wormald, 1999).)

Using this link, the lower bound is easily obtained. It suffices to note that, with positive probability, $G_{n,n/2}$ contains a tree component $T$ whose size is between $n^{2/3}/2$ and $2n^{2/3}$ (see Janson et al. (2000), Theorem 5.20). This tree is a subtree of $MWST(K_n)$, so $diam(MWST(K_n)) \geq diam(T)$. Conditioned on its size, such a tree is a Cayley tree (uniform labeled tree), and hence has expected diameter $\Theta(n^{1/3})$ (Rényi and Szekeres, 1967; Flajolet and Odlyzko, 1982). Therefore, $\mathbf{E}\{diam(MWST(K_n))\} = \Omega(n^{1/3})$.

The upper bound is much more delicate. To obtain it, we in fact study the random graph process $G_{n,p}$ (Stepanov, 1970a,b; Janson et al., 2000; Bollobás, 2001): assign an independent $[0,1]$-uniform edge weight $w(e)$ to each edge $e$ of $K_n$, and for all $p \in [0,1]$, set $G_{n,p} = \{f | w(f) \leq p\}$. Our preference for this model over $G_{n,m}$ is due to the fact that it can be analyzed via the BFS-based process seen in Chapter 3. For this edge weighting, $e \in MWST(G)$ precisely if $e$ is a cutedge of $G_{n,w(e)}$. This implies that the vertex sets of the components of $G_{n,p}$ are precisely the vertex sets of the components of the forest $F_{n,p} = MWST(K_n) \cap \{e | w(e) \leq p\}$ built by Kruskal's algorithm. Actually it implies something stronger: $MWST(K_n) \cap \{e | w(e) \leq p\}$ consists exactly of the unique $MWST$s of the connected components of $G_{n,p}$ under the given weighting. It is this fact which allows us to determine the diameter of $MWST(K_n)$.

The results of the previous chapter guide our analysis. We shall take a snapshot of $F_{n,p}$ for an increasing sequence of $p$ and examine how this graph evolves. Due to the above connection with $G_{n,p}$, we know that for $p = 1 + \epsilon/n$, there is already a component of $F_{n,p}$

of size $\Omega(n)$ (the component corresponding to $H_{n,p}$). This fact is crucial to our analysis. Essentially, rather than looking at $F_{n,p}$, we focus on the diameter of $MWST(K_n) \cap H_{n,p}$ for $p = 1/n + \Omega(n^{-4/3})$. To track the diameter of this increasing (for inclusion) sequence of graphs, we use the following fact. For a graph $G = (V, E)$, we write $lp(G)$ for the length of the longest path of $G$. The subgraph of $G$ induced by a vertex set $U \subset V$ is denoted $G[U]$.

**Lemma 56.** *Let $G, G'$ be graphs such that $G \subset G'$. Let $H \subset H'$ be connected components of $G, G'$ respectively. Then* $\mathrm{diam}(H') \leq \mathrm{diam}(H) + 2\mathrm{lp}(G'[V - V(H)]) + 2$.

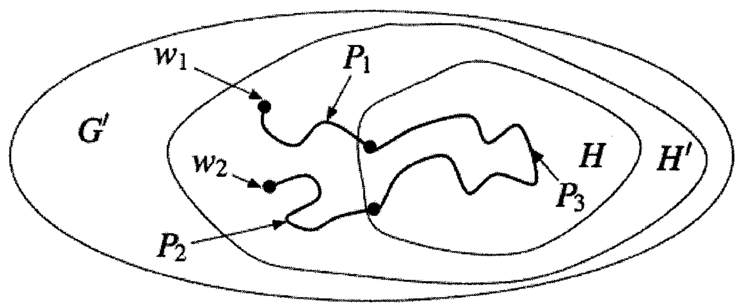


Figure 4.1: The path $P = P_1 \cup P_2 \cup P_3$ from $w_1$ to $w_2$ in $H'$.

*Proof.* For any $w_1$ and $w_2$ in $H'$, let $P_i$ be a shortest path from $w_i$ to $H$ ($i = 1, 2$), and let $P_3$ be a shortest path in $H$ joining the endpoint of $P_1$ in $H$ to the endpoint of $P_2$ in $H$. Then $P_1 \cup P_2 \cup P_3$ is a path of $H'$ from $w_1$ to $w_2$ of length at most $diam(H) + 2lp(G'[V - V(H)]) + 2$ (See Figure 4.1). □

If $p < p'$ and $H_{n,p} \subseteq H_{n,p'}$, then Lemma 56 implies that $diam(MWST(H_{n,p'}))$ is at most $diam(MWST(H_{n,p})) + 2lp(G_{n,p'}[V - V(H_{n,p})])$. We consider an increasing sequence $1/n < p_0 < p_1 < \ldots < p_t < 1$ of values of $p$ at which we take a snapshot of the random graph process. Specifically, we fix some large constant $F$, set $f_i = (5/4)^i F$, stopping at the first integer $t$ for which $f_t \geq n^{1/3}/\log n$, and choose $p_i = 1/n + f_i/n^{4/3}$ (the reason for this choice will become clear). This is similar to Łuczak's method of considering "moments" of the graph process (Łuczak, 1990). For each $p_i$, we consider the largest component $H_{n,p_i}$ of $G_{n,p_i}$. We define $dt_i$ to be the diameter of $MWST(K_n) \cap H_i$.

For $1 \le i < t$, we say $G_{n,p_i}$ is *well-behaved* if

(I) $|H_{n,p_i}| \ge (3/2)n^{2/3} f_i$ and the longest path of $H_{n,p_i}$ has length at most $f_i^4 n^{1/3}$, and

(II) the longest path of $G_{n,p_{i+1}}[V - V(H_{n,p_i})]$ has length at most $n^{1/3}/\sqrt{f_i}$

If $G_{n,p_i}$ is well-behaved then by Lemma 56, $dt_{i+1} - dt_i \le 2n^{1/3}/\sqrt{f_i}$. Let $i^*$ be the smallest integer for which $G_{n,p_j}$ is well-behaved for all $i^* \le j < t$ or $i^* = t$ if $G_{n,p_{t-1}}$ is not well-behaved. We have deterministically that

$$\begin{aligned} dt_t - dt_{i^*} &\le \sum_{i=i^*}^{t-1} n^{1/3}/\sqrt{f_i} \\ &\le F^{-1/2} n^{1/3} \sum_{i=1}^{t-1} (4/5)^{i/2} = O(n^{1/3}). \end{aligned} \tag{4.1}$$

A simple argument given at the end of this section yields that $\mathbf{E}\{MWST(K_n) - dt_t\} = O(\log^6 n)$. Thus,

$$\mathbf{E}\{MWST(K_n)\} = \mathbf{E}\{dt_{i^*}\} + O(n^{1/3}). \tag{4.2}$$

It remains only to bound $\mathbf{E}dt_{i^*}$. The key to doing so is to show that for all $j$ between 0 and $t-1$

$$\mathbf{P}\{i^* = (j+1)\} \le 4e^{-\sqrt{f_j/8}} \tag{4.3}$$

Using (4.3) together with (4.1) and the fact that the longest path has length no longer than

$n$ yields that

$$\begin{aligned}
\mathbf{E}\{dt_{i^*}\} &\leq f_0^4 n^{1/3} + n\mathbf{P}\{i^* = t\} + \sum_{i=1}^{t-1} f_i^4 n^{1/3} \mathbf{P}\{i^* = i\} \\
&\leq F^4 n^{1/3} + n(4e^{-(n^{1/3}/8\log n)^{1/2}}) + \sum_{i=1}^{t-1} f_i^4 n^{1/3} \left(4e^{-(f_{i-1}/8)^{1/2}}\right) \\
&\leq F^4 n^{1/3} + O(1/n) + 4n^{1/3} \sum_{i=1}^{t-1} f_i^4 e^{-(f_{i-1}/8)^{1/2}} = O(n^{1/3}).
\end{aligned}$$

Combining this with (4.2) completes the upper bound of Theorem 55. To prove (4.3), we note that if $i^* = j+1$ and $j > 0$, then one of (I) or (II) fails for $G_{n,p_j}$. We shall show that

(A) $\mathbf{P}\left\{\text{(I) fails for } G_{n,p_j}\right\} \leq e^{-\sqrt{f_j}}$, and

(B) $\mathbf{P}\left\{\text{(II) fails for } G_{n,p_j}\right\} \leq 3e^{-\sqrt{f_j/8}}$.

This implies (4.3) since $3e^{-\sqrt{f/8}} + e^{-\sqrt{f}} < 4e^{-\sqrt{f/8}}$ for all $f > 0$.

We prove (B) using the tail bounds on the size of $H_{n,p}$ that we proved in Chapter 3. We combine these bounds with existing knowledge about the diameter of $G_{n,p}$ for *subcritical* $p$ (Łuczak, 1998), together with the fact that for $p > 1/n$, $p - 1/n = o(1/n)$, the structure of $G_{n,p}$, minus its giant component, is very similar to the structure of a subcritical random graph (Bollobás, 1984; Łuczak, 1990) (we mentioned this fact in Section 3.1).

We remind the reader that the *excess* of $G$ is the quantity $|E(G)| - |V(G)|$; trees, for example, have excess $-1$. Rényi and Szekeres (1967) and Flajolet and Odlyzko (1982) have studied the moments of the height of uniformly random labeled trees, and Łuczak (1995) has provided information about the precise number of such trees with a given height. This latter result can be used to prove tail bounds on the lengths of longest paths in uniformly random labeled graphs with small excess. We will prove (A) by combining the bounds on the size and excess of $H_{n,p}$ from Chapter 3 with these latter bounds.

We now return to a description of the final stage of the proof, in which we establish that $\mathbf{E}\{MWST(K_n) - dt_t\} = O(\log^6 n)$. It is convenient to think of growing the $MWST$ in a different fashion at this point. Let $H_t = H_{n,p_t}$ and consider an arbitrary component $C$ of $G_{n,p_t}[V - V(H_t)]$. The edge $e$ with one endpoint in $C$ and the other endpoint in some other component of $G_{n,p_t}$ and minimizes $w(e)$ subject to this is a cutedge of $G_{n,w(e)}$. Therefore $e$ is necessarily an edge of $MWST(K_n)$.

Let $E$ be the event that $|H_t| > n/\log n$ and every other component of $G_{n,p_t}$ has longest path of length at most $n^{1/6}\sqrt{\log n}$. If $E$ does not occur then one of (1) or (2) fail for $G_{n,p_t}$, so (A) and (B) tell us that

$$\mathbf{P}\left\{\bar{E}\right\} \le e^{-(n^{1/3}/\log n)^{1/2}} + 3e^{-(n^{1/3}/8\log n)^{1/2}} = O(1/n). \tag{4.4}$$

Since the edge weights are i.i.d., the second endpoint of $e$ is uniformly distributed among vertices not in $C$. If $E$ holds, it follows that with probability at least $|H_t|/n > 1/\log n$, the second endpoint is in $H_t$. If the second endpoint is not in $H_t$, we can think of $C$ joining another component to create $C'$. The component $C'$ has longest path of length at most $2n^{1/6}\sqrt{\log n}$.

(As an aside, note that $MWST(C')$ is not necessarily a tree created by Kruskal's algorithm, as there may well be edges leaving $C'$ which have weight less than $w(e)$. The technique of growing the $MWST$ of a graph by focussing on the cheapest edge leaving a *specific* component, rather than the cheapest edge joining *any* two components, is known as Prim's tree growing method (Jarńik, 1930; Prim, 1957).)

Conditional upon this choice of $e$, the edge $e'$ leaving $C'$ which minimizes $w(e')$ is also in $MWST(K_n)$. Again, with probability at least $1/\log n$ the second endpoint lies in $H_t$. If not, $C'$ joins another component to create $C''$ with longest path of length at most $3n^{1/6}\sqrt{\log n}$.

Continuing in this fashion, we see that the probability the component containing $C$ has longest path of length greater than $rn^{1/6}\sqrt{\log n}$ when it joins to $H_t$ is at most $(1-1/\log n)^r$. In particular, the probability that it has length greater than $n^{1/6}(\log n)^{7/2}$ is at most $(1 - 1/\log n)^{\log^3 n} = o(1/n^2)$.

Since $C$ was chosen arbitrarily and there are at most $n$ such components, with probability $1-o(1/n)$ none of them has longest path of length greater than $n^{1/6}(\log n)^{7/2}$ before joining $H_t$. It follows from Lemma 56 that with probability $1-o(1/n)$, $diam(MWST(K_n)) - dt_t \leq 2n^{1/6}(\log n)^{7/2} + 2$. Since $diam(MWST(K_n))$ never exceeds $n$, it follows that

$$\mathbf{E}\left\{diam(MWST(K_n)) - dt_t|E\right\} = O(n^{1/6}(\log n)^{7/2}),$$

so

$$\begin{aligned}\mathbf{E}\left\{diam(MWST(K_n)) - dt_t\right\} &\leq \mathbf{E}\left\{diam(MWST(K_n)) - dt_t|E\right\} + n\mathbf{P}\left\{\bar{E}\right\} \\ &= O(n^{1/6}(\log n)^{7/2}). \qquad (4.5)\end{aligned}$$

In Section 4.2 we use results from Chapter 3 to prove (B). In Section 4.3 we derive tail bounds on the diameters of random treelike graphs. Finally, in Section 4.4 we use these tail bounds and results from Chapter 3 to prove (A).

## 4.2 The proof of (B)

Let $\mathcal{H}$ be the set of all labeled connected graphs $H$ with vertex set $V(H) \subset \{v_1,\ldots,v_n\}$ for which $H$ has between $(3/2)fn^{2/3}$ and $(5/2)fn^{2/3}$ vertices. For $H \in \mathcal{H}$, let $C_H$ be the event that in the random graph process, $H$ is a connected component of $G_{n,p}$, and let $Bad$ be the event that in the random graph process, *no* element of $\mathcal{H}$ is a connected component of $G_{n,p}$. For *any* event $E$ we may write

$$\mathbf{P}\{E\} \leq \mathbf{P}\{Bad\} + \sum_{H\in\mathcal{H}} \mathbf{P}\{E|C_H\}\mathbf{P}\{C_H\}.$$

If *Bad* occurs then $G_{n,p}$ has *no* component of size between $(3/2)fn^{2/3}$ and $(5/2)fn^{2/3}$, so by Theorem 47 applied with $c = 1/2$, $\mathbf{P}\{Bad\} \leq 85e^{-f^3/2^{10}\cdot 100} \leq e^{-f}$ for $f$ large enough. Therefore,

$$\begin{aligned}\mathbf{P}\{E\} &\leq e^{-f} + \sum_{H\in\mathcal{H}} \mathbf{P}\{E|C_H\}\mathbf{P}\{C_H\}.\\ &\leq e^{-f} + \mathbf{E}\{|\{H : C_H \text{ holds }\}|\}(\max_{H\in\mathcal{H}}\mathbf{P}\{E|C_H\}).\end{aligned}$$

Applying Theorem 54 to bound the above expectation, presuming $f$ is large enough that $2e^{-f^2/2} < 1$, for $n$ large enough we have that

$$\mathbf{P}\{E\} \leq e^{-f} + 2\max_{H\in\mathcal{H}}\mathbf{P}\{E|C_H\}. \tag{4.6}$$

Let $p = 1/n + f/n^{4/3}$ and let $p' = 1/n + (5/4)f/n^{4/3}$, and recall that $H_{n,p}$ is the largest component of $G_{n,p}$. We will apply equation (4.6) to the event *Long* that some component of $G_{n,p'}[V - V(H_{n,p})]$ has longest path of length at least $n^{1/3}/f^{1/4}$.

For any graph $H \in \mathcal{H}$, the graph $G_{n,p'}[V - V(H)]$ is $G_{n',p'}$ for some $n' \leq n - (3/2)fn^{2/3}$, and so

$$\begin{aligned}p' &= \frac{1}{n} + \frac{(5/4)f}{n^{4/3}} \leq \frac{1}{n'}\left(1 - \frac{(3/2)f}{n^{1/3}}\right) + \frac{(5/4)f}{n^{4/3}}\\ &< \frac{1}{n'} - \frac{(1/4)f}{n^{4/3}} \leq \frac{1}{n'} - \frac{(1/8)f}{(n')^{4/3}},\end{aligned} \tag{4.7}$$

for $n$ large enough. Let *Large*$(H)$ be the event that either

(a) $G_{n,p'}[V - V(H)]$ has a component of size larger than $8n^{2/3}/f$, or a complex component of size greater than $f$, or

(b) $G_{n,p'}[V - V(H)]$ has a tree or unicyclic component of size at most $8n^{2/3}/f$ and longest path of length at least $36n^{1/3}\log f/\sqrt{f}$.

$G_{n,p'}[V - V(H)]$ is a subcritical random graph by (4.7). For $f$ large enough, Theorem 52 applied with $k = f/10$, say, therefore yields that (a) occurs with probability at most $3e^{-f/10}$. Theorem 52 also yields that (b) occurs with probability at most $e^{-\sqrt{f/8}}$. As $3e^{-f/10} \leq e^{-\sqrt{f/8}}$ for $f$ large enough, this yields $\mathbf{P}\{Large(H)\} \leq 2e^{-\sqrt{f/8}}$ for $f$ large enough.

If $C_H$ occurs but (a) does not then $G_{n,p}$ certainly has no component of size larger than $H$ so $H = H_{n,p}$. Also, for $f$ large enough $36n^{1/3}\log f/\sqrt{f} < n^{1/3}/f^{1/4}$, so for such $f$, if $C_H$ occurs and $Large(H)$ does not occur then $Long$ does not occur. Furthermore, $Large(H)$ is independent of $C_H$ as the two events are determined by disjoint sets of edges. Therefore,

$$\mathbf{P}\{Long|C_H\} \leq \mathbf{P}\{Large(H)|C_H\} = \mathbf{P}\{Large(H)\} \leq 2e^{-\sqrt{f/8}},$$

which combined with (4.6) applied with $E = Long$ yields

**Lemma 57.** *There exists $F > 1$ such that for $f > F$, for $n$ large enough,* $\mathbf{P}\{Long\} \leq 3e^{-\sqrt{f/8}}$.

This proves the bound (B) stated in the introduction.

## 4.3 Longest paths in random treelike graphs

As mentioned in the introduction, information about the excess of a random connected graph gives us information about its diameter. This is, in essence, because a random graph with only a few more edges than vertices is "treelike"; in this section we make this idea precise.

### 4.3.1 The diameter of uniform trees

We first collect the required bounds on the diameter of trees. A uniform random rooted tree of size $s$ is a tree chosen uniformly at random from among all rooted labeled trees with $s$ nodes. Rényi and Szekeres (1967) and Flajolet and Odlyzko (1982) have calculated the asymptotics of the moments of the height $H_s$ of a uniform random rooted tree $R_s$ of size $s$ and provided sharp information about the number of uniformly random rooted trees of size $s$ and height $c\sqrt{s}$ for constant $c$. Through combinatorial arguments, Łuczak (1995) has extended these results to count the number of such trees when $c = c(s)$ is $\omega(1)$. The version of Łuczak's result that we need can be stated as:

**Theorem 58** (Łuczak (1995), p. 299). *There is $C > 0$ such that for $s$ large enough, for all $t \geq C\sqrt{s}$,*

$$\mathbf{P}\{H_s = t\} \leq \frac{e^{-t^2/4s}}{\sqrt{s}}.$$

In fact, this theorem is weaker than what Łuczak proved, but it is easier to state and suffices for our purposes. We have as an immediate consequence:

**Corollary 59.** *There is $C > 0$, such that for $s$ large enough, for all $c \geq C$,*

$$\mathbf{P}\left\{H_s \geq c\sqrt{s}\right\} \leq 2e^{-c^2/4}.$$

*Proof.* By Theorem 58, we have that for $c \geq C$,

$$\begin{aligned}
\mathbf{P}\left\{H_s \geq c\sqrt{s}\right\} &\leq \sum_{t=\lceil c\sqrt{s}\rceil}^{s} \frac{1}{\sqrt{s}} e^{-t^2/4s} \\
&\leq \sum_{i=0}^{\lceil\sqrt{s}-c\rceil} \sum_{t=\lceil(c+i)\sqrt{s}\rceil}^{\lceil(c+i+1)\sqrt{s}\rceil-1} \frac{1}{\sqrt{s}} e^{-t^2/4s} \\
&\leq \frac{\sqrt{s}+1}{\sqrt{s}} \sum_{i=0}^{\lceil\sqrt{s}-c-1\rceil} e^{-(c+i)^2/4} \\
&\leq 2e^{-c^2/4},
\end{aligned}$$

as long as $c$ and $s$ are large enough. □

There is a natural $s$-to-1 map from rooted trees of size $s$ to unrooted trees of size $s$, obtained by "unrooting". Clearly, if $T_s$ is an unrooted tree corresponding to $R_s$ via this map, then $lp(T_s) = lp(R_s) \leq 2H_s$. As a consequence,

**Lemma 60.** *Let $T_s$ be a uniformly random unrooted tree (a* Cayley *tree) on $s$ nodes. Then there is $C > 0$ such that for $s$ large enough, for all $c > C$*

$$\mathbf{P}\left\{\mathrm{lp}(T_s) \geq 2c\sqrt{s}\right\} \leq \mathbf{P}\left\{H_s > c\sqrt{s}\right\} \leq 2e^{-c^2/4}. \tag{4.8}$$

Lemma 60 is the key fact about random trees that allows us to bound the lengths of the longest paths of uniformly random connected tree-like graphs. We now focus our attention on bounding longest paths in such graphs. In doing so, it is useful to describe them in a way that emphasize some underlying tree structures.

### 4.3.2 Describing graphs with small excess

Given a connected labeled graph $G$ with excess $q$, define the *core* $C = C(G)$ of $G$ to be the maximum induced subgraph of $G$ which has minimum degree 2. To see that the core is indeed unique, we note that it is precisely the graph obtained by repeatedly removing vertices of degree 1 from $G$ until no such vertices exist (so in particular, if $G$ is a tree then $C$ is empty). It is clear from the latter fact that $G[V - V(C)]$ is a forest, so if $v_i \in V - V(C)$, then there is a unique shortest path in $G$ from $v_i$ to some $v_j \in V(C)$. We thus assign to each vertex $v_j \in V(C)$ the set of labels

$$L_{v_j} = \{j\} \cup \{i | \text{the shortest path from } v_i \text{ to } C \text{ ends at } v_j\}.$$

We next define the *kernel* $K = K(G)$ to be the multigraph obtained from $C(G)$ by replacing all paths whose internal vertices all have degree 2 in $C$ and whose endpoints have degree at least three in $C$ by a single edge (see, e.g., Janson et al., 2000). If $q < 1$ we agree that the kernel is empty; otherwise the kernel has minimum degree 3 and precisely $q$ more edges than vertices. It follows that the kernel always has at most $2q$ vertices and at most $3q$ edges. We denote the multiplicity of edge $e$ in $K$ by $m(e)$. We think of $K$ as a simple graph in which edge $e$ has positive integer weight $m(e)$, to emphasize the fact that parallel edges are indistinguishable. We may keep track of what vertices correspond to *edges* of $K(G)$ as we did for *vertices* of $C(G)$: if $P_1, \ldots, P_{m(e)}$ are paths of $C(G)$ corresponding to edge $e = xy$ of $K(G)$, we let $L_e^i = \bigcup_{v \in V(P_i)-x-y} L_v$ (if $P_i = xy$ then $L_e^i = \emptyset$) and assign a *set of sets of labels* $\{L_e^1, \ldots, L_e^{m(e)}\}$ to $e$. We emphasize that permuting the order of $P_1, \ldots, P_{m(e)}$ does not change the label of $e$.

Given a labeled graph $G$, the above reduction yields a labeled multigraph $K$ and sets $L_v$ for each vertex of $K$, $\{L_e^1, \ldots, L_e^{m(e)}\}$ for each edge of $K$. Conversely, any graph with nonempty labeled kernel $K$ to which such sets have been assigned can be described uniquely in the following way:

- For all $v_i \in V(K)$, let $T_{v_i}$ be a labeled tree with labels from $L_{v_i}$.
- For all $e = xy \in V(K)$, and all $i = 1, 2, \ldots, m(e)$, let $T_e^i$ be a labeled tree with labels from $L_e^i$ (if $L_e^i = \emptyset$ then $T_e = \emptyset$ - this can occur for at most one $i \in \{1, 2, \ldots, m(e)\}$). If $L_e^i \neq \emptyset$, our description depends on whether $e$ is a loop, i.e., on whether $y = x$:
  - If $x \neq y$ then mark an element of $L_e^i$ with an **X** and mark an element of $L_e^i$ with a **Y**. We allow that the same element of $L_e^i$ receives both markers.
  - If $x = y$ then place two markers of type **X** on elements of $L_e^i$. Again, we allow that the same element of $L_e^i$ receives both markers.

Observe that in marking elements of $L_e^i$, if $e = xy$ and $x \neq y$ then there are $|L_e^i|^2$ ways to place the markers. If $e = xx$ then there are $|L_e^i| + \binom{|L_e^i|}{2} = (|L_e^i| + 1)|L_e^i|/2$ ways to place the markers as we may either choose an element of $L_e^i$ and place both **X** markers on it, or we may choose two distinct elements of $L_e^i$ and place an **X** marker on each.

We obtain $G$ from this description as follows:

1. for all $v_j \in V(K)$, identify the vertices $v_j \in V(K)$ and $v_j \in T_{v_j}$, then
2. for all loops $e = xx \in E(K)$, choose a copy of $e$ for each nonempty tree $T_e^i$, $1 \leq i \leq m(e)$. Remove this edge and let $x$ be adjacent to the vertices in $T_e^i$ marked with **X**.
3. for all $e = xy \in E(K)$ with $x \neq y$, choose an edge $xy$ for each nonempty tree $T_e^i$, $1 \leq i \leq m(e)$. Remove this edge, then let $x$ (respectively $y$) be adjacent to the vertex in $T_e^i$ marked with **X** (respectively **Y**).

Clearly labeled graphs with distinct labeled kernels are not identical. Now, let $G$, $G'$ be graphs with the same labeled kernel $K$. If for some $v \in V(K)$, $L_v \neq L'_v$ or for some $e \in E(K)$, $\{L_e^1, \ldots, L_e^{m(e)}\} \neq \{L_e'^1, \ldots, L_e'^{m(e)}\}$, then $G, G'$ are not identical. Hence, given a labeled kernel $K$, and sets of labels $\{L_v | v \in V(K)\}$, $\bigcup_{e \in E(K)} \{L_e^1, \ldots, L_e^{m(e)}\}$, and distinguished elements of the nonempty sets $L_e^i$ as described above, there are

$$\prod_{v \in V(K)} |L_v|^{|L_v|-2} \prod_{e \in E(K)} \prod_{i: L_e^i \neq \emptyset} |L_e^i|^{|L_e^i|-2}$$

possible graphs, corresponding to the choices of a tree for each set $L_v$ and for each set $L_e^i$. It follows that if $G$ is a uniformly random connected labeled graph with $p$ vertices and excess $q \geq 1$ specified by its kernel $K$ and a description as above, then conditional on the sizes of their elements, the sets $\mathcal{T}_V = \{T_v | v \in V(K)\}$ and $\mathcal{T}_E = \bigcup_{e \in E(K)} \{T_e^1, \ldots, T_e^{m(e)}\}$ must be uniformly random amongst all such sets. As a consequence, conditional on their sizes, the unrooted labeled trees in $\mathcal{T}_V$ and in $\mathcal{T}_E$ must be uniformly random; i.e., they are simply Cayley trees.

Labeled unicyclic graphs (graphs with excess 1) have empty kernels but nonempty cores; they can be described in a similar but simpler way. Suppose we are given a labeled graph $G$ with unique cycle $C$. We let $T_1$ be the unique maximal tree containing vertex $v_1$ and containing exactly one element $v^*$ of $C$ – set $\mathcal{T}_V = \{T_1\}$ and mark $v^*$. The vertex $v^*$ has exactly two distinct neighbours $w^*$, $x^*$ in the tree $T_2$ induced by the vertices in $V(G) - V(T_1)$; we let $\mathcal{T}_E = \{T_2\}$ and mark $w^*$, $x^*$. Given trees $T_1$, $T_2$ such that $v_1 \in V(T_1)$, $T_1$ contains one marked vertex $v^*$ and $T_2$ contains two marked vertices $w^*, x^*$, we may construct a unicyclic graph $G$ by letting $w^*$ and $x^*$ be adjacent to $v^*$. The only difference between this bijection and that given for graphs with nonempty kernel is that now we need to mark a vertex in the tree in $\mathcal{T}_V$. As above, this bijection shows that conditional on their sizes, the trees $T_1$ and $T_2$ are Cayley trees.

### 4.3.3 The diameter of graphs with small excess

With this latter fact in hand, it is easy to prove bounds on $lp(G)$. Recall that the *net excess* of a connected graph $G$ is equal to the excess of $G$, plus 1.

**Lemma 61.** *Let $G$ be a uniformly random labeled connected graph on $s$ vertices and with net excess $q$. Then there is $C$ such that for $s$ large enough, for all $c \geq C$,*

$$\mathbf{P}\left\{\text{lp}(G) \geq 2(5q+1)c\sqrt{s} + 10q\right\} \leq \max\{10q, 2\}e^{-c^2/4}. \tag{4.9}$$

*Proof.* The bound holds by (4.8) if $q = 0$. If $q > 0$ then let the sets $\mathcal{T}_V$ and $\mathcal{T}_E$ be defined as above, and let $\mathcal{T} = \mathcal{T}_V \cup \mathcal{T}_E$ - then $|\mathcal{T}| \leq 5q$ as if $q \geq 2$, the kernel has at most $2q$ vertices and at most $3q$ edges, counting multiplicity, and if $q = 1$ then $|\mathcal{T}_V| = |\mathcal{T}_E| = 1$. Trivially, any path $P$ in $G$ is composed of paths from the trees in $\mathcal{T}$ together with edges of $G$ that are not edges of some tree in $\mathcal{T}$. For a given tree $T$, if $P$ does not have an endpoint in $T$ then it must enter and exit $T$ at most once, i.e., the intersection of $P$ with $T$, if nonempty, is itself a path. $P$ may also enter one or two of the trees without leaving them – such trees must contain an endpoint of $P$. If the endpoints are in distinct trees then the intersection of these trees with $P$ are both paths; if the endpoints are in the same tree then that tree's intersection with $P$ consists of two paths.

(In fact, $P$ can *not* enter every tree. If $q > 1$, for example, then the set of vertices and edges of the kernel that have trees intersecting $P$ can not itself contain a cycle in the kernel. We crudely bound the length of $P$ by supposing that it may contain a path from every tree and two paths from at most one tree, so at most $(5q+1)$ paths from trees of $\mathcal{T}$ in total.)

Each time the path $P$ enters or exits a tree, it uses an edge of $G$ that is not an edge of a tree in $\mathcal{T}$. By the definition of the trees in $\mathcal{T}$, there are precisely two such edges for each nonempty tree $T_e^i$; thus there are at most $10q$ such edges in total. We thus have

$$\mathbf{P}\left\{lp(G) \geq 2(5q+1)c\sqrt{s}+10q\right\} \leq \mathbf{P}\left\{\max_{T\in\mathcal{T}} lp(T) \geq 2c\sqrt{s}\right\}. \tag{4.10}$$

We choose $C$ large enough so that if $|T| \geq C\sqrt{s}$ and $c \geq C$ then Lemma 60 applies to $T$ with this choice of $c$. For $c \geq C$, for all $T \in \mathcal{T}$ either $|T| < C\sqrt{s}$, in which case $\mathbf{P}\left\{lp(T) \geq 2c\sqrt{s}\right\} = 0$, or $|T| \geq C\sqrt{s}$, in which case since $|T| \leq s$, there is $c' \geq c$ such that $2c\sqrt{s} = 2c'\sqrt{|T|}$. In the latter case, $\mathbf{P}\left\{lp(T) \geq 2c\sqrt{s}\right\} = \mathbf{P}\left\{lp(T) \geq 2c'\sqrt{|T|}\right\} \leq 2e^{-c^2/4}$ by Lemma 60. Therefore, by a union bound applied to the right-hand-side of (4.10) we have

$$\mathbf{P}\left\{lp(G) \geq 2(5q+1)c\sqrt{s}+10q\right\} \leq 5q(2e^{-c^2/4}) = 10qe^{-c^2/4}. \qquad \square$$

## 4.4 The proof of (A)

We apply Lemma 61 to bound $lp(H_{n,p})$. First, let $D$ be the event that $H_{n,p}$ has size greater $(5/2)fn^{2/3}$, which we denote by $s$, or excess greater than $150f^3$, which we denote by $q$. If $D$ occurs then either

(a) $H_p$ has size greater than $s$ or excess greater than $q$ or

(b) $H_p \neq H_{n,p}$.

By Theorem 47, the event (a) occurs with probability at most $85e^{-f^3/2^{10}\cdot 100}$, which is at most $e^{-f}$ for $f$ large enough. By Theorem 53, the probability that (b) occurs is most $2e^{-f^2/2}$, also at most $e^{-f}$ for $f$ large enough, so $\mathbf{P}\left\{D\right\} \leq 2e^{-f}$. Letting $E_p$ be the event that

$lp(H_{n,p}) > f^4 n^{1/3}$ we have

$$\mathbf{P}\{E_p\} \le \mathbf{P}\{\overline{D}\} + \mathbf{P}\{E_p|D\} \le 2e^{-f} + \mathbf{P}\{E_p|D\} \le 2e^{-f} + \mathbf{P}\{E_p|D\}. \qquad (4.11)$$

Furthermore, $2(5q+1)c\sqrt{s}+10q < 5000cf^{7/2}n^{1/3}$ for $f$ large enough. It follows by applying Lemma 61 with $c = f^{1/2}/5000$ (which is at least $C$ for $f$ large enough) that

$$\mathbf{P}\{E_p|D\} \;\le\; 1500f_r^3 e^{-c^2/4} \le 1500 f_r^3 e^{-f/(10000^2)} \le e^{-f/2^{30}},$$

for $f$ large enough. By this bound and (4.11), we have $\mathbf{P}\{E_p\} \le 2e^{-f} + e^{-f/2^{30}}$, which is at most $e^{-\sqrt{f}}$ for $f$ large enough. This proves the bound (A) of the introduction, and completes the proof.

## 4.5 Conclusion

We have pinned down the growth rate of the diameter of the minimum weight spanning tree of $K_n$ whose edges are weighted with i.i.d. $[0,1]$-uniform random variables. We did so by using an equivalence between Kruskal's algorithm for growing the minimum weight spanning tree of $K_n$ and a restricted version of the random graph process for $G_{n,p}$. Theorem 55 raises a myriad of further questions. Two very natural questions arising directly from our result are: does $\mathbf{E}\{diam(MWST(K_n))\}/n^{1/3}$ converge to a constant? What constant?

Theorem 55 seems related not only to the diameter of minimum spanning trees, but also to the diameter of $G_{n,p}$ itself. This latter problem still seems difficult when $p$ gets closer to $1/n$ (Chung and Lu, 2001). A key difference between the analysis required for the two problems is captured by the fact that there is some (random) $p^*$ such that for $p \ge p^*$, the diameter of $G_{n,p}$ is decreasing, whereas the diameter of $F_{n,p}$ is increasing for all $0 \le p \le 1$. At some point in the range $(p - 1/n) = o(1/n)$, the diameters $G_{n,p}$ and $F_{n,p}$ diverge; the precise behavior of this divergence is unknown. If the expected diameter of $G_{n,p}$ is unimodal, for example, then it makes sense to search for a specific probability $p^{**}$ at which the expected diameters of $G_{n,p}$ and $F_{n,p}$ cease to have the same order. In this case, what can we say about $|p^* - p^{**}|$? For $p = (1+\epsilon)/n$ and $\epsilon > 0$ constant, the diameter of $G_{n,p}$ is concentrated on a finite number of values, whereas it follows from results of Łuczak et al. (1994) that this is not the case in $G_{n,p}$ for $p = 1/n + O(1/n^{4/3})$. How does this behavior change as $p$ increases through the critical window? Answering such questions would seem to be a prerequisite to a full understanding of the diameter of $G_{n,p}$ in the critical range.

Can this method be used to study the diameters of minimum weight spanning trees of other graphs? Bollobás et al. (1992) have exhibited a modified branching process for growing random subgraphs of the hypercube. Can a corresponding random walk be found, allowing their approach to be combined with ours to find the diameter of a random minimum weight spanning tree of the hypercube?

# Chapter 5

# Maxima in branching random walks

A branching random walk (or BRW) starts with a single particle $r$ placed at position 0 on the real line. This particle splits into a random, finite number of children according to some distribution $B$ (so $B \in \mathbb{N}$), which we call the *branching distribution.* Each of its children is independently and randomly displaced according to a second, real-valued random variable $E$, which we call the *step size.* These children form the *first generation* of the branching random walk. Each child $v$ then splits independently into a random number of children, again according to the distribution $B$, and these children are independently and randomly displaced from the position of $v$ according to $E$ to form the second generation. This operation is iterated infinitely. We note that if $B$ were identically 1, then this would simply be a random walk on $\mathbb{R}$ distributed like $E$.

We may equivalently define branching random walks in terms of a branching process (or *Galton-Watson process*), which is defined by a single parameter, a non-negative integer valued random variable $B$. We start with a single node, the *zeroth generation* $G_0 = \{r\}$ This node $r$ has a random number of child nodes according to $B$; these offspring form the first generation $G_1 = \{v_1, \ldots, v_B\}$. Each of these children independently have a random number of offspring according to the distribution $B$; the collection of these offspring form the second generation $G_2$. We continue in this manner to define $G_3, G_4, \ldots$; the union of these nodes and the edges that link child to parent form a possibly infinite tree $T = T(B)$.

To define a BRW from a branching process, we choose a second random variable $E$ and assign an independent copy of $E$ to each edge of the random tree $T$. The resulting structure is equivalent to that given by the first formulation of the BRW, as we may see by thinking of each node $v$ as having a position $S_v$ on the real line which is by the sum of the edge labels on the path connecting $r$ to $v$. (For the formal details of a probabilistic construction of branching random walks, see, e.g., Harris (1963).)

One of the most well-studied parameters associated with branching random walks is the *minimum after $n$ steps*, which we denote $M_n^b$. In the branching process formulation, we may define $M_n^b$ as the minimum value of $S_v$ over all nodes $v$ having unweighted depth $n$ in $T$. Viewed as a point process on the line, this minimum is the position of the leftmost particle of the $n$'th generation. (We set $M_n^b = \infty$ if $G_n = \emptyset$, i.e., if the process does not survive for $n$ generations. Clearly, if $M_n^b = \infty$ for some $n$ then $M_{n'}^b = \infty$ for all $n' > n$.) One of the first properties of $M_n^b$ to be investigated was when it obeys a *"law of large numbers"*, i.e., for what branching random walks is there a constant $\gamma$ such that $M_n^b/n \to \gamma$ (for the moment we decline to specify what sort of convergence we mean when we write "$\to$").

If $\mathbf{E}B \leq 1$ then with probability 1, there is $n$ such that $G_n^b = \emptyset$ (see Athreya and Ney, 1972) (except in the special case that $B = 1$ deterministically, in which case $M_n$ is just the sum of $n$ independent identically distributed random variables and is well understood); if such an $n$ exists we say the BRW *does not survive.* In this case $M_n^b/n \to \infty$ almost surely. If $\mathbf{E}B = \mu > 1$ then with positive probability, the BRW survives; we denote the event that the BRW survives by $\mathcal{S}$. When the BRW survives, we may ask whether $M_n^b/n$ has a *finite* limit. To see when we might expect such a limit to exist, we first note that there is a random variable $W$ such that $G_n^b/\mu^b \to W$ almost surely, and if $\mu > 1$ then $\mathbf{P}\{W > 0\} = \mathbf{P}\{\mathcal{S}\}$ (see Athreya and Ney, 1972). Intuitively, this suggests that if the BRW survives then $|G_n^b|$ almost surely exhibits exponential growth. It seems natural, then, that $E$ should need to satisfy some form of exponential lower tail bound in order for $M_n^b/n$ to converge to a finite limit.

In a sequence of papers, each building on the results of the last, Hammersley (1974), Kingman (1975) and Biggins (1976b) proved a law of large numbers for $M_n^b$ in a very general setting. They showed that if $E_1, \ldots, E_B$ are the locations of the particles in the first generation (so $E_1, \ldots, E_B$ are independent and distributed as $E$), $\mathbf{E}B > 1$, and there is $c > 0$ for which $1 < \mathbf{E}\left\{\sum_{i=1}^{B} e^{-cE_i}\right\} < \infty$, then there is a finite constant $\gamma$ such that $M_n^b/n \to \gamma$ as $n \to \infty$ almost surely given that the BRW survives. This result is now known as the Hammersley-Kingman-Biggins theorem (we have stated their theorem in slightly less than its full generality). The condition on the sum $\mathbf{E}\left\{\sum_{i=1}^{B} e^{-cE_i}\right\}$ essentially imposes the exponential tail bounds we suggested would be needed for such a result to hold.

When Hammersley (1974) initiated research into the behavior of $M_n$, he posed several questions to which complete answers remain unknown. In particular, he asked if more detailed information about $M_n - \gamma n$ than that given by the above law of large numbers can be found, about the behavior of the differences $\mathbf{E}M_{n+1} - \mathbf{E}M_n$ as $n$ becomes large, and about the higher moments of $M_n$. In this chapter we answer Hammersley's questions for a class of branching random walks with certain "nice" properties – we will be more explicit about our results shortly. See also Biggins (1977); Bramson (1978b); Biggins (1979); Durrett (1983); Dekking and Host (1991); McDiarmid (1995); Bachmann (2000) for further investigations into the behavior of $M_n^b$ and of Hammersley's questions. We will discuss certain results from these papers in the course of the chapter.

When the step size $E$ is non-negative, $E$ may be viewed as a "time to birth": the label on edge $vw$ is the time it takes for $v$ to produce child $w$. We may then ask: what particles are born before time $t$? This set of particles is precisely the set of nodes of $v$ for which $S_v \leq t$. As $E$ is non-negative, if $S_v > t$ for some node $v$, then for all descendants $w$ of $v$, $S_w > t$. When we view the branching random walk in its branching process formulation, then, the set of nodes corresponding to the particles born before time $t$ form a subtree of $T$, which we denote $T_t^b$. The *depth* of a node $v$ is the number of edges on the path from $r$ to $v$ – the *height* $H_t^b$ of $T_t^b$ is the (unweighted) greatest depth of any node of $T_t^b$. The random variables $H_t^b$ and $M_n^b$ are closely linked: namely, we have the relations

$$M_n^b = \inf\{m \mid H_m^b \geq n\} \quad \text{and} \quad H_t^b = \max\{n \in \mathbb{N} \mid M_n^b \leq t\}. \tag{5.1}$$

It follows that results on the distribution of $M_n^b$ yield results on the distribution of $H_t^b$ and vice-versa.

The trees $T_t^b$ are quite similar to a family of trees arising in the analysis of algorithms, called *ideal trees.* Suppose we are given a fixed integer $d \geq 2$ and a random vector $\mathcal{X} = (X_1, \ldots, X_d)$ satisfying $\sum_{i=1}^d X_i = 1$ and whose entries are identically distributed as some variable $X \geq 0$. (Such a vector is called a *split vector* (Broutin and Devroye, 2005) – we may think of it as splitting the interval $[0,1]$ into sub-intervals of length $X_1, \ldots, X_d$. We emphasize that the component random variables of $\mathcal{X}$ are *not* necessarily independent – in fact, since they are identically distributed and sum to 1, they can only be independent if they are all deterministically equal to $1/d$.)

We consider the infinite $d$-ary tree $T_\infty^d$ with root $r$ and assign a copy $\mathcal{X}_v$ of $\mathcal{X}$ to each node $v$ of $T_\infty^d$. We think of the children of each node $v$ as being ordered (say as $w_1, \ldots, w_d$), and associate the elements of $\mathcal{X}_v = (X_{v,1}, \ldots, X_{v,d})$ to the edges $vw_1, \ldots, vw_d$. (We will sometimes refer to the *label of an edge* $e$ and write $X_e$ without reference to the split vector of which $X_e$ is a component.) Finally, we assign each node $v$ a label $L_v$ which is the product of the edge labels on the path from $r$ to $v$. For $t \geq 0$, the *ideal tree* $T_t^i$ is the subtree of $T_d^\infty$ consisting of all nodes $v$ for which $L_v \geq d^{-t}$; we denote the height of $T_t^i$ by $H_t^i$.

To link ideal trees and branching random walks, we transform the edge labels $X_e$, letting $E_e = -\log_d X_e$, so that $E_e$ takes values in $[0, \infty]$. Defining the vertex labels $S_v$ just as for the branching random walk, it follows that $T_t^i$ is the subtree of $T_\infty^d$ consisting of nodes $v$ for which $S_v \leq t$. $T_t^i$ is very similar to a branching random walk, but has the key difference that there is dependence between certain edge labels. (Also, the branch factor is deterministically $d$, whereas for the branching random walk we allowed a random number of children $B$.)

Ideal trees were introduced by Devroye (1986), who used them to find the first order term of the height of random binary search trees; his lower bound exploited the close connection between ideal trees and branching random walks. In this work he studied the ideal tree with split vector $\mathcal{X} = (U, 1-U)$, where $U$ is a uniform $[0,1]$ random variable (we remark that $1-U$ is also a uniform $[0,1]$ random variable). He showed in particular that for this split vector, the height $H_t^i$ of the ideal tree $T_t^i$ satisfies $\mathbf{E}H_t^i/t \to c^* \ln 2$, where $c^*$ is the unique solution in $(2, \infty)$ of the equation $c \ln(2e/c) = 1$. Later, Devroye and Reed (1995) expanded on this work, showing that $\mathbf{E}H_t^i - (c^* \ln 2)t = O(\ln t)$, and $\mathbf{Var}\{H_t^i\} = O(\ln^2 t)$. Substantially expanding on the same approach, Reed (2003) showed that $\mathbf{Var}\{H_t^i\} = O(1)$. Reed additionally proved that $\mathbf{E}H_t^i = (c^* \ln 2)t - 3 \ln t / 2 \ln(c^*/2) + O(1)$, and proved exponential tail bounds for $\mathbf{P}\{|H_t^i - \mathbf{E}H_t^i| \geq x\}$. After being apprised of Reed's work, Drmota (2003) proved using completely different techniques that $\mathbf{Var}\{H_t^i\} = O(1)$, and additionally proved that the distribution function for $H_t^i - \mathbf{E}H_t^i$ converges uniformly in $n$ to some implicitly defined distribution function $\Phi$. (In fact, the primary objective of all the above work was to prove results about the height of random binary search trees, not ideal trees. We have stated the results in terms of ideal trees as they are the focus of this chapter. Chauvin and Drmota (2005) have recently used Drmota's approach to show that the variance of the height of random $m$-ary search trees is $O(1)$.)

As they will be important in the remainder of the chapter, we now give a very brief sketch of some of the key ideas from Devroye (1986), Devroye and Reed (1995) and Reed (2003). We consider the transformed edge labels given by $E = -\log_2 U$ as above. Since $U$ is a uniform $[0,1]$ random variable, $E$ is an exponential mean 1 random variable. Fix an integer $h$ and some path $r = v_0, v_1, \ldots, v_h$ ending at depth $h$. Let the transformed labels on the path to $v_h$ be $E_1, \ldots, E_h$. Letting $S_i = E_1 + \ldots + E_i$ for $1 \leq i \leq h$, the $S_i$ form a random walk (note also that $S_i = -\log L_{v_i}$). The probability that $v_h$ is in $T_t^i$ is precisely the probability that $S_h \leq t$. Denote by $N_{t,h}$ the set of nodes at depth $h$ that are in $T_t^i$; by linearity of expectation and symmetry, $\mathbf{E}|N_{t,h}| = 2^h \mathbf{P}\{S_h \leq t\}$. It follows that when $\mathbf{P}\{S_h \leq t\} = o(2^{-h})$, we have $\mathbf{E}N_{t,h} = o(1)$, so $\mathbf{P}\{H_t^i \geq h\} = o(1)$. Furthermore, bounding $\mathbf{P}\{S_h \leq t\}$ is straightforward as $S_h$ is a sum of $h$ independent exponential mean 1 random variables, i.e., it has a gamma distribution with parameter $h$. Classic results on the gamma distribution then directly yield bounds on $\mathbf{P}\{S_h \leq t\}$. In fact, letting $h^* > 0$ be the smallest integer for which $\mathbf{P}\{H_t^i \geq h^*\} < 2^{-h^*}$, it turns out that the bounds given by the above approach are enough to show that $\mathbf{E}H_t^i \leq h^* + O(1)$. (This was the technique used by Devroye (1986) to prove an upper bound on $\mathbf{E}H_t^i$.)

To prove a lower bound on $\mathbf{E}H_t^i$, it is *not* enough to find an integer $h$ for which $\mathbf{E}|N_{t,h}| = \Omega(1)$. Indeed, it turns out that for $h^*$ as defined above, $\mathbf{E}|N_{t,h^*}| = \Omega(1)$ but there is $\alpha > 0$ such that $\mathbf{P}\{|N_{t,h^*}| \geq 1\} = \mathbf{P}\{H_t^i \geq h^*\} = o(d^{-\alpha t})$. (It follows immediately that $\mathbf{E}\{|N_{t,h^*}| \mid |N_{t,h^*}| \geq 1\} = \Omega(d^{\alpha t})$ – in other words, knowing that the height of $T_t^i$ is at least $h^*$ increases the expected number of nodes in $T_t^i$ at depth $h^*$ by at least $d^{\alpha t}$.)

The lower bound from Devroye and Reed (1995) is loosely based on the fact that for *any* non-negative, integer random variable $Z$, $\mathbf{E}Z = \mathbf{E}\{Z|Z \geq 1\}\mathbf{P}\{Z \geq 1\} + 0$, so as long as $\mathbf{P}\{Z \geq 1\} > 0$ we may write

$$\mathbf{P}\{Z \geq 1\} = \frac{\mathbf{E}Z}{\mathbf{E}\{Z|Z \geq 1\}}, \tag{5.2}$$

so studying $\mathbf{P}\{Z \geq 1\}$ amounts to studying how much the expected value of $Z$ increases, given that $Z$ is at least 1. They were thus motivated to search for a random variable $Z_h$ for which (1) if $Z_h \geq 1$ then $T_t^i$ has height at least $h$, and (2) knowing that $Z_h$ is at least 1 does not increase the expected value of $Z_h$ by "very much".

In order define the random variable $Z_h$ used by Devroye and Reed, we first discuss how the behavior of the random walk ending at $v_h$ affects the number of other nodes of $T_t^i$ at depth $h$. Recall that we fixed a path $v_0, v_1, \ldots, v_h$; for $1 \leq i \leq h$, the node $v_{i-1}$ has a child $v_i$ on the path and another child, which we call $w_i$. Every element of $T_t^i$ at depth $h$ aside from possibly $v_h$ is in one of the subtrees rooted at some $w_i$ (for $1 \leq i \leq h$), and the expected number of elements in the subtree rooted at $w_i$ depends crucially on the sum $S_{i-1} = -\log L_{v_{i-1}}$. This dependence is intuitively clear, since $L_{v_{i-1}}$ and $L_{w_i}$ differ only by a factor of $X_{v_{i-1}w_i}$, which is very likely $O(1)$. Furthermore, the subtree of $T_t^i$ rooted at $w_i$ is distributed as $T_{t/L_{w_i}}^i$, so, generally speaking, the larger $S_{i-1}$, the smaller $t/L_{w_i}$ and the smaller the chance that $w_i$ has a descendent in $T_t^i$ at depth $h$. It is thus plausible that if $S_i$ is "large" for most $1 \leq i \leq h$ then given such information, the conditional expected number of nodes of $T_t^i$ at depth $h$ is not too large.

This idea motivated Devroye and Reed to call $v_h$ a *good* node if $v_h \in T_t^i$ (so $S_h \leq t$) *and additionally* $S_i \geq i \cdot S_h/h$ for all $1 \leq i \leq h$ (we extend this definition to all nodes at depth $h$ by symmetry). We denote by $G_{n,h}$ the set of good nodes at depth $h$. The random variable $|G_{n,h}|$ clearly satisfies (1) above.

To calculate $\mathbf{E}|G_{n,h}|$, Devroye and Reed used a ballot theorem. By the standard rotation argument, the probability that $S_h \leq t$ and additionally $S_i \geq i \cdot S_h/h$ for all $1 \leq i \leq h$, is $\Theta(\mathbf{P}\{S_h \leq t\}/h)$. By linearity of expectation and symmetry, the expected number of good nodes at depth $h$ is $2^h\mathbf{P}\{S_h \leq t\}/h$, so again by the properties of the gamma distribution, it is straightforward to find the smallest value $h'$ for which $\mathbf{E}|G_{n,h'}| \leq 1$ and to show that in fact

$\mathbf{E}|G_{n,h'}| = \Theta(1)$ – it turns out that $h'$ is $h^* - O(\log h^*)$. By a more formal version of the above discussion, they then showed that for any node $v$ at depth $h$ in $T_\infty^d$, $\mathbf{E}\{|G_{n,h}| | v \in G_{n,h}\} \leq 1 + O(h^2\mathbf{E}|G_{n,h}|)$, which has the flavour of condition (2), above. Combining this fact with a more sophisticated variant of (5.2) arising from an application of the second moment method, some well-known facts about the gamma distribution, and a standard amplification argument, was enough for them to show that $\mathbf{E}H_t^i \geq h' - O(\log h^*) = h^* - O(\log h^*)$.

In fact, it turns out that $\mathbf{E}H_t^i = h' + O(1)$. To show this, Reed (2003) exploited the properties of exponential random variables to prove a ballot theorem very similar to that of Section 2.5. This gave him rather precise control over the behavior of the random walk $S_1, \ldots, S_h$ and the probability it stays near its conditioned mean. In particular, he was able to show that a random walk with iid exponential steps that stays above its conditioned mean very likely stays "well above" its conditioned mean most of the time. In other words, if $v_h$ is good, then not only is $S_i \geq i \cdot S_h/h$ for all $i$, but in fact it is quite likely that $S_i \geq i \cdot S_h/h + f(i,h)$ for some function $f(i,h)$ that is fairly large when both $i$ and $h - i$ are much larger than 1 (i.e., $f(i,h)$ is fairly large except "at the ends" of the random walk $S_1, \ldots, S_h$). He is thereby able to strengthen the result of Devroye and Reed, proving that in fact $\mathbf{E}\{|G_{n,h}| | |G_{n,h}| \geq 1\} \leq 1 + O(\mathbf{E}|G_{n,h}|)$. Arguing as did Devroye and Reed but using this stronger bound, he shows that $\mathbf{E}H_t^i \geq h' - O(1)$. By an argument also based on the ballot theorem, which is essentially a careful application of several union bounds but whose detailed outline we postpone, he shows a matching upper bound and so proves that $\mathbf{E}H_t^i = h' + O(1)$.

The ballot theorem of Section 2.5 is the key new ingredient that allows us to use Reed's approach to study the heights and the minima of a much more general family of random trees and branching random walks.

## 5.1 Ideal Branching Random Walks

Suppose we are given a branching random walk with branching distribution $B$ and step size $E$. Suppose further that there is an integer $d \geq 2$ such that $B \leq d$, and additionally $E$ is non-negative. Let $T$ be the weighted tree corresponding to this BRW. There is a natural way to obtain $T$ as a subtree of $T_\infty^d$ (which we remind the reader is the infinite $d$-ary tree with root $r$). At each node $v$ of $T_\infty^d$ with children $w_1, \ldots, w_d$, we assign an independent copy $B_v$ of $B$. We label the edges $vw_1, \ldots, vw_B$ with independent copies $E_{vw_1}, \ldots, E_{vw_B}$ of $E$, and label the remaining edges with labels $E_{vw_{B+1}} = \ldots = E_{vw_d} = +\infty$. For a node $v$ of $T$, the *weight* of the path from $r$ to $v$ is the sum of the edge labels along this path. The subtree of $T_\infty^d$ containing all nodes connected to the root by a path of finite weight is distributed precisely as $T$. (We observe that $T_\infty^d$ with such weights is not necessarily a branching process, as the labels $(E_{vw_1}, \ldots, E_{vw_d})$ are possibly infinite, and, more importantly, are not necessarily independent – knowing $E_{vw_1} = \infty$, for example, gives us information about $B$ and thus about the probability that the other edge labels are finite.)

We now introduce the objects we study in this chapter, of which both the weighted tree $T_\infty^d$ of the previous paragraph, and the split trees we saw above, are examples. Suppose we are given a vector $\mathcal{E} = (E_1, \ldots, E_d)$ of possibly dependent random variables each distributed as some random variable $E$ that takes values in $[0, \infty]$ and that $\inf\{x \geq 0 \mid \mathbf{P}\{E \leq x\} > 0\} = 0$, i.e. that $E$ takes values arbitrarily close to zero. We note that these conditions on $E$ are equivalent to requiring that $E$ is bounded from below as we may then replace $E$ by the variable $E - \inf\{x \mid \mathbf{P}\{E \leq x\} > 0\}$.

We call a vector $\mathcal{E}$ satisfying these conditions an *ideal vector*. As above, we consider a copy of the infinite $d$-ary tree $T_\infty^d$ in which each node $v$ is assigned an independent copy $\mathcal{E}_v$ of $\mathcal{E}$ – we denote this tree $T_d(\mathcal{E})$, or $T(\mathcal{E})$ when $d$ is clear from context. For each node $v$, we associate the component random variables of $\mathcal{E}_v$ with the edges $vw_1, \ldots, vw_d$ from $v$ to its children. We will use both the notation $\mathcal{E}_v = (E_{v,1}, \ldots, E_{v,d})$ and $\mathcal{E}_v = (E_{vw_1}, \ldots, E_{vw_d})$.

We call $T(\mathcal{E})$ an *ideal branching random walk.* We note that every split tree may be viewed as an ideal branching random walk: given a split vector $\mathcal{X} = (X_1, \ldots, X_d)$, we may define an associated ideal vector $\mathcal{E} = (E_1, \ldots, E_d)$ by letting $E_i = -\log_d X_i$ and possibly applying an additive shift to the $E_i$. Furthermore, consider any BRW with branching distribution $B$ satisfying the deterministic bound $0 \leq B \leq d$ and with step size and step size $E$ that is bounded from below. We may view such a BRW as contained in ideal branching random walk, by letting $E_1, \ldots, E_B$ be independent copies of $E - E - \inf\{x \mid \mathbf{P}\{E \leq x\} > 0\}$, letting $E_{B+1}, \ldots, E_d$ all equal $+\infty$, taking a uniformly random permutation $\sigma$ of $1, \ldots, d$ and setting $\mathcal{E} = (E_{\sigma(1)}, \ldots, E_{\sigma(d)})$. (We remark that taking a random permutation ensures that the component random variables of $\mathcal{E}$ are identically distributed.) By analogy with branching random walks, we say that $T(\mathcal{E})$ *survives* if there is an infinite path $r = v_0, v_1, v_2, \ldots$ along which all edge labels are finite. We denote the event that $T(\mathcal{E})$ survives by $\mathcal{S}$, and say that $T(\mathcal{E})$ is *supercritical* if $\mathbf{P}\{E < \infty\} > 1/d$ (so if $T(\mathcal{E})$ is supercritical then $\mathbf{P}\{\mathcal{S}\} > 0$).

As we did for ordinary branching random walks, we assign each node $v$ of $T(\mathcal{E})$ a label $S_v$ which is the sum of the edge labels on the path from $r$ to $v$. Just as for branching random walks and for split trees, we may define the trees $\{T_t\}_{t\geq 0}$, $T_t$ containing all nodes $v$ with $S_v \leq t$; we let $H_t$ be the height of $T_t$. For $n = 1, 2, \ldots$, we define $M_n$ to be the minimum label of any node $v$ at depth $n$. Also just as for ordinary branching random walks, we have the relationships

$$M_n = \inf\{m \mid H_m \geq n\} \quad \text{and} \quad H_t = \max\{n \in \mathbb{N} \mid M_n \leq t\}, \tag{5.3}$$

for all $n = 1, 2, \ldots$ and for all $t \geq 0$. The aim of the remainder of the chapter is to pin down $\mathbf{E}M_n$ to within $O(1)$ for a broad class of ideal BRWs that in particular captures all supercritical branching random walks with bounded branching distribution and bounded steps, as well as all split trees. We say $T(\mathcal{E})$ has *bounded steps* if there is a constant $A > 0$ such that $E < A$ when $E$ is finite. In this chapter we prove:

**Theorem 62.** *Suppose we are given an supercritical ideal branching random walk $T(\mathcal{E})$ with bounded steps, where $\mathcal{E} = (E_1, \ldots, E_d)$ has component random variables distributed as $E$. Suppose additionally that $\mathbf{Var}\{E\} > 0$, and that $\mathbf{P}\{E = 0\} \neq 1/d$. Then there are $\tau^* \geq 0$, $\beta^* \geq 0$ depending only on $\mathcal{E}$ such that $\mathbf{E}\{M_n|\mathcal{S}\} = \tau^* n + \beta^* \ln n + O(1)$ for all integers $n > 0$. Furthermore, there are $c > 0$, $\delta > 0$ such that for all $x \geq 0$ and all $n = 1, 2, \ldots$, $\mathbf{P}\{|M_n - \tau^* n - \beta^* \ln n| \geq x|\mathcal{S}\} \leq ce^{-\delta x}$.*

To give an idea of how we will prove Theorem 62, we first consider the special case that $\mathbf{P}\{E = 0\} > 1/d$ and $E$ is never $\infty$ (so there is some constant $A$ for which $E \leq A$). In this case $\mathbf{P}\{S\} = 1$ so the conditioning in Theorem 62 vanishes. We consider the related branching process $T_0$ in which the set of children of a node is the set of its children in $T(\mathcal{E})$ for which the displacement is 0. Clearly, this walk survives with positive probability, so there is a positive probability $p_0$ that $M_n = 0$ for every $n$.

Suppose that we want to bound the probability that $M_n$ is greater than $x$, conditioned on survival. If $n$ is at most $x/A$ then every node at depth $n$ has label at most $nA \leq x$, so $\mathbf{P}\{M_n > x\} = 0$. for larger $n$, we first observe that for any node $v$ at depth $\lfloor x/A \rfloor$, the tree rooted at $v$ whose nodes are the descendants $w$ of $v$ with $S_w = S_v$ is distributed precisely as $T_0$; we temporarily denote this tree $T_0(v)$.

The tree $T_0(v)$ survives with probability $p_0$, and if $T_0(v)$ survives then in particular, there is a descendent $w$ of $v$ at depth $n$ in $T(\mathcal{E})$ for which $S_w = S_v$, so $M_n \leq S_w = S_v \leq x$. It follows that

$$\mathbf{P}\{M_n > x\} \leq \mathbf{P}\left\{ \bigcap_{v \text{ at depth } \lfloor x/A \rfloor} \{T_0(v) \text{ does not survive.}\} \right\}.$$

Since the subtrees rooted at distinct nodes at depth $\lfloor x/A \rfloor$ are independent and there are

$d^{\lfloor x/A \rfloor}$ such nodes, it follows that

$$\begin{aligned} \mathbf{P}\{M_n > x\} &\leq \prod_{v \text{ at depth } \lfloor x/A \rfloor} \mathbf{P}\{T_0(v) \text{ does not survive.}\} \\ &= (1-p_0)^{d^{\lfloor x/A \rfloor}}, \end{aligned} \tag{5.4}$$

which in particular proves the tail bound of Theorem 62 and also immediately implies that $\mathbf{E}M_n = O(1)$. The key to the above line of reasoning is the idea of analyzing the subtrees of $T(\mathcal{E})$ rooted at depth $\lfloor x/A \rfloor$ independently in order to strengthen our probability bound. We will hereafter refer to this technique as an *amplification* argument. McDiarmid (1995) uses this idea in much the same fashion as above in his analysis of the minima of branching random walks; it also plays a key role in both (Devroye and Reed, 1995) and (Reed, 2003).

When $\mathbf{P}\{E = 0\} > 1/d$ but $E$ is possibly infinite, we still have that $\mathbf{P}\{T_0 \text{ survives}\} = p_0$ for some $p_0 > 0$. For $x \geq 0$ and $n \geq x/A$, we temporarily let $N_x$ be the set of nodes at depth $\lfloor x/A \rfloor$ with finite labels (so with labels at most $A\lfloor x/A \rfloor \leq x$). Using an amplification argument just as we did in deriving (5.4) immediately yields that for any integer $c > 0$,

$$\mathbf{P}\{M_n > x \mid |N_x| = c\} \leq (1-p_0)^c. \tag{5.5}$$

So, to handle this case, we really need to analyze the distribution of the number of nodes at a given level of a supercritical branching process.

It turns out that there is $\epsilon > 0$ depending only on $E$ such that $\mathbf{P}\{0 < |N_x| < \epsilon x\}$ decreases exponentially in $x$. In Section 5.1.1, we will show how such a bound be easily proved by growing $T(\mathcal{E})$ using a slight variant of the breadth-first-search-based exploration process first seen in Chapter 3 and analyzing a corresponding random walk. (Such bounds for $|N_x|$ are well-known; other proofs can be found in, e.g., Athreya and Ney (1972) or McDiarmid (1995).) We will then use this bound and an easy amplification argument such as that sketched above to prove exponential tail bounds for $\mathbf{P}\{M_n \geq x \mid \mathcal{S}\}$ and thereby prove Theorem 62 in the case $\mathbf{P}\{E=0\} > 1/d$. Combined with the ballot theorem and a little more work, this line of argument will also yield lower tail bounds for $M_n$, which we will need for the case $\mathbf{P}\{E=0\} < 1/d$.

We remark that the case $\mathbf{P}\{E=0\} > 1/d$ of Theorem 62 is a result of Biggins (1976a). We have included a proof of this case as it is easier and gives a clear idea of the approach we will take in handling the case $\mathbf{P}\{E=0\} < 1/d$. In tackling the case $\mathbf{P}\{E=0\} < 1/d$, it will in fact be more convenient to study $H_t$ than to study $M_n$. We will show:

**Theorem 63.** *Suppose we are given an supercritical ideal branching random walk $T(\mathcal{E})$ with bounded steps, where $\mathcal{E} = (E_1, \ldots, E_d)$ has component random variables distributed as $E$. Suppose additionally that $\mathbf{Var}\{E\} > 0$, and that $\mathbf{P}\{E=0\} < 1/d$. Then there are $\tau > 0$, $\beta > 0$, and $A > 0$ depending only on $\mathcal{E}$ such that $\mathbf{P}\{|H_t - \tau t - \beta \ln t| \leq A\} = \Omega(1)$ for all $t \geq 1$.*

For any $\tau > 0$, $\beta > 0$, letting $f(t) = \tau t - \beta \ln t$, there is $a > 0$ such that $|f(t)/\tau + (\beta/\tau)\ln f(t) - t| \leq a$ as long as $t \geq 1$. Letting $f^*(t) = t/\tau + (\beta/\tau)t$, it follows from this fact and from Theorem 63 that $\mathbf{P}\left\{H_{f^*(n)+A+a} > n\right\} = \Omega(1)$ and $\mathbf{P}\left\{H_{f^*(n)-A-a} < n\right\} = \Omega(1)$. By the first relationship in (5.3), it follows that $\mathbf{P}\{M_n \leq f^*(n) + A + a\} = \Omega(1)$ and $\mathbf{P}\{M_n \geq f^*(n) - A - a\} = \Omega(1)$. Thus, if we can prove that when $\mathbf{P}\{E=0\} < 1/d$ there is *some* value $b_n$ for which we have exponential tail bounds for $\mathbf{P}\{|M_n - b_n| \geq x|\mathcal{S}\}$, we must have $b_n = \mathbf{E}\{M_n|\mathcal{S}\} + O(1)$. Combining such tail bounds with Theorem 63, it follows that we must also have $b_n = f^*(n) + O(1)$, which proves Theorem 62 in the case that $\mathbf{P}\{E=0\} < 1/d$, with $\tau^* = 1/\tau$, $\beta^* = \beta/\tau$.

From Theorems 62 and 63 and from (5.3), we also immediately obtain

**Corollary 64.** *Under the conditions of Theorem 63, $\mathbf{E}\{H_t|\mathcal{S}\} = \tau t - \beta \ln t + O(1)$, and there are $c > 0$, $\delta > 0$ such that for all $x \geq 0$ and all $n = 1, 2, \ldots$, $\mathbf{P}\{|H_n - \mathbf{E}H_n| \geq x \mid \mathcal{S}\} \leq$*

$ce^{-\delta x}$.

We note that a recent series of papers (Broutin and Devroye, 2005; Broutin et al., submitted, 2006) has established the asymptotics of the first order term of the height of a very general family of trees; their techniques can in particular be used without modification to prove that $H_n/n \to \tau$ for ideal branching random walks. Broutin et. al. also treat the case where the edges of $T_\infty^d$ have *lengths* in addition to the weights given above, and prove a law of large numbers for the resulting "weighted height".

Before proceeding to the proofs of our results, we briefly discuss a probabilistic structure closely related to branching random walks: *branching Brownian motion* (BBM). In this model, an initial particle starts at position 0 on the real line and begins a standard Brownian motion. The particle decays according to an exponential mean 1 clock; when the clock goes off, the particle splits into two, each of which continues an independent Brownian motion and each of which independently decays according to an exponential mean 1 clock. This process is continued forever.

The pride of place among results on branching Brownian motion goes to Bramson (1978a, 1986), who studied the maximum position $M_t$ of any particle at time $t$. Bramson first showed that $M_t$, adjusted by its median $m_t$, converges in distribution to a certain "travelling wave", which is a distribution formally defined as one of a family of solutions to a certain partial differential equation called the KPP equation (see Bachmann (1998) for details). He then found the value of $m_t$, showing that there is a constant $c$ such that

$$m_t = \sqrt{2}t - 3\ln t/2 + c + o(1).$$

The similarity of this equation to that found by Reed for the expected height of random binary search trees and to that appearing in Theorem 62 is too great to be ignored. In fact, the lines of argument in Bramson (1978a) and in Reed (2003) are strikingly similar. Reed's argument is essentially a discrete-time analog of Bramson's argument, which proceeds by finding precise estimates for the behavior of a Brownian motion conditioned on its value at time $t$, and in particular in bounding the probability that such a Brownian motion strays far from its conditioned mean. Perhaps surprisingly, though the overall structure of the arguments is very similar, they differ in almost every detail, and there is no obvious way to move directly from Bramson's results to discrete equivalents. Indeed, Biggins (1997) noted that though it was widely expected that analogous results to Bramson's should hold for branching random walks, at that time there had been essentially no progress on proving such results. As with much of the theory of random walks, it seems that the treatment of discrete and continuous time differs substantially even when the resulting facts are very similar in nature.

In Section 5.1.1, we prove Theorem 62 in the case that $\mathbf{P}\{E=0\} > 1/d$. In Section 5.2, we give a sketch of some of the key ideas behind our proof of Theorem 63 that fleshes out the high-level discussion of Devroye and Reed's results given above. In Section 5.3 we develop two key lemmas based on the ballot theorems of Chapter 2. Finally, in Sections 5.4 and 5.5, respectively, we prove lower tail bounds and upper tail bounds on $H_n$ that together prove Theorem 63 and thereby establish Theorem 62 in the case $\mathbf{P}\{E=0\} < 1/d$. First, however, we fix a few aspects of the notation that will recur in the remainder of the section and prove Theorem 62 in the case that $\mathbf{P}\{E=0\} > 1/d$.

## Notation

We recall that our object of study is a copy of the infinite $d$-ary tree $T_\infty = T_\infty^d$ with root $r$. We have assigned each edge $e = vw$ a label $E_{vw}$; we additionally assign each node $v$ a label $S_v$ that is the sum of the edge labels on the path from $r$ to $v$. Let $X_\infty^h$ be the nodes of $T_\infty$ at depth $h$, and let $N_{n,h}$ be the set of nodes of $T_n$ at depth $h$ (we remark that for the remainder of the section we use the index $n$ instead of $t$ – so $n$ is not necessarily integer). We observe that the height $H_n$ of $T_n$ is just the largest $h$ for which $N_{n,h} \neq \emptyset$, and that for any integer $h \geq 0$, $M_h$ is the smallest $n \geq 0$ for which $N_{n,h} \neq \emptyset$

For the sake of our analysis, it will be useful to fix a distinguished infinite path $P$ in $T_\infty$ with nodes $r = v_0, v_1, \ldots$. We denote the labels along this path by $E_1, E_2, \ldots$; recall that these labels are iid and distributed as some random variable $E$, $0 \leq E \leq \infty$. We further let $S_i = S_{v_i} = E_1 + \ldots + E_i$. We remind the reader that $v_i \in T_n$ precisely if $S_i \leq n$.

Each node $v_i$ has one child $v_{i+1}$ in $P$ – let its other children be called $v_{i+1}^{(1)}, \ldots, v_{i+1}^{(d-1)}$. Denote by $T_\infty^{i,j}$ the subtree of $T_\infty$ rooted at $v_i^{(j)}$ (note that $T_\infty^{i,j}$ is isomorphic to $T_\infty$). Let $T_n^{i,j} = T_\infty^{i,j} \cap T_n$ (which may be empty), and let $N_{n,h}^{i,j}$ be the set of nodes of $N_{n,h}$ that are contained in $T_n^{i,j}$.

### 5.1.1 Bounding $M_n$ when $\mathbf{P}\{E = 0\} > 1/d$

As discussed above, we let $T_0$ be the subtree of $T(\mathcal{E})$ consisting of all descendants of the root for which the displacement is 0. This tree is distributed as a supercritical branching process and therefore has some positive probability of survival $p_0 > 0$. We let the maximum step size be $A$ and observe that for any $x \geq 0$ and $n \geq x/A$, if $T(\mathcal{E})$ contains $k$ nodes with finite label at depth $\lfloor x/A \rfloor$ (i.e. $|N_{x,\lfloor x/A \rfloor}| = k$) then for $n \geq x/A$, the probability that $M_n$ is greater than $x$ is at most $p_0^k$, so we really want to study the distribution of the size of $N_{x,\lfloor x/A \rfloor}$ as $x$ grows.

The set $N_{x,\lfloor x/A \rfloor}$ is just the set of nodes with finite labels in $T(\mathcal{E})$, and is therefore distributed as the set of vertices at depth $\lfloor x/A \rfloor$ of a supercritical branching process. We will prove using the random walk-based techniques developed in Chapter 3 that for such a set, there are $\epsilon > 0$,

$c_0 > 0$, and $\delta_0 > 0$ such that for all $x \geq 0$,

$$\mathbf{P}\left\{0 < N_{x,\lfloor x/A \rfloor} \leq \epsilon x\right\} \leq c_0 e^{-\delta_0 x}. \tag{5.6}$$

(As noted in the discussion proceeding the statement of Theorem 62, this is a classic result in the theory of branching processes that can be found in Athreya and Ney (1972) or in McDiarmid (1995).) Since if $\mathcal{S}$ occurs then in particular $|N_{x,\lfloor x/A \rfloor}| > 0$ for all $x$, it follows that for $x$ large enough, for $n \geq x/A$,

$$\begin{aligned} \mathbf{P}\{M_n > x, \mathcal{S}\} &\leq \mathbf{P}\left\{0 < |N_{x,\lfloor x/A \rfloor}| \leq \epsilon x\right\} + \mathbf{P}\left\{M_n > x \mid |N_{0,\lfloor x/A \rfloor}| > \epsilon x\right\} \\ &\leq c_0 e^{-\delta_0 x} + p_0^{\epsilon x} \leq c_1 e^{-\delta x}, \end{aligned} \tag{5.7}$$

for some $c_1 > 0$ and $\delta > 0$. Finally, since the ideal BRW is supercritical, we have $\mathbf{P}\{\mathcal{S}\} = p > 0$. Letting $c = c_1/p$, by the previous fact and by (5.7) we thus have $\mathbf{P}\{M_n > x|\mathcal{S}\} \leq ce^{-\delta x}$ for $n \geq \lfloor x/A \rfloor$. For $n < x/A$, if $\mathcal{S}$ occurs then $M_n \leq An < x$, so $\mathbf{P}\{M_n > x|\mathcal{S}\} = 0$. This proves the exponential tail bounds of Theorem 62 and also proves that $\mathbf{E}\{M_n|\mathcal{S}\} = O(1)$, which completes the proof of Theorem 62 in the case $\mathbf{P}\{E = 0\} > 1/d$.

To prove (5.6), we explore the subtree of $T(\mathcal{E})$ with finite labels using the breadth-first-search exploration process of Section 3.2. Namely, we maintain a set (which we now think of as a *queue*) of open vertices $\mathcal{O}_i$. We begin with $\mathcal{O}_1 = \{r\}$, the root of $T(\mathcal{E})$. A step of the exploration process consists of removing a vertex $v_{i+1}$ from the front of the queue $\mathcal{O}_i$, letting $C_{i+1}$ be the set of children $w$ of $v_{i+1}$ for which the edge label $E_{v_{i+1}w}$ is finite, and adding the elements of $C_{i+1}$ to the back of the queue to form $\mathcal{O}_{i+1}$. If ever $\mathcal{O}_i = \emptyset$ then we set $\mathcal{O}_j = \emptyset$ for all $j \geq i$.

We now point out several properties of this exploration process which follow immediately from its definition. First of all, this process explores $T(\mathcal{E})$ level-by-level; at any given time $i$, there is an integer $h \geq 0$ for which $\mathcal{O}_i$ is entirely contained within $X_\infty^h \cup X_\infty^{h+1}$. Furthermore, as a step of the exploration process can only increase the depth of the deepest explored node by 1, if $\mathcal{O}_i$ contains an element of $X_\infty^h$ then $i \geq h$. Finally, for a given integer $h > 0$, if $\mathcal{O}_i$ contains elements of $X_\infty^{h-1}$ and $\mathcal{O}_{i+1}$ does not, then $\mathcal{O}_{i+1}$ is entirely contained within $X_\infty^h$; more strongly, in this case $\mathcal{O}_{i+1}$ is precisely $N_{Ah,h}$. In particular, it follows from these facts that for a given $x$, if $|N_{x,\lfloor x/A \rfloor}| \leq k$ then there is $i \geq \lfloor x/A \rfloor$ for which $|\mathcal{O}_i| \leq k$. As a consequence we can bound the size of $N_{x,\lfloor x/A \rfloor}$ from below by studying the behavior of $|\mathcal{O}_i|$ as $i$ grows.

For $i = 0, 1, 2, \ldots$, we let $S_i = |\mathcal{O}_i|$; the sequence $S_0, S_1, \ldots$ forms a random walk. Let $X_0 = 0$; for $i > 0$, if $\mathcal{O}_i > 0$ then let $X_{i+1} = C_{i+1} - 1$ and if $\mathcal{O}_i = 0$ then let $X_i = 0$. With the random variables $X_0, X_1, \ldots$ so defined, $S_i = \sum_{j=0}^{i} X_j$ for all $i \geq 0$. We remark that given that $\mathcal{O}_i > 0$, $X_{i+1}$ is an integer-valued random variable taking values in $\{-1, 0, 1, \ldots, d-1\}$ (with distribution $B$, say). We couple $S$ to a random walk $S'$ for which $S'_0 = 1$ and, for which, for $i \geq 0$, $S'_{i+1} = S'_i + X'_{i+1}$ , where $X'_1, X'_2, \ldots$ are independent random variables distributed as $B$. (So when $S_i > 0$, $X'_{i+1} = X_{i+1}$ and $X'_{i+1} + 1$ is precisely the number of children $w$ of $v_{i+1}$ for which $E_{v_{i+1}w}$ is finite.) We emphasize that until $S$ hits zero, $S_i$ and $S'_i$ are identical, and we let $t_0$ be the first time that $S_{t_0} = 0$, or let $t_0 = \infty$ if $S$ is never zero. Using this coupling and the facts from the previous paragraph, for any $x \geq 0$ and $v > 0$, we have

$$\begin{aligned} \mathbf{P}\left\{0 < N_{x,\lfloor x/A \rfloor} < v\right\} &\leq \mathbf{P}\left\{\exists i \geq \left\lfloor \frac{x}{A} \right\rfloor \text{ s.t. } 0 < S_i < v\right\} \\ &= \mathbf{P}\left\{\exists i \geq \left\lfloor \frac{x}{A} \right\rfloor \text{ s.t. } t_0 > i \text{ and } 0 < S'_i < v\right\} \\ &\leq \mathbf{P}\left\{\exists i \geq \left\lfloor \frac{x}{A} \right\rfloor \text{ s.t. } 0 < S'_i < v\right\}. \end{aligned} \tag{5.8}$$

The steps of the random walk $S'$ have positive mean – we choose $0 < \epsilon < 1/2A$ so that $\mathbf{E}B > 2\epsilon A$ and thus $\mathbf{E}S'_i > 1 + (2\epsilon A)i$. Letting $v = \epsilon x$, we in particular have $\mathbf{E}S'_{\lfloor x/A \rfloor} \geq 1 + (2\epsilon A)\lfloor x/A \rfloor \geq 2\epsilon x$, so if $i = \lfloor x/A \rfloor + j$ then $\mathbf{E}S'_i \geq 2\epsilon x + (2\epsilon A)j$. Combined with (5.8), this yields

$$\begin{aligned}\mathbf{P}\left\{0 < N_{x,\lfloor x/A\rfloor} < v\right\} &\le \mathbf{P}\left\{\exists j \ge 0 \text{ s.t. } S'_{\lfloor x/A\rfloor+j} \le \mathbf{E}S'_{\lfloor x/A\rfloor+j} - \epsilon x - (2\epsilon A)j\right\} \\ &\le \sum_{j=0}^{\infty}\mathbf{P}\left\{S'_{\lfloor x/A\rfloor+j} \le \mathbf{E}S'_{\lfloor x/A\rfloor+j} - \epsilon(x+2Aj)\right\}. \end{aligned} \tag{5.9}$$

Since $S'$ has bounded step size ($-1 \le X_i \le d-1$ for all $i$), we can easily derive (5.6) from (5.9) by, e.g, Chernoff bounds.

In fact, a much stronger statement than (5.6) is true; essentially, we may replace the probability $\mathbf{P}\left\{0 < N_{x,\lfloor x/A\rfloor} \le \epsilon x\right\}$ by $\mathbf{P}\left\{0 < N_{x,\lfloor x/A\rfloor} \le (1+\epsilon)^{\lfloor x/A\rfloor}\right\}$ in (5.6) and the inequality remains valid. This has been proved by McDiarmid (1995), who was then able to use this stronger inequality to prove exponential tail bounds for $\mathbf{P}\{M_n < x|\mathcal{S}\}$ when $x < \mathbf{E}\{M_n|\mathcal{S}\}$ (in the same work he pinned down $\mathbf{E}\{M_n|\mathcal{S}\}$ to within $O(\log n)$ in the case $\mathbf{P}\{E=0\} < 1/d$). Of course, exponential lower tail bounds hold trivially when $\mathbf{E}\{M_n|\mathcal{S}\} = O(1)$, which is why we did not need to prove such bounds in the case $\mathbf{P}\{E=0\} > 1/d$. We *will* need lower tail bounds when considering the case $\mathbf{P}\{E=0\} < 1/d$; with more work, we could derive such bounds using the above random walk approach. We decline to do so, instead using McDiarmid's result, which we now state:

**Lemma 65** (McDiarmid (1995)). *Given an supercritical ideal branching random walk $T(\mathcal{E})$ with bounded steps, for all $n = 1, 2, \ldots$, let $b_n = \inf\{t : \mathbf{P}\{M_n \le t\} \ge \mathbf{P}\{\mathcal{S}\}/2\}$. Then there are $c > 0, \delta > 0$ such that for all $x \ge 0$ and all $n$, $\mathbf{P}\{|M_n - b_n| > x \mid \mathcal{S}\} < ce^{-\delta x}$.*

(McDiarmid originally stated this result for non-negative branching random walks with bounded steps, but his proofs apply without change to ideal branching random walks. See also Athreya (1994) for related results.) As discussed just after the statement of Theorem 63, combining Theorem 63 with tail bounds such as those of Lemma 65 immediately proves both the case $\mathbf{P}\{E=0\} < 1/d$ of Theorem 62 and, as a consequence, Corollary 64. The remainder of the chapter is therefore devoted to proving Theorem 63.

## 5.2 A few key ideas

For the remainder of the chapter, we will be primarily concerned with analyzing the behavior of the random walk $S_1, S_2, \ldots$ given by the labels of the distinguished path $v_1, v_2, \ldots$ – from studying this random walk we will prove both the upper and the lower bounds of Theorem 63. It turns out that it will usually be more convenient to study the *negative* random walk $S'$ given by $S'_i = -S_i = -S_{v_i}$. (For general nodes $v \in T_\infty$, we likewise let $S'_v = -S_v$.)

For a given $h$, $\mathbf{P}\{v_h \in T_n\}$ is just the probability $S'_h \geq -n$. By linearity of expectation and by symmetry, the *expected* number of nodes of depth $h$ that are in $T_n$ is $d^h$ times this probability. As we discussed when summarizing the results of Devroye and Reed, it seems a likely bet that the expected height of $T_n$ will be close to the smallest value $h^*$ such that $\mathbf{P}\{v_{h^*} \in T_n\} \leq 1/d^{h^*}$, i.e., the first level for which $\mathbf{E}N_{n,h} \leq 1$. We call the value $h^*$ the *breakpoint*.

Our guess that the expected height is near the breakpoint is not perfectly accurate, but it is not far off; as we shall see, the expected height is $\Theta(n)$, and the breakpoint $h^*$ is within $O(\log n)$ of the expected height. Furthermore, this analysis will highlight several of the key ideas behind the stronger results to follow. For this reason, we devote the remainder of this section to an exploration of the breakpoint and what it can tell us about the height of $T_n$.

### 5.2.1 The breakpoint

The object of this section is to pin down the value $h^*$ and, more importantly, to bound the probability $\mathbf{P}\{S'_h \geq c - n\}$ for "small" $c$ and for $h$ near $h^*$. This information is a straightforward consequence of the asymptotics for large deviations appearing in Appendix A. (For the reader who recalls the use of these results in Section 2.5, what follows should be somewhat familiar.) We wish to find the smallest $h$ for which $\mathbf{P}\{S'_h \geq -n\} \leq 1/d^h$. The results of the appendix yield:

**Lemma 66.** *Let $h^* = h^*(n)$ be the smallest integer for which $\mathbf{P}\{S'_{h^*} \geq -n\} \leq 1/d^{h^*}$. Then there are constants $\tau > 0$, $\alpha > 0$, and $\gamma > 0$ such that*

*(a) $h^* = \tau n - \frac{\ln n}{2\alpha} + O(1)$ and $\mathbf{P}\{S'_{h^*} \geq -n\} = \Theta\left(\frac{1}{d^{h^*}}\right)$, and*

*(b) if $c, c' = o(\sqrt{n})$ then*

$$\mathbf{P}\left\{S'_{h^*+c'} \geq c - n\right\} = \Theta\left(\frac{e^{-\alpha c' - \gamma c}}{d^{h^*+c'}}\right),$$

*where for any $g(n)$ tending to infinity with $n$, the order notation $\Theta$ is uniform over $|c|, |c'| \leq \sqrt{n}/g(n)$.*

*Proof.* We consider the random walk $\hat{S}_0, \hat{S}_i, \ldots, \hat{S}_h$ given by $\hat{S}_i = S'_i/\mathbf{E}E = \sum_{j=1}^{i}(-E_i/\mathbf{E}E)$; this is a non-positive random walk with mean $-1$, so the results of Appendix A apply to $\hat{S}$. Since $\mathbf{P}\{-E_i/\mathbf{E}E = 0\} < 1/d$, by Corollary 83 there is a real number $t_0 > 0$ and a function $\Lambda(t)$ that is infinitely differentiable in an open neighbourhood of $t_0$ for which $t_0\Lambda'(t_0) - \Lambda(t_0)$, which we denote $f(t_0)$, equals $\ln d$. Furthermore, by Corollary 84, letting $\tau = -(\Lambda'(t_0)\mathbf{E}E)^{-1}$, $\alpha = -\Lambda(t_0)$, and $c_0 = t\sqrt{2\pi\Lambda''(t)}$, it is the case that $\tau > 0$, $\alpha > 0$, $c_0 > 0$, and letting $h = \tau n$, letting $a = -\lceil \frac{\ln n}{2\alpha} \rceil$, we have

$$\begin{aligned}
\mathbf{P}\left\{S'_{h+a} \geq -n\right\} &= \mathbf{P}\left\{\hat{S}_{h+a} \geq -n/\mathbf{E}E\right\} \\
&= (1+o(1))\frac{c_0 e^{-\alpha a - hf(t_0)}}{\sqrt{h+a}} = \Theta\left(\frac{c_0 e^{-\ln n/2}}{d^h\sqrt{h}}\right) = \Theta\left(\frac{1}{d^h}\right).
\end{aligned}$$

Corollary 84 also yields that increasing (resp. decreasing) $a$ by 1 decreases (resp. increases) $\mathbf{P}\left\{S'_{h+a} \geq -n\right\}$ by a constant factor; it follows that $h^*$ is within $O(1)$ of $h+a$ and that $\mathbf{P}\left\{S'_{h+a} \geq -n\right\} = \Theta(1/d^h)$, i.e., (a) holds. Finally, letting $\gamma = t_0$, statement (b) and the comment that follows are a direct consequence of Corollary 85 applied with this specific choice of $h^*$. □

For our purposes, the main point of the uniformity result in Lemma 66 is that it allows us to "take the order notation outside the sum"; we will often do so to avoid excessive definition of arbitrary constants, and will not always remark that we are using part of Lemma 66.

From this lemma, we can immediately derive upper tail bounds on $\mathbf{P}\{H_n \geq h^* + i\}$, which by symmetry and a union bound is at most $d^{h^*+i}\mathbf{P}\left\{S'_{h+i} \geq -n\right\}$. In particular, it follows from Lemma 66 (b) that there is $\epsilon > 0$ such that $\mathbf{P}\{H_n \geq h^* + i\} = O((1-\epsilon)^i)$ for $i = o(\sqrt{n})$. It is tempting to try to bound the expected value from below using Lemma 66 as well. Lemma 66 tells us that the expected number of nodes of $T_n$ at level $h^* - i$ is $\Omega(e^{\alpha i})$, so is growing exponentially fast as $i$ increases (and $i = o(\sqrt{n})$). This fact on its own is not enough to bound $\mathbf{E}H_n$ from below, however. Indeed, it turns out that $\mathbf{E}\{H_n\} - h^*$ is *not* $O(1)$.

We know, however, that if we can find some integer $h'$ within $o(\sqrt{n})$ of $h^*$ for which $\mathbf{P}\{|N_{n,h'}| \neq 0\} = \Omega(1)$, then by Lemma 65 we have exponential tail bounds on $\mathbf{P}\{|N_{n,h'-i}| = 0\}$. To find such a value $h'$, for each integer $h$ with $h - h^* = o(\sqrt{n})$, we will define a set of "good" nodes $G_{n,h}$ that is a subset of $N_{n,h}$. We mentioned such "good" nodes in the introduction to this chapter, in our earlier discussion of Devroye and Reed's proof technique. Essentially, a good node is a node $v$ of $N_{n,h}$ for which the random walk ending at $v$ stays below its conditioned mean (in fact, we will impose one other, rather insignificant condition as part of the definition of good nodes). The definition of good will ensure that if $\mathbf{E}|G_{n,h}|$ is large then $\mathbf{P}\{|G_{n,h}| \geq 1\}$ is $\Omega(1)$, so we will be able to find lower bounds on $\mathbf{E}H_n$ by studying $\mathbf{E}|G_{n,h}|$. This is the subject of the next section.

### 5.2.2 Good nodes

Recall that we have fixed the infinite path $v_0, v_1, \ldots$ in $T_\infty$, and that these vertices have received labels $S'_0, S'_1, \ldots$. For $h \le h^*$ for which $h - h^* = o(\sqrt{n})$, we say that node $v_h$ is a *good* node of $T_n$ if the following properties hold:

(G1) $S'_h \ge -n$, (so $v_h \in T_n$),

(G2) $S'_h \le 2 - n$, and

(G3) $S'_i \le i \cdot S'_h / h$ for all $0 < i < h$; adopting the language of Section 2.5, we will say $S'_h$ *stays below 0* and write $S'_h$ *bel* 0 for this event.

We extend this definition to all nodes $x$ at depth $h$ by symmetry, and denote the set of good nodes of $T_n$ at depth $h$ by $G_{n,h}$. The key to this section, and indeed to the whole chapter, is an analysis of $G_{n,h}$ along the lines sketched in the opening paragraph of this section.

Condition (G2) says that $S'_h$ is not too large. By Lemma 66 (b), the probability that $S'_h \le 2 - n$ given that $S_h \ge -n$ is $\Theta(1)$. By replacing the (G2) by the condition $S_h \le C - n$ for some large $C$, we could make this probability arbitrarily close to 1. Intuitively, therefore, we can think that the "typical" node in $T_n$ at depth $h$ satisfies something like (G2).

Condition (G3) is the key to the definition. In the language of random walks, it says the random walk $S'$ is staying below its (conditional) mean up to time $h$. In the current setting, it states that along the path to $v_h$, the labels $S'_i$ are decreasing "at least as fast as they should". Because these labels are decreasing quickly, it will follow that knowing that $v_h$ is in $T_n$ does not increase the expected number of nodes of $T_n$ in the subtrees hanging off the path $v_0 v_1 \ldots v_h$ by very much. (G3) is precisely the sort of condition we studied in detail in Section 2.5; we saw there that $\mathbf{P}\{S'_h \text{ bel } 0 | S'_h\} = \Theta(1/h)$. Combined with our discussion of

(G2), it follows that $\mathbf{E}|G_{n,h}| = \Theta(\mathbf{E}|N_{n,h}|/h)$. We will eventually show that if $\mathbf{E}|G_{n,h}| = \Omega(1)$ then $\mathbf{P}\{|G_{n,h}| > 0\} = \Omega(1)$. As a warmup, we prove:

**Fact 67.** *For $h < h^*$ for which $h^* - h = o(\sqrt{n})$, if $\mathbf{E}|G_{n,h}| = \Omega(1/h^2)$ then $\mathbf{P}\{|G_{n,h}| \geq 1\} = \Omega(1/h^2) = \Omega(1/n^2)$.*

*Proof of Fact 67.* To prove a lower bound on $\mathbf{P}\{|G_{n,h}| \geq 1\}$, we would like to use (5.2); however, it turns out to be much easier to calculate $\mathbf{E}\{|G_{n,h}| \mid v_h \in G_{n,h}\}$ than to calculate $\mathbf{E}\{|G_{n,h}| \mid |G_{n,h}| \geq 1\}$. As a consequence, instead of (5.2) we use a version of the second moment method often called the Erdős-Chung inequality (see Chung and Erdös (1952) and also Devroye and Reed (1995) and Alon and Spencer (2000), Chapter 2), which in our setting can be stated as follows: for any random set $S \subseteq X_\infty^h$,

$$\mathbf{P}\{|S| \geq 1\} \geq \frac{\mathbf{E}|S|}{1 + \sup_{v \in X_\infty^h} \mathbf{E}\{|S| \mid v \in S\}}, \tag{5.10}$$

for any $h = 1, 2, \ldots$ (we will apply (5.10) both here and later in the chapter). Let $S = G_{n,h}$. By the symmetry of the ideal vector $\mathcal{E}$, $\mathbf{E}\{|G_{n,h}| \mid v \in G_{n,h}\}$ is identical for all $v \in X_\infty^h$, so we may replace $\sup_{v \in X_\infty^h} \mathbf{E}\{|G_{n,h}| \mid v \in G_{n,h}\}$ by $\mathbf{E}\{|G_{n,h}| \mid v_h \in G_{n,h}\}$ and obtain

$$\mathbf{P}\{|G_{n,h}| \geq 1\} \quad \geq \quad \frac{\mathbf{E}|G_{n,h}|}{1 + \mathbf{E}\{|G_{n,h}| \mid v_h \in G_{n,h}\}}. \tag{5.11}$$

It follows that to prove Fact 67, it suffices to show that $\mathbf{E}\{|G_{n,h}| \mid v_h \text{ is good}\}$ is $O(h^2\mathbf{E}|G_{n,h}|)$. Recalling the definitions from the beginning of Section 5.2, we have

$$\begin{aligned}
\mathbf{E}\{|G_{n,h}| \mid v_h \text{ is good}\} &\leq \mathbf{E}\{|N_{n,h}| \mid v_h \text{ is good}\} \\
&\leq 1 + \sum_{i=0}^{h-1}\sum_{j=1}^{d-1} \mathbf{E}\{|N_{n,h}^{i,j}| \mid v_h \text{ is good}\} \\
&= 1 + \sum_{i=0}^{h-1} (d-1)\mathbf{E}\{|N_{n,h}^{i,1}| \mid v_h \text{ is good}\}. \qquad (5.12)
\end{aligned}$$

We write $h = h^* - c$, where $0 < c = o(\sqrt{n})$ by our choice of $h$. Since $v_h$ is good, for $0 \le i < h$, $S'_i \le (2-n)(i/h) \le 2 - ni/h$. It follows that for such $i$, $N_{n,h}^{i+1,1}$ is at most the number of nodes $x$ in $X_\infty^h$ that are in $T_n^{i+1,1}$ and for which the sum $S'_{x,h-i}$ of the negatives of the edge labels on the path from $v_{i+1,1}$ to $x$ is at least $-n(h-i)/h - 2$. Letting $n_i = n(h-i)/h$, we have that

$$\begin{aligned} h - i &= \frac{hn_i}{n} = \frac{(h^*-c)n_i}{n} = \frac{h^* n_i}{n} - \frac{cn_i}{n} \\ &= \tau n_i - \frac{n_i \ln n}{2n\alpha} + O(1) - \frac{c(h-i)}{h} \\ &\ge h^*(n_i) - \frac{c(h-i)}{h} + O(1). \end{aligned} \tag{5.13}$$

Next, let $c'$ be such that $h - i = h^*(n_i) - c'$. By (5.13), we thus have $c' = h - h^*(n_i) - i \ge O(1) - c(h-i)/h$. By Lemma 66 (b), therefore,

$$\mathbf{P}\left\{S'_{x,h-i} \ge -n_i - 2\right\} = \mathbf{P}\left\{S_{x,h^*-c^*} \ge -n_i - 2\right\} =' O\left(\frac{e^{\alpha c(h-i)/h}}{d^{h-i}}\right),$$

so by linearity of expectation,

$$\mathbf{E}\left\{|N_{n,h}^{i+1,1}| \mid v_h \text{ is good}\right\} = O(e^{c\alpha(h-i)/h}) = O(e^{c\alpha})$$

Plugging this into (5.12), yields $\mathbf{E}\{|G_{n,h}| \mid v_h \text{ is good}\} = 1 + O(he^{c\alpha})$. Since $\mathbf{E}|G_{n,h}| = \Theta(\mathbf{E}|N_{n,h}|/h)$, it follows from Lemma 66 (b) that $\mathbf{E}|G_{n,h}| = (1+o(1))e^{c\alpha}/h$, so if $\mathbf{E}|G_{n,h}| = \Omega(1/h^2)$ then $\mathbf{E}\{|G_{n,h}| \mid v_h \in G_{n,h}\} = O(h^2 \mathbf{E}|G_{n,h}|)$ by (5.11), as desired. □

By Fact 67 and a standard amplification argument, we could show that $\mathbf{P}\{|N_{n,h^*-a\ln n}| \ge 1\} = \Omega(1)$ for some $a = O(1)$ and thereby show that $\mathbf{E}H_n \ge h^* - a\ln n$; we will not do so as we are headed towards stronger bounds. It turns out that the expected height of $T_n$ is within $O(1)$ of the depth at which we expect $\Theta(1)$ *good* nodes of $T_n$. The key to proving this result is to strengthen Fact 67. We will eventually show that for values of $h$ for which

$h^* - h = o(\sqrt{n})$, $\mathbf{E}\{|G_{n,h}| \mid v_h \in G_{n,h}\} = \mathbf{E}|G_{n,h}| + O(1)$, and exhibit an integer $h'$ in this range which we can write as $h' = \tau n - \beta \ln n + O(1)$ for some constant $\tau > 0$, $\beta > 0$, and for which $\mathbf{E}\{|G_{n,h'}|\} = \Omega(1)$. This immediately implies by (5.10) and by symmetry that $\mathbf{P}\{|G_{n,h'}| > 0\} = \Omega(1)$. Since $G_{n,h} \subseteq N_{n,h}$, it follows that $\mathbf{P}\{|N_{n,h'}| > 0\} = \Omega(1)$, so $\mathbf{P}\{H_n \geq h'\} = \Omega(1)$, which proves the lower half of the bound of Theorem 63.

To prove that $\mathbf{E}\{|G_{n,h}| \mid v_h \in G_{n,h}\} = \mathbf{E}|G_{n,h}| + O(1)$, we will need to better understand the effect of conditioning on the event $\{v_h \text{ is good}\}$, and in particular what this conditioning implies about the difference of $S'_i$ from its conditioned mean, for $1 \leq i \leq h$. We will tackle this problem with the aid of the conditional ballot theorem results of Section 2.5. Proving that $\mathbf{P}\{|N_{n,h'+i}| \neq 0\}$ decays rapidly with $i$, and thereby proving the upper half of the bound of Theorem 63, will also rely crucially on the conditional ballot theorem. We therefore devote the next section to collecting the information about the conditional behavior of the random walk $S'_1, \ldots, S'_h$ that we will need to accomplish the aims of this chapter.

## 5.3 Staying low

The aim of this section is to gather two results about the conditional behavior of the random walk $S'_1, \ldots, S'_h$ – these results address the difference between $S'_i$ and its conditional mean (conditioned on $S'_h$), for $1 \leq i \leq h$. The first of these bounds we already proved in Section 2.5; the second is a straightforward consequence of the first. We remind the reader of some terminology from Section 2.5: we say that $S'_h$ *stays below* $a$ if, for all $0 < i < h$, $S'_i < S'_h(i/h) + a$. $E$ is a *lattice random variable* if there is $r \neq 0$ for which $rE$ is integer; otherwise it is *non-lattice*. The following is a slight weakening of Theorem 36, restated in terms of the random walk $S'$.

**Lemma 68.** *For all constants $c > 0$, $0 < \epsilon < \mathbf{E}E$ and all $a, r$ with $1 \leq a = O(h^{1/5})$ and*

$r = O(h^{1/5})$, *if $E$ is non-lattice then*

$$\mathbf{P}\left\{S'_h \textit{ bel } a \mid -\epsilon h + r \le S'_h \le -\epsilon h + r + c\right\} = O\left(\frac{a^5}{h}\right).$$

*Furthermore, if $E$ is lattice then the above equation holds for all $c$ for which $cE$ is integer.*

For the remainder of the chapter, we assume $E$ is non-lattice to avoid unedifying case analysis; all results hold in the lattice case with virtually identical proofs. Using this lemma, we can bound the conditional probability that $S'_h$ spends much time near its average, given that it stays below its average. For a given $a > 0$, let $B_a$ be the event that there is some $k$ for which $a^{33} \le k \le h - a^{33}$ and for which $S'_k \ge S'_h(k/h) - \min(k, h-k)^{1/33}$. For each fixed $k$ in this range let $B_{a,k} = \{S'_k \ge S'_h(k/h) - \min(k, h-k)^{1/33}\}$. Then:

**Lemma 69.** *For all constants $c > 0$, $0 < \epsilon < -\mathbf{E}E$, all integers $a > 1$ and all $r = O(h^{1/5})$,*

$$\mathbf{P}\left\{S'_h \textit{ bel } a, B_a \mid -\epsilon h + r \le S'_h \le -\epsilon h + r + c\right\} = O\left(\frac{1}{ha^5}\right).$$

*Proof.* We assume $r = 0$ for simplicity; an identical proof yields the result for general $r$. We denote the conditional probability $\mathbf{P}\{\cdot \mid -\epsilon h \le S'_h \le -\epsilon h + c\}$ by $\mathbf{P}^c\{\cdot\}$. Fix $k$ and $a$ as above and let $k_h = \min(k, h-k)$. Given that $-\epsilon h \le S'_h \le -\epsilon h + c$, if $B_{a,k}$ and $S'_h$ *bel* $a$ are to occur then it must be the case that $-\epsilon k - k_h^{1/33} \le S'_h(k/h) \le -\epsilon k + c + a + 1$. More strongly, there must be some integer $i$ such that $\lfloor -k_h^{1/33} \rfloor \le i \le \lceil c + a \rceil$ and for which the following events occur:

- $A_{i,k}$ is the event that $\{-\epsilon k + i \le S'_k \le -\epsilon k + i + 1\}$ and $\{S'_k \textit{ bel } a + k_h^{1/33} + c\}$, and
- $E_{i,k}$ is the event that $\{-\epsilon(h-k) - i - 1 \le S'_h - S'_k \le \epsilon(h-k) - i + c\}$ and, letting $S^*_j = S'_j - S'_k$ for $k < j < h$, $\{S^*_{h-k} \textit{ bel } a + k_h^{1/33} + c\}$.

Since $B_{a,k} \cap \{S'_h \ bel\ a\} \subseteq \bigcup_{i=\lfloor -k_h^{1/33} \rfloor}^{\lceil c+a \rceil} A_{k,i} \cap E_{k,i}$, it follows by a union bound that

$$\mathbf{P}^c\{S'_h \ bel\ a, B_a\} \ \leq \ \sum_{k=a^{33}}^{h-a^{33}} \sum_{i=\lfloor \epsilon k - k_h^{1/33} \rfloor}^{\lceil \epsilon k + c + a \rceil} \mathbf{P}^c\{E_{k,i}, A_{k,i}\} \tag{5.14}$$

We will show that for each $k$, for all $i$ in the above range

$$\mathbf{P}^c\{A_{k,i}, E_{k,i}\} \ = \ O\left(\frac{(a + k_h^{1/33} + c)^{10} h^{1/2}}{(k(h-k))^{3/2}}\right) \tag{5.15}$$

Presuming this for the moment that (5.15) holds, using this bound in (5.14) yields

$$\begin{aligned}
\mathbf{P}^c\{S'_h \ bel\ a, B_a\} \ &\leq \ O\left(\sum_{k=a^{33}}^{h-a^{33}} \sum_{i=\lfloor \epsilon k - k_h^{1/33} \rfloor}^{\lceil \epsilon k + c + a \rceil} \frac{(a + k^{1/33} + c)^{10} h^{1/2}}{(k(h-k))^{3/2}}\right) \\
&= \ O\left(\sum_{k=a^{33}}^{\lceil h/2 \rceil} \sum_{i=\lfloor \epsilon k - k^{1/33} \rfloor}^{\lceil \epsilon k + c + a \rceil} \frac{(a + k^{1/33} + c)^{10} h^{1/2}}{(k(h-k))^{3/2}}\right) \\
&= \ O\left(\sum_{k=a^{33}}^{\lceil h/2 \rceil} \frac{(a + k^{1/33} + c)^{11} h^{1/2}}{(k(h-k))^{3/2}}\right) \\
&= \ O\left(\frac{1}{h} \sum_{k=a^{33}}^{\lceil h/2 \rceil} \frac{(a + k^{1/33} + c)^{11}}{k^{3/2}}\right).
\end{aligned}$$

Since $a \leq k^{1/33}$, $(a + k^{1/33} + c)^{11} = O(k^{1/3})$, so the last fraction above is $O(1/k^{7/6})$. Therefore, the whole sum is $O(a^{33/6}) = O(a^5)$, so we have $\mathbf{P}^c\{S'_h \ bel\ a, B_a, 0\} = O(1/a^5 h)$ as desired. It therefore remains to prove (5.15).

By the strong Markov property, $A_{k,i}$ and $E_{k,i}$ are independent, so we have $\mathbf{P}\{E_{k,i}, A_{k,i}\} = \mathbf{P}\{E_{k,i}\}\mathbf{P}\{A_{k,i}\}$. By Lemma 68, it follows that

$$\mathbf{P}\{A_{k,i} \mid -\epsilon k + i \leq S'_k \leq -\epsilon k + i + 1\} = O\left(\frac{(a + k_h^{1/33} + c)^5}{k}\right) \tag{5.16}$$

and

$$\mathbf{P}\left\{E_{k,i} | -\epsilon(h-k)-i-1 \leq S^*_{h-k} \leq -\epsilon(h-k)-i+c\right\} = O\left(\frac{(a+k_h^{1/33}+c)^5}{h-k}\right). \quad (5.17)$$

By the asymptotic estimates for large deviations found in Appendix A (or, for the reader who recalls the details of Section 2.5, by Lemma 35), it is immediate that

$$\frac{\mathbf{P}\left\{i \leq S'_k + \epsilon k \leq i+1\right\} \cdot \mathbf{P}\left\{-i-1 \leq S^*_{h-k} + \epsilon(h-k) \leq -i+c\right\}}{\mathbf{P}\left\{0 \leq S'_h + \epsilon h \leq c\right\}} = \Theta\left(\frac{h^{1/2}}{(k(h-k))^{1/2}}\right).$$

Combining this equation with the two equations (5.16) and (5.17), this proves (5.15) and completes the proof. □

We remark that if $h$ is within $O(\log n)$ of $\tau n$ then taking $\epsilon = 1/\tau$, $r = \epsilon h - n$, we have $r = O(\log h) = O(h^{1/5})$ and furthermore, the event $\{-\epsilon h + r \leq S'_h \leq -\epsilon h + r + c\}$ is precisely the event $\{-n \leq S'_h \leq -n + c\}$. In the remainder of the section, we will often use Lemmas 68 and 69 to bound probabilities of the form $\mathbf{P}\left\{\cdot | -n \leq S'_h \leq -n+c\right\}$ for $h$ with $h - \tau n = O(\log n)$, without bothering to explicitly derive the values $\epsilon$ and $r$.

## 5.4 The lower bound

The proof of the lower bound is a refinement of the line of argument of Section 5.2: we consider the special set of good nodes, find the depth $h'$ at which we expect $\Theta(1)$ good nodes, and use the properties of the random walks ending at good nodes to show that with probability $\Omega(1)$, $T_n$ has a good node, and hence *some* node, at depth $h'$.

We denote by $h' = h'(n)$ the smallest depth at which we expect at most 1 good node of $T_n$. Since for any $h$, $\mathbf{E}|G_{n,h}| = \Theta(\mathbf{E}|N_{n,h}|/h) = \Theta(d^h \mathbf{P}\left\{S'_h \geq -n\right\}/h)$, by Lemma 66 (b),

$h'(n) = h^*(n) - \alpha^{-1}\ln n + O(1)$, which by Lemma 66 (a) is $\tau n - (3/2)\alpha^{-1}\ln n + O(1)$. We let $\beta = (3/2)\alpha^{-1}$, so that $h'(n) = \tau n - \beta \ln n + O(1)$. (In the remainder of the chapter we show that Theorem 63 holds with this choice of $\tau$ and of $\beta$.)

By our choice of $\tau$ and of $\beta$, we have $\mathbf{E}|G_{n,h'}| = \Theta(1)$. By following chain of reasoning described above, we will prove:

**Lemma 70.** $\mathbf{P}\{|N_{n,h'}| \geq 1\} = \Omega(1)$

As noted at the end of Section 5.2.2, the lower half of the bound of Theorem 62 immediately follows from a bound such as that of Lemma 70.

*Proof of Lemma 70.* The chain of reasoning in the proof is quite similar to that of the proof of Fact 67. For a node $x$ at depth $h$, $\mathbf{P}\{x \text{ is good}\} = \Theta(1/h)\mathbf{P}\{-n \leq x \leq 2-n\}$. Since $h' - \tau n$ is $O(\log n)$, we may apply Lemma 69 to obtain that for all integers $a > 1$, the probability that there is some $a^{33} < k < h' - a^{33}$ for which

$$S'_i \geq S'_h\left(\frac{i}{h'}\right) - \min\{k, h'-k\}^{1/33}$$

is $O(1/ha^5)\mathbf{P}\{-n \leq x \leq 2-n\}$ and is therefore $O(\mathbf{P}\{x \text{ is good}\}/a^5)$. It follows that there is $C$ such that if $x$ is good then with probability at least $1/2$, for all $C \leq k \leq h' - C$,

$$S'_i \;\leq\; S'_h\left(\frac{i}{h'}\right) - \min\{k, h'-k\}^{1/33}. \tag{5.18}$$

If $x$ is good and additionally satisfies this condition, we say that $x$ is *well-behaved.* We denote the set of well-behaved nodes at depth $h$ by $W_{n,h}$. We emphasize that every well-behaved node is good, and every good node is in $T_n$. Furthermore, since each good node is well-behaved with probability at least $1/2$, it follows from the definition of $h'$ that $\mathbf{E}W_{n,h'} = \Omega(1)$. We claim that $\mathbf{E}\{|W_{n,h'}| \mid v_{h'} \in W_{n,h'}\} = \mathbf{E}W_{n,h'} + O(1)$. Applying (5.10) with $S =$

$W_{n,h'}$, it immediately follows that $\mathbf{P}\{|W_{n,h'}| \geq 1\} = \Omega(1)$, so since $W_{n,h'} \subset N_{n,h'}$, we have $\mathbf{P}\{|N_{n,h'}| \geq 1\} = \Omega(1)$. We thus turn to proving our claim about $W_{n,h'}$, namely, that

$$\mathbf{E}\{|W_{n,h'}| \mid v_{h'} \text{ is well-behaved}\} = \mathbf{E}|W_{n,h'}| + O(1).$$

We now remind the reader of some notation from earlier in the section. We recall that node $v_{i-1}$ has child $v_i$ that is on the distinguished path $P$, and that its remaining children are $v_i^{(1)}, \ldots, v_i^{(d-1)}$. Node $v_i^{(j)}$ is the root of a subtree of $T_\infty$ that we denoted $T_\infty^{i,j}$.

Let the set of well-behaved nodes at depth $h$ that are in $T_\infty^{i,j}$ be $W_{n,h}^{i,j}$. To simplify notation, let $W_i = W_{n,h'}^{i,1}$ and denote the functions $\mathbf{P}\{\cdot \mid v_{h'} \text{ is well-behaved}\}$ and $\mathbf{E}\{\cdot \mid v_{h'} \text{ is well-behaved}\}$ by $\mathbf{P}^w\{\cdot\}$ and $\mathbf{E}^w\{\cdot\}$, respectively. Finally, for each $i$ fix an arbitrary node $x_i$ at depth $h$ that is a descendent of $v_i^{(1)}$, let the partial sums of the negatives of the labels on the path from $v_i^{(1)}$ to $x_i$ be $S'_{x_i,1}, S'_{x_i,2}, \ldots, S'_{x_i,h'-i-1}$. We remark that the edge labels contributing to the sum $S'_{x_i,h'-i-1}$ are a subset of the edge labels contributing to the vertex label $S'_{x_i}$; more precisely,

$$S'_{x_i} \;=\; S'_{x_i,h'-i-1} + S'_i - E_{v_{i-1}v_i^{(1)}} \tag{5.19}$$

We now mimic the portion of the proof of Fact 67 that leads to (5.12), in our case for the particular value $h = h'$. By symmetry, we have

$$\begin{aligned}\mathbf{E}^w\{W_{n,h'}\} &= 1 + \sum_{i=0}^{h'-1}\sum_{j=1}^{d-1}\mathbf{E}^w\left\{W_{n,h}^{i,j}\right\} \\ &= 1 + \sum_{i=0}^{h'-1}(d-1)\mathbf{E}^w\{W_i\}. \end{aligned} \tag{5.20}$$

Since $d = O(1)$ and, for a given $i$, $W_{h'-i} \leq i^d$, it follows that for any integer $0 < c = O(1)$, $\sum_{i=h'-c}^{h'-1}(d-1)\mathbf{E}^w\{W_i\} \leq \sum_{i=1}^{c}(d-1)i^d = O(1)$. By this fact, by (5.20), and by the fact that $(d-1)$ is constant, to prove the lemma it therefore suffices to show that

$$\sum_{i=1}^{h'-C} \mathbf{E}^w\{|W_i|\} = O(1), \tag{5.21}$$

where $C$ is the same constant as in the definition of well-behaved.

By symmetry, $\mathbf{E}^w\{|W_i|\} = d^{h'-i}\mathbf{P}^w\{x_i \in W_i\}$. In order for $x_i \in W_i$ to occur, we must in particular have that $S'_{x_i} \geq -n$, so by (5.19), we must have $S'_{x_i,h'-i-1} \geq -n - S'_i = c - n$ for some $c \geq 0$. Furthermore, $S'_{x_i,h'-i-1}$ is distributed as $S'_{h'-i-1}$ and is independent of $E_1, \ldots, E_h$, and therefore of $S'_i$. This independence will allow us use the bounds of Lemma 66 to bound the conditional probability that $S'_{x_i,h'-i-1} \geq c - n$. When $i$ is far from 1 and from $h' - C$, the bounds on $S_i$ given by the fact that $v_{h'}$ is well-behaved will ensure that $c$ is large enough that the conditional probability that $S'_{x_i,h'-i-1} \geq c - n$ is extremely small. By slightly modifying this same approach, we will prove similar bounds when $i$ is near 1 or near $h' - C$; summing these bounds will prove (5.21). We now turn to the details.

For any $i$, letting $n_i = n(h'-i)/h'$ as in the proof of Fact 67 and mimicking the derivation of (5.13) gives

$$\begin{aligned} h' - i - 1 &= \frac{h' n_i}{n} + O(1) \\ &= \tau n_i - \frac{\beta n_i \ln n}{n} + O(1) \\ &\geq h'(n_i) + O(1), \end{aligned}$$

So since $S'_{x_i,h'-i-1}$ is distributed as $S'_{h'-i-1}$ and is independent of $E_1, \ldots, E_h$, for any $c = o(\sqrt{n_i})$, by Lemma 66 (b) we have

$$\begin{aligned} \mathbf{P}^w\left\{S'_{x_i,h'-i-1} \geq c - n_i\right\} &= \mathbf{P}\left\{S'_{h'-i-1} \geq c - n_i\right\} = \Theta(\mathbf{P}\left\{S'_{h'(n_i)} \geq c - n_i\right\}) \\ &= \Theta\left(\frac{n_i e^{-c\gamma}}{d^{h'(n_i)}}\right) = \Theta\left(\frac{n_i e^{-c\gamma}}{d^{h'-i}}\right). \end{aligned} \tag{5.22}$$

For any $i > C$, since $v_{h'}$ is well-behaved we have

$$S'_i \leq S'_{h'}(i/h') - \min\{i, h'-i\}^{1/33} \leq 2 - \min\{i, h'-i\}^{1/33} - \frac{ni}{h'}.$$

For for such an $i$, therefore, in order for $x_i$ to be in $T_n$, we must have

$$S'_{x_i,h'-i-1} \geq \min\{i, h'-i\}^{1/33} - 2 - \frac{n(h'-i)}{h'} = \min\{i, h'-i\}^{1/33} - 2 - n_i. \qquad (5.23)$$

In particular, when $\min\{i, h'-i\}^{1/33} \geq 2\ln n/\gamma$, applying (5.22) with $c = \min\{i, h'-i\}^{1/33} - 2$ yields that $\mathbf{P}^w\{x_i \in T_n\} = O(n_i e^{-2\ln n}/d^{h'-i}) = O((d^{h'-i}n)^{-1})$, so by linearity of expectation and symmetry, $\mathbf{E}^w\{W_i\} = O(1/n)$ for such $i$. Letting $t = t(n) = \lceil (2\ln n/\gamma)^{33} \rceil$, then,

$$\sum_{i=t}^{h'-t} \mathbf{E}^w\{W_i\} = O\left(\frac{h'-2t}{n}\right) = O(1). \qquad (5.24)$$

This bounds the bulk of the sum (5.21); it remains to consider the cases when $i$ is either close to 1 or close to $h' - C$.

**Case 1** ($h' - t \leq i \leq h' - C$)**:** let $k = h' - i$, so $C \leq k < t$. Since $k$ is so small, $n_i = n(h'-i)/h = nk/h \leq k/\tau + O(1)$. By (5.22) and (5.23), we thus have

$$\mathbf{P}^w\{x_i \in T_n\} \quad = \quad O\left(\frac{ke^{-\gamma k^{1/33}}}{d^k}\right). \qquad (5.25)$$

By linearity of expectation and by symmetry,

$$\mathbf{E}^w\{W_i\} = d^k \mathbf{P}^w\{x_i \in T_n\} = O(ke^{-\gamma k^{1/33}}) = O(k^{-3}).$$

Therefore, $\sum_{i=h'-t}^{h'-C} \mathbf{E}^w\{W_i\} = O(\sum_{k=C}^{t} k^{-3}) = O(1)$.

**Case 2** ($1 \leq i < t$)**:** by (5.19), we know that $S'_{x_i,h'-i-1} = S'_{x_i} - (S'_i - E_{v_{i-1}v_i^{(1)}})$. We condition on the value of $S_i' - E_{v_{i-1}v_i^{(1)}}$ (which is necessarily less than $2 - ni/h'$ as $v_{h'}$ is well-behaved), and show that for *any* $s \geq ni/h' - 2$,

$$\mathbf{P}^w\left\{x_i \in W_{n,h'} \mid -(s+1) \leq S_i' - E_i^{(1)} < -s\right\} = O(d^{-(h'-i)}i^{-3}). \tag{5.26}$$

This implies that $\mathbf{E}^w\{|W_{n,h'}|\} = d^{h'-i}\mathbf{P}^w\{x_i \in W_{n,h'}\} = O(1/i^3)$, for all $1 \leq i \leq t$, so $\sum_{i=1}^{t}\mathbf{E}^w\{W_i\} = O(1)$. It thus remains to prove (5.26).

For $x_i \in W_{n,h'}$ to occur given that $-(s+1) \leq S_i' - E_i^{(1)} < -s$, we must have that $S'_{x_i,h'-i-1} \geq s - n(h'-i)/h' = s - n_i$. If $s \geq ni/h' + 2\ln n/\gamma$ then since $i \leq t = O((\ln n)^{33})$, it follows that $2\ln n/\gamma = o(\sqrt{n_i})$. By applying (5.22) with $c = 2\ln n/\gamma$, we therefore have

$$\begin{aligned}\mathbf{P}^w\left\{x_i \text{ is well-behaved} \mid -(s+1) \leq S_i' - E_i^1 < -s\right\} &\leq \mathbf{P}\left\{S'_{h'-i-1} \geq 2\ln n/\gamma - n_i\right\}\\ &= O(d^{-(h'-i)}n^{-1}).\end{aligned}$$

It follows that for such $i$ and $s$, $\mathbf{E}^w\left\{W_i \mid -(s+1) \leq S_i' - E_i^{(1)} < -s\right\} = O(n^{-1}) = O(i^{-3})$. Finally, if $s \leq ni/h' + 2\ln n/\gamma t$ then letting $a = s - ni/h' \geq -2$, in order for $x_i$ to be well-behaved (in fact, in order for it to be *good*), it must be the case that

(i) $a - n(h'-i)/h' \leq S'_{x_i,h'-i-1} \leq a + 3 - n(h'-i)/h'$, and

(ii) the random walk $S'_{x_i,1}, \ldots, S'_{x_i,h'-i-1}$ stays below $a+3$.

We again emphasize that the random walk $S'_{x_i,1}, \ldots, S'_{x_i,h'-i-1}$ is independent of $E_1, \ldots, E_{h'}$. By (5.22), the probability of (i) given $E_1, \ldots, E_{h'}$ is $O(n_i e^{-a\gamma}/d^{h'-i})$. By Lemma 68, the conditional probability of (ii) given (i) and $E_1, \ldots, E_{h'}$ is $O(a^5/n_i)$. Combining these two bounds yields

$$\mathbf{P}^w\left\{x_i \text{ is well-behaved} \mid -(s+1) \leq S_i' - E_i^{(1)} < -s\right\} = O\left(\frac{a^5 e^{-a\gamma}}{d^{h'-i}}\right).$$

If $i \le C$ then $i^3 = O(1)$, so since $a \ge -2$ and $\gamma$ is constant, $a^5 e^{-a\gamma} d^{-(h'-i)} = O(d^{-(h'-i)}) = O(d^{-(h'-i)} i^{-3})$, and (5.26) holds. If $i \ge C$, then as $v_{h'}$ is well-behaved, $-(s+1) \le S_i' - E_i^{(1)} \le S_i' \le 2 - ni/h' - i^{1/33}$, so $a = s - ni/h' \ge i^{1/33} - 3$. In this case, $a^5 e^{-a\gamma}$ is $O(i^{1/11} e^{-\gamma i^{1/33}})$, which is $O(1/i^3)$, and again (5.26) holds. This completes the proof. □

## 5.5 The upper bound

We know that the naive approach to proving an upper bound on $\mathbf{P}\{H_n \ge h' + i\}$, namely, bounding $\mathbf{P}\{v_{h'+i} \in T_n\}$, then applying a union bound, will not work; it is the approach we used in Section 5.2, and only yields an upper bound of $h^* + O(1)$ on the expected height. To prove our upper bound, then, we will certainly have to take into consideration the dependence between the labels of different nodes at depth $h'+i$. Equivalently, we will have to understand the dependence between the $d^{h'+i}$ different *random walks* from the root to depth $h' + i$.

We observe that we can *easily* prove strong enough bounds for *one* group of potential nodes of $T_n$, namely, the set of *good* nodes at depth $h' + i$: by the standard rotation argument (or by Lemma 68), $\mathbf{P}\{v_h \in G_{n,h'+i}\} = O(\mathbf{P}\{v_h \in N_{n,h'+i}\} / (h' + i))$, from which strong bounds on $\mathbf{P}\{G_{n,h'+i} \neq \emptyset\}$ follow directly from Lemma 66 (b) and a union bound over all nodes at depth $h' + i$.

We prove our general bound by splitting the nodes of $T_n$ at depth $h' + i$ into *many* groups. In each group the behavior of the random walks leading to the nodes of the group will be in some sense homogeneous, in that any random walk leading to a node in the group will satisfy some specified constraint on its difference from its conditioned mean. This homogeneity will yield that the accuracy of our ballot-theorem-plus-union-bound based argument to bound the probability that any given group is non-empty will be greatly improved. In essence, this is because our conditions on the random walks will tell us precisely how to optimize our applications of the lemmas from Section 5.3 for these specific nodes. When we recombine the bounds we obtain for the individual groups (again by union bound), the result will be strong enough to yield exponential upper tail bounds on $\mathbf{P}\{N_{n,h'+i} \neq \emptyset\}$ for small $i$:

**Lemma 71.** *There are $A_1 > 0$, $\epsilon' > 0$ such that for $0 \leq i = O(n^{1/5})$, $\mathbf{P}\{N_{n,h'+i} \neq \emptyset\} \leq A_1(1-\epsilon')^i$.*

We note that the upper half of the bound of Theorem 63 immediately follows Lemma 71, as if $N_{n,h'+i} = \emptyset$ then $H_n < h' + i$. Since $h' - (\tau n - \beta \ln n) = O(1)$, we may choose $A = O(1)$ large enough that $A_1(1-\epsilon')^{(h'-(\tau n - \beta \ln n))+A} \leq 1/4$, say. Lemma 71 then implies that

$$\mathbf{P}\{H_n \geq \tau n - \beta \ln n + A\} = \mathbf{P}\{H_n \geq \tau n - \beta \ln n + (h' - (\tau n - \beta \ln n)) + A\} \leq 1/4.$$

(In fact, the bound of Lemma 71 is much stronger than we need in order to prove the upper half of the bound of Theorem 63.) In proving Lemma 71, we will repeatedly use the following fact, which is a simple consequence of the definition of $h'$ and Lemma 66 (b), and which we restate for convenience:

**Fact 72.** *For $h = h' + i$ and $0 < i = o(\sqrt{n})$ $\mathbf{P}\{S'_h \geq -n\} = \Theta(\mathbf{P}\{2 - n \geq S'_h \geq -n\}) = \Theta\left(\frac{ne^{-\alpha i}}{d^h}\right)$.*

We note that we need only prove the bound of Lemma 71 for $i$ larger than any fixed constant, as we may presume the bound holds for $i$ small by our choice of $A_1$. We may also restrict our attention to $i \leq (1+\alpha^{-1}) \log n$, say, as if $i > (1+\alpha^{-1}) \log n$ then $h' + i \geq h^* + \log n + O(1)$ and the result is implied by Lemma 66 (b). For the remainder of this section, we assume $i$ is an integer between 0 and $h^* - h' + \log n$, and let $h = h' + i$. By Lemma 66 (b), for any node $x$ at depth $h$ $\mathbf{P}\{x \in N_{n,h}, S'_x \leq 2 - n\} = \Theta(\mathbf{P}\{x \in N_{n,h}\})$, so it also suffices to prove that $\mathbf{P}\{\exists x \in N_{n,h}, S'_x \leq 2 - n\} = O((1-\epsilon)^i)$.

We now proceed to define the "homogeneous" groups discussed above, and bound the probabilities they are non-empty, in a sequence of claims. The proof of each claim will consist of straightforward applications of Lemma 66 and the lemmas of Section 5.3, and Lemma 71 will be a trivial consequence of the bounds of the claims.

Let $a = \lfloor e^{\alpha i/10} \rfloor$, and let $A_a$ be the set of nodes $x$ of $T_n$ at depth $h$ for which $S'_x \leq 2 - n$ and for which the negative random walk ending at $x$ *stays below* $a$.

**Claim 73.** $\mathbf{P}\{A_a \neq \emptyset\} = O(e^{-i\alpha/2})$.

*Proof.* By Lemma 68,

$$\mathbf{P}\{v_h \in A_a \mid -n \leq S'_h \leq 2 - n\} = O\left(\frac{a^5}{h}\right) = O\left(\frac{e^{i\alpha/2}}{h}\right).$$

By Fact 72, it follows that $\mathbf{P}\{v_h \in A_a\} = O\left(ne^{-i\alpha/2}/hd^h\right) = O(e^{-\alpha i/2}/d^h)$. The claim follows by symmetry and by linearity of expectation. □

Since $h^*(n) = \tau n - (2\alpha)^{-1} \log n$, there is $c^*$ such that for any $h' \leq h \leq h^* + \log n$ and $k \leq h$

$$\left| h^*\left(\frac{nk}{h}\right) - k \right| \leq c^* \log n, \tag{5.27}$$

and for any $k \leq h/\log h$,

$$\left| h^*\left(\frac{nk}{h}\right) - k \right| \leq c^*. \tag{5.28}$$

Let $c = \log^5 n$, and let $H_c$ consist of the nodes $x$ of $T_n$ at depth $h$ for which $S'_x \leq 2 - n$ and for which the negative random walk ending at $x$ *does not stay below* $c$. Then

**Claim 74.** $\mathbf{P}\{H_c \neq \emptyset\} = O\left(e^{-i}\right)$.

*Proof.* If $H_c$ is not empty, then for some $0 < k < h$ there is a node $y$ at depth $k$ for which the sum of the negatives of the labels on the path to $y$ is at least $-nk/h + c$. When $k \geq c^3$, by (5.27) and by Lemma 66 (b),

$$\begin{aligned}\mathbf{P}\left\{S'_k \ge c - \frac{nk}{h}\right\} &\le \mathbf{P}\left\{S'_{h^*(nk/h)-\lceil c^* \log n\rceil} \ge c - \frac{nk}{h}\right\}\\ &= O\left(\frac{e^{\alpha c^* \log n - \gamma c}}{d^k}\right) = O\left(\frac{1}{n^{\log^4 n} d^k}\right),\end{aligned}$$

for $n$ large enough, as $\gamma c - \alpha c^* \log n = \omega(\log^4 n)$. When $i^4 \le k \le c^3$, we have $k \le h/\log h$ as long as $n$ is large enough, so by (5.28) and Lemma 66 (b),

$$\begin{aligned}\mathbf{P}\left\{S'_k \ge c - \frac{nk}{h}\right\} &\le \mathbf{P}\left\{S'_{h^*(nk/h)-c^*} \ge k^{1/3} - \frac{nk}{h}\right\}\\ &= O\left(\frac{e^{\alpha c^* - \gamma k^{1/3}}}{d^k}\right) = O\left(\frac{1}{e^{\gamma k^{1/3}} d^k}\right).\end{aligned}$$

When $k \le i^4$, $c - nk/h \ge c - k - c^* \ge c - ((1+\alpha^{-1})\log n)^4 - c^* > 0$ for $n$ large enough, so $\mathbf{P}\{S'_k \ge c - nk/h\} = 0$. By the above bounds and by symmetry, we thus have

$$\begin{aligned}\mathbf{P}\{H_c \ne \emptyset\} &\le \sum_{k=1}^{h-1} d^k \mathbf{P}\left\{S'_k \ge c - \frac{nk}{h}\right\}\\ &\le O\left(\sum_{k=i^4}^{c^3} \frac{1}{e^{\gamma k^{1/3}} d^k}\right) + O\left(\sum_{k=c^3+1}^{h-1} \frac{1}{n^{\log^4 n}}\right)\\ &= O\left(e^{-i}\right) + O(n^{1-\log^4 n}) = O(e^{-i}),\end{aligned}$$

as $i = O(\log n)$. □

For each integer $b$ with $a \le b \le c$, we say that $v_h$ is in the set $M_b$ if $v_h \in T_n$, $S'_h \le 2 - n$, *and* $S'_h$ *is below* $b+1$ *but not below* $b$. Note that this implies $S'_k \ge b - nk/h$ for some $0 < k < h$. We say that $v_h$ is in $M_b^{mid}$ if $v_h \in M_b$ and additionally $S'_k \ge b - nk/h$ for some $b^{33} \le k \le h - b^{33}$. For $k < b^{33}$ (resp. $k > h - b^{33}$), we say that $v_h$ is in $M_b^k$ if $v_h \in M_b$ and additionally, *$k$ is the smallest value for which* $S'_k \ge b - nk/h$. We extend these definitions to the other nodes $x$ at depth $h$ by symmetry. We remind the reader that $a = \lfloor e^{\alpha i/10} \rfloor$ and that $c = \log^5 n$.

**Claim 75.** $\mathbf{P}\left\{\exists b,\ a \le b \le c \ s.t.\ M_b^{mid} \ne \emptyset\right\} = O(e^{-\alpha i})$

*Proof.* Fix $a \le b \le c$. By Lemma 69,

$$\mathbf{P}\left\{v_h \in M_b^{mid} \mid -n \le S_h' \le 2-n\right\} = O\left(\frac{1}{hb^5}\right),$$

so by Fact 72, $\mathbf{P}\left\{v_h \in M_b^{mid}\right\} = O(ne^{-\alpha i}/hb^5 d^h) = O(e^{-\alpha i}/b^5 d^h)$. By symmetry and by linearity of expectation, $\mathbf{P}\left\{M_b^{mid} \ne \emptyset\right\} = O(e^{-\alpha i}/b^5)$. The claim follows by summing over $b$. □

**Claim 76.** $\mathbf{P}\left\{\exists b,\ a \le b \le c, 1 \le k \le b^{33} \ s.t.\ M_b^k \ne \emptyset\right\} = O(e^{-i})$

*Proof.* Fix $b$ and $k$ as above. For each node $x$ at depth $k$, let $W_x$ be the set of descendents of $x$ in $M_b^k$. By symmetry and a union bound, $\mathbf{P}\left\{M_b^k \ne \emptyset\right\} \le d^k \mathbf{P}\left\{W_{v_k} \ne \emptyset\right\}$. If $W_{v_k}$ is nonempty, then necessarily $S_k' \ge b - nk/h$. Since $k \le b^{33} \le c^{33} = (\log n)^{165}$, for $n$ large enough $k \le h/\log h$, so by (5.28), $k \ge h^*(nk/h) - c^*$.

If $k \le \tau b/2$, then since $n/h = \tau + o(1)$, we have $b - nk/h > 0$ for $n$ large enough, so $\mathbf{P}\left\{W_{v_k} \ne \emptyset\right\} = 0$. If $\tau b/2 < k \le b^{33}$, then $b \ge k^{1/33} \ge (\tau b/2)^{1/33}$. By Lemma 66 (b) we therefore have

$$\begin{aligned}
\mathbf{P}\left\{W_{v_k} \ne \emptyset\right\} &\le \mathbf{P}\left\{S_k' \ge b - \frac{nk}{h}\right\} \le \mathbf{P}\left\{S_k' \ge \left(\frac{\tau b}{2}\right)^{1/33} b^{1/33} - \frac{nk}{h}\right\} \\
&= O\left(\frac{e^{\alpha c^* - \gamma(\tau b/2)^{1/33} b^{1/33}}}{d^k}\right) = O\left(\frac{e^{-\gamma(\tau b/2)^{1/33}}}{d^k}\right).
\end{aligned}$$

A union bound over $1 \le k \le b^{33}$ thus gives

$$\mathbf{P}\left\{M_b^k \ne \emptyset\right\} = O\left(b^{33} e^{-\gamma(\tau b/2)^{1/33}}\right) = O(e^{-b^{1/40}}).$$

We sum this bound over $b$ between $a$ and $c$, and crudely bound the whole sum by writing $O(\sum_{b=a}^{c} e^{-b^{1/40}}) = O(e^{-a^{1/41}})$. Since $a^{1/41}$ is at least $i$ as long as $i$ is large enough, the claim follows. □

**Claim 77.** $\mathbf{P}\left\{\exists b,\ a \le b \le c, 1 \le k \le b^{33} \text{ s.t. } M_b^{h-k} \neq \emptyset\right\} = O(e^{-i})$

*Proof.* Fix $b$ and $k$ as above, and let $r = h - k$. For each node $x$ at depth $r$, let $W_x$ be the set of descendents of $x$ in $M_b^k$. By symmetry and a union bound, $\mathbf{P}\left\{M_b^{h-k} \neq \emptyset\right\} \le d^{h-k}\mathbf{P}\left\{W_{v_r} \neq \emptyset\right\}$. Suppose $W_{v_r}$ is nonempty – then necessarily $b - nr/h \le S'_r \le 2 + (b + 1) - nr/h$ *and in addition* $S'_r$ *bel* $b$ (or else $r$ is not the *first* time $S'$ exceeds its mean by $b$).

Since $r = h - k = h - O((\log n)^{165}) = h' + i - O((\log n)^{165})$, an easy calculation mimicking many we have seen before shows that $r - h'(nr/h) = i + O(1)$. Therefore, Lemma 66 (b) implies that

$$\mathbf{P}\left\{b - nr/h \le S'_r \le b + 3 - nr/h\right\} = O\left(\frac{ne^{-\alpha b}}{d^r}\right).$$

Furthermore, by Lemma 68 and the fact that $r \ge h - b^{33} = \Omega(h) = \Omega(n)$, we have

$$\mathbf{P}\left\{S'_r \text{ bel } b \mid b - nr/h \le S'_r \le b + 3 - nr/h\right\} = O\left(\frac{b^5}{r}\right) = O\left(\frac{b^5}{n}\right).$$

Combining these bounds yields that

$$\mathbf{P}\left\{W_{v_r} \neq \emptyset\right\} = O\left(\frac{b^5 e^{-\alpha b}}{d^r}\right) = O\left(\frac{e^{5\ln b - \alpha b}}{d^{h-k}}\right),$$

so by symmetry and a union bound, $\mathbf{P}\left\{M_b^{h-k} \neq \emptyset\right\} = O(e^{5\ln b - \alpha b})$. Just as in the proof of Claim 76, summing this bound over $b$ yields the result. □

We are now prepared for

*Proof of Lemma 71.* $H_n \ge h$ is the event that there is a vertex of $T_n$ at depth $h$, i.e., that

$N_{n,h} \neq \emptyset$. It is immediate from the definitions of the sets $A_a$, $H_c$, $M_b^{mid}$ and $M_b^k$ that any vertex in $N_{n,h}$ is either in $A_a$, or in $H_c$, or in $M_b^{mid}$ (for some integer $a \leq b \leq c$), or in $M_b^k$ (for some integer $a \leq b \leq c$ and some integer $k$ for which either $1 \leq k \leq b^{33}$ or $h - b^{33} \leq k \leq h$). Applying Claims 73, 74, 75, 76, and 77, respectively, to bound each of these events, it follows that

$$\mathbf{P}\{H_n \geq h\} = O(e^{-\alpha i/2}) + O(e^{-i}) + O(e^{-\alpha i}) + O(e^{-i}) + O(e^{-i}) = O((1-\epsilon')^i)$$

for some $\epsilon' > 0$. This completes the proof. □

## 5.6 Conclusion

There are several natural questions raised by our results. Our results in particular yield the expected value of the minimum of a branching random walk with bounded degree and bounded step sizes that have zero probability of extinction, to within $O(1)$. Both the requirement of bounded degree and that of bounded step sizes, should be able to be relaxed to some degree while maintaining that the conclusions of Theorem 62 still hold. In fact, an approach very similar to that seen above can be used to prove such a result when we have "strong enough" exponential upper tail bounds on the step size $E$ (which is a condition quite similar to but stronger than that required for the Hammersley-Kingman-Biggins theorem); a proof of this fact will eventually be presented elsewhere. What are necessary and sufficient conditions for such a result to hold?

The most obvious natural complement to the results of this section would be a treatment of the case $\mathbf{P}\{E = 0\} = 1/d$. In this case it is known (McDiarmid, 1995) that $\mathbf{Var}\{M_n\} = O(\log n)$. However, Bramson (1978b) has shown examples of branching random walks for which in our terminology $\mathbf{P}\{E = 0\} = 1/d$ and for which there is a constant $c$ such that

$M_n - c \log \log n$ converges in distribution to a non-degenerate random variable. It follows that any result akin to that of Theorem 62 that held when $\mathbf{P}\{E = 0\} = 1/d$ would certainly require a different kind of additive renormalization than that of Theorem 62. For what functions $f(n)$ are there branching random walks with step size $E$ for which $\mathbf{P}\{E = 0\} = 1/d$ and and $M_n - f(n)$ converges in distribution to a non-degenerate random variable?

# Chapter 6

# Conclusion: A percolation-theoretic perspective

In the second half of this thesis, particularly in Chapters 4 and 5, we explored the connection between ballot theorems (and, more generally, random walks) and the heights of random trees. The perspective of *first-passage percolation* provides another way to look at our results, and yields a raft of research questions that may deserve study using a ballot theorem-based approach. To explain this angle in a little more detail, we first provide an (extremely abridged) introduction to the key ideas of first-passage percolation.

Suppose we are given a connected graph $G = (V, E)$ with non-negative edge weights given by a function $w : E \to [0, \infty)$, and a fixed node $r \in V$ that we call the *origin*. (We assume for the moment that $G$ is finite but will later relax this restriction.) Given a path $P = v_0 v_1 \ldots v_k$, the *weight* $w(P)$ of $P$ is $\sum_{i=0}^{k-1} w(v_i v_{i+1})$. Given any node $v \neq r$ in $G$, the *weighted distance* from $r$ to $v$ is the weight of the smallest weight path from $r$ to $v$ – we denote this weighted distance $d(r, v)$. Finally, for any $t \geq 0$, we let $G_r(t)$ be subgraph of $G$ induced by all nodes $v$ with $d(r, v) \leq t$, and call $G_r(t)$ the *first passage percolation cluster* of depth $t$ rooted at $r$.

We hereafter assume that $G$ and $w$ are such that for all $v$, there is a *unique* path $P$ for which $w(P) = d(r,v)$, i.e., a unique smallest weight path from $r$ to $v$, and denote this path $P_v$; we furthermore presume that if $v \neq w$ then $d(r,v) \neq d(r,w)$. (We say that in this case the weight function $w$ is *reasonable.*) We let $T_r$ be the union over all $v \in V$ of the paths $P_v$, and call $T_r$ the *shortest path tree* for $(G,w)$. For $t \geq 0$ we let $T_r(t)$ be the intersection of $T_r$ and $G_r(t)$.

To see that $T_r$ *is* a tree, fix any vertex $v \in V$. Then $v \in G_r(w(P_v))$ and, for any $t < w(P_v)$, $v \notin G_r(t)$. Since $G$ is finite, by our assumption that $d(r,w) \neq d(r,v)$ for $w \neq v$ it follows that there is $0 \leq t < w(P_v)$ for which $G_r(w(P_v)) = G_r(t) - \{v\}$. Letting $v^*v$ be the last edge on the path $P_v$, necessarily $v^* \in G_r(t)$ – it follows that $T_r(w(P_v))$ is just the graph $T_r(t)$ together with the edge $v^*v$. It is then easy to prove by induction that $T_r$ is indeed a tree.

Clearly, if $0 \leq s \leq t$ then $G_r(s) \leq G_r(t)$; since $G$ is connected we also have $\cup_{t\geq 0} G_r(t) = G$. The family of graphs $\{G_r(t)\}_{t\geq 0}$ defines an increasing graph process, in the sense that if $0 \leq s \leq t$ then $G_r(s) \subseteq G_r(t)$, and since $G$ is connected there is some value $t^*$ for which $G_r(t^*) = G$. Similarly, $\{T_r^s(t)\}_{t\geq 0}$ is an increasing graph process, and $T_r = T_r(t^*)$ is a spanning tree for $G$. In fact, in showing that $T_r$ was a tree we showed the stronger statement that $T_r(t)$ is a tree for all $t$, so in particular is a spanning tree for $G_r(t)$.

If the edge weights given by $w$ are *random*, then $\{G_r(t)\}_{t\geq 0}$ and $\{T_r(t)\}_{t\geq 0}$ are *random* graph processes, which we can investigate in the same spirit as we investigated the with the graph processes of Chapter 4.

We begin by supposing that $G = K_n$ and that the edge weights $w(e)$ are given by independent uniform $[0,1]$ random variables: as with $G_{n,p}$ we may ask how $G_r(t)$ and $T_r(t)$ grow as $t$ grows from 0 to 1. In fact, for this problem it has been more common to study $K_n$ with *exponential mean* 1 edge weights. For such a weighting Janson (1999) has that for any fixed $r$, letting $t^*$ be the smallest $t$ for which $G_r(t) = K_n$, it is the case that $t^* \log n/n \rightarrow 2$ in probability.

In other words, for this choice of $G$ and of $w$, for any $\epsilon > 0$ the first passage percolation cluster $G_r(t)$ becomes the whole graph at time between $(2 - \epsilon)n/\log n$ and $(2 + \epsilon)n/\log n$ with probability tending to 1 as $n$ tends to infinity.

When $G = K_n$ and $w$ weights the edges of $G$ with iid random variables (and all edge weights are distinct with probability 1), we derived the height of another spanning tree of $K_n$ in Chapter 4: the *minimum weight* spanning tree of $G$. It turns out that the structure of the shortest path tree $T_r$ is much simpler than the structure of the minimum weight spanning tree of $K_n$; we now explain why. Given any time $t \geq 0$ and corresponding tree $T_r(t)$, the next edge to attach to $T_r(t)$ in the graph process is equally likely to connect any pair of vertices of $V(T_r(t))$ and $V(K_n) - V(T_r(t))$. In particular, the location at which the next edge connects to $T_r(t)$ is equally likely to be any vertex of $T_r(t)$. Trees constructed in such a fashion called are called *random recursive trees*, and are extremely well-studied; see (Smythe and Mahmoud, 1995) for a survey of results on their structure. In particular, their height is known to be asymptotic to $e \ln n$ (Moon, 1974). Together with the results of Chapter 4, this demonstrates in particular that for edge weights as above, the shortest path tree $T_r$ is *not* distributed like the minimum weight spanning tree of $K_n$.

When random recursive trees arise as shortest path trees for $K_n$, many new problems arise: for example, what is the total weight of $T_r$? What is the *weighted* height of $T_r$? What is the average weighted distance between nodes in $T_r$? These and many other questions have been investigated in a sequence of papers by van der Hofstad et al. (2001, 2002, 2005a,b), as well as by Janson (1999). This viewpoint also provokes new questions for the minimum weight spanning tree. What is its *weighted* diameter? If we consider the method of Prim (1957) for growing the MWST (which grows the MWST starting from a single node – see Page 4.1), then we may think of the starting node as the root – what then is the expected weighted height of the MWST?

We may also consider the subject of Chapter 5 in a first passage percolation-theoretic light. To do so, we relax the restriction imposed early in the conclusion that $G$ is finite, and additionally allow that the weight function $w$ take value infinity. In Chapter 5 we considered infinite $d$-ary rooted trees whose edges $e$ were assigned "exponential labels" $L_e$ taking values in $[-\infty, 0]$. Let $G = T_d^\infty$ and let $w(e) = -L_e$ – then in the language of Chapter 5, for $n \geq 0$ $G_r(n)$ is just the tree $T_n$. Theorem 62, then, can be viewed as answering a question about the moments of first-passage percolation on trees.

Can the techniques of Chapter 5 be used to study first-passage percolation on other graphs? The variance of the depth of the first-passage percolation cluster in $\mathbb{Z}^2$ with exponential mean 1 edge weights is a well-studied open problem (see Kesten (1993b); Pemantle and Peres (1994); Talagrand (1995); Newman and Piza (1995); Benjamini et al. (2003), among other work). To date the best known upper bound on this variance is $O(t/\log t)$ and the best known lower bound is of order $\Omega(\log t)$. Benjamini et al. (2003) state that physical evidence suggests the variance of the depth is order $t^{1/3}$, and that recent work on increasing subsequences in random permutations of IID random variables (Zeitouni and Deuschel, 1999; Johansson, 2000) supports this hypothesis. If the variance is indeed of order $t^{1/3}$, then rigorous upper bounds and lower bounds that approach the "correct" answer to the problem both remain elusive. Might the approach of Chapter 5 provide some insight into its solution?

# Appendix A

# The requisite results on large deviations

The results we use concern the so-called *exact asymptotics* for large deviations. They address the first order term of $\mathbf{P}\{S_n > \mathbf{E}S_n + \epsilon n\}$ for positive $\epsilon$, for sums of random variables which individually satisfy certain conditions on their moments. The result we state is considerably less general than the strongest known results; we have simplified the setting as much as possible given the demands of the thesis.

We assume we are given iid random variables $X_1, X_2, \ldots$, distributed as $X$, where $X$ is a non-positive, extended real-valued random variable with mean $-1$. We furthermore assume for simplicity that there is no real number $r \neq 0$ for which $rX$ assumes only integer values (i.e., $X$ is not a *lattice* random variable). There are versions of all results from this section which apply to lattice random variables, and work just as well for all our purposes; we set aside a formal treatment of such random variables purely to avoid an unenlightening and gratuitous technical burden.

The *cumulant generating function*, or *rate function*, of $X$ is defined as $\Lambda = \Lambda_X(t) = \log \mathbf{E}\left\{e^{tX}\right\}$ (we will omit the subscript $X$ when obvious from context). We first state a few properties of the function $\Lambda$ which we will need in our discussion.

**Fact 78.** $\Lambda(t) \leq 0$ *for* $t \geq 0$ *and* $\lim_{t\to\infty} \Lambda(t) = -\infty$.

*Proof.* This is immediate as $X$ is non-positive and has negative mean. □

We let $t^* = \inf\{t \mid \Lambda(t) < \infty\}$ – the previous fact implies $t^* \leq 0$.

**Fact 79.** *The function* $\Lambda$ *is convex, and for any* $t \in (t^*, \infty)$*,* $\Lambda$ *is infinitely differentiable in an open neighbourhood of* $t$.

*Proof.* See (Dembo and Zeitouni, 1992, Lemma 2.2.5 and Exercise 2.2.24). □

**Fact 80.** $\lim_{t\to 0^+} \Lambda'(t) \geq -1$ *and, if* $t^* < 0$ *then* $\Lambda'(0) = -1$.

*Proof.* By definition, $\Lambda(0) = \log \mathbf{E}\left\{e^{0X}\right\} = 0$, and for any $t > 0$, by Jensen's inequality $\Lambda(t) = \log \mathbf{E}\left\{e^{tX}\right\} \geq \mathbf{E}\left\{tX\right\} = -t$. This establishes the first claim as $\Lambda'(t)$ is increasing. By an identical argument, if $t^* < t < 0$ then $\Lambda(t) \geq -t$, so $\Lambda'(0) \leq -1$; the second claim follows. □

Let $x^* = \sup\{x \mid \mathbf{P}\left\{X > x\right\} > 0\}$ – then we have

**Fact 81.** $\Lambda'$ *is increasing and* $\lim_{t\to\infty} \Lambda'(t) = x^*$.

*Proof.* The fact that $\Lambda'$ is increasing follows from Fact 79. To see the second claim, fix any $\epsilon > 0$ and let $p = p(\epsilon) = \mathbf{P}\left\{x^* - \epsilon \leq X \leq x^*\right\}$. We have $\Lambda(t) = \log(\mathbf{E}\left\{e^{tX}\mathbf{1}_{[X \geq x^* - \epsilon]}\right\} +$

$\mathbf{E}\left\{e^{tX}\mathbf{1}_{[X<x^*-\epsilon]}\right\})$. As $X$ is non-positive and has negative mean, the first expectation above dominates the second, i.e.,

$$\lim_{t\to\infty}(\Lambda(t) - \log \mathbf{E}\left\{e^{tX}\mathbf{1}_{[X\geq x^*-\epsilon]}\right\}) = 0.$$

Finally, since $pt(x^*-\epsilon) \leq \log \mathbf{E}\left\{e^{tX}\mathbf{1}_{[X\geq x^*-\epsilon]}\right\} \leq ptx^*$, it follows that for all $\delta > 0$ there exists $t_\delta > 0$ such that for all $t \geq t_\delta$,

$$pt(x^*-\epsilon) - \delta \leq \log \mathbf{E}\left\{e^{tX}\mathbf{1}_{[X\geq x^*-\epsilon]}\right\} \leq ptx^* + \delta.$$

As we may make $\epsilon$ and $\delta$ arbitrarily small and $\Lambda'$ is increasing, the result follows. □

For $i > 0$, let $S_i = X_1 + \ldots + X_i$. The following result was proved by Badahur and Rao (1960) and is stated here approximately as found in (Dembo and Zeitouni, 1992, Theorem 3.7.4 and Exercise 3.7.10).

**Theorem 82** (Badahur and Rao (1960)). *For any $0 < a \leq o(\sqrt{n})$ and $t > 0$*

$$\mathbf{P}\left\{S_n \in [n\Lambda'(t), n\Lambda'(t) + a]\right\} = (1 + o(1))\frac{(1-e^{-ta})e^{-n(t\Lambda'(t)-\Lambda(t))}}{t\sqrt{\Lambda''(t)2\pi n}}, \quad \text{(A.1)}$$

*and given any $g(n)$ tending to infinity with $n$, the convergence $o(1)$ is uniform in $a \leq \sqrt{n}/g(n)$.*

(The uniformity statement does not appear in Dembo and Zeitouni (1992), but fact implied by the proof given there.) We note that, letting $f(t) = t\Lambda'(t) - \Lambda(t)$, $f(0) = 0$ and $f'(t) = t\Lambda''(t)$ is positive when $t$ is positive; therefore the second exponent in (A.1) is always negative, so these estimates are non-trivial. We also have

**Corollary 83.** *$f$ is increasing and $\lim_{t\to\infty} f(t) = \ln(1/\mathbf{P}\left\{X = x^*\right\})$ (where we interpret the right-hand-side as $\infty$ if $\mathbf{P}\left\{X = x^*\right\} = 0$).*

*Proof.* Since for any $\epsilon > 0$, $\mathbf{P}\{S_n \geq x^*n\} \geq \mathbf{P}\{X = x^*\}^n$, it is immediate from Theorem 82 that $\lim_{t\to\infty} f(t)$ must be at most $\ln(1/\mathbf{P}\{X = x^*\})$. To see that in fact equality holds, choose any small $\epsilon > 0$ and any $c > \epsilon$. If we are to have $S_n \geq (x^* - \epsilon)n$ then at most an $\epsilon/c$ proportion of the $X_i$ can satisfy $X_i \leq x^* - c$. Letting $p = p(c) = \mathbf{P}\{X \geq x^* - c\}$, it follows that $\mathbf{P}\{S_n \geq (x^* - \epsilon)n\} \leq \mathbf{P}\{\text{Bin}(n,p) \geq (1 - \epsilon/c)n\}$. Since $p(c) \to \mathbf{P}\{X = x^*\}$ as $c \to 0$, letting $c \to 0$ and $\epsilon \to 0$ in such a manner that $1 - \epsilon/c \to 1$, it follows from standard binomial estimates that $\lim_{t\to\infty} f(t)$ must be at least $\ln(1/\mathbf{P}\{X = x^*\})$. □

By studying $\Lambda$ and its derivatives, we can use this theorem to derive a relation between the asymptotics of $\mathbf{P}\{S_n > n\Lambda'(t)\}$ and $\mathbf{P}\{S_m > n\Lambda'(t)\}$, the precise form of which depends on $n - m$. We show:

**Corollary 84.** *When $b = o(n^{1/2})$, for $n$ large enough*

$$\mathbf{P}\{S_{n-b} \geq n\Lambda'(t)\} = (1 + o(1))e^{b(f(t)-\Lambda(t))}\mathbf{P}\{S_n \geq n\Lambda'(t)\},$$

*and for any function $g$ tending to infinity with $n$, the convergence $o(1)$ is uniform for $b \leq \sqrt{n}/g(n)$.*

*Proof.* Let $m = n - b$, and suppose $m$ is small enough that there is $s > 0$ for which $n\Lambda'(t) = m\Lambda'(s)$. Since $\Lambda'$ is negative and increasing, if $m > n$ then such an $s$ certainly exists, $s \geq t$, and $\Lambda$ is infinitely differentiable on an open neighbourhood of $s$. Furthermore, since $\Lambda$ is continuous and $t > 0$, there is $\epsilon > 0$ such that if $m \geq (1 - \epsilon)n$ again we can find such a value $s$. Furthermore, by choosing $\epsilon$ small enough we may again ensure that $\Lambda$ is infinitely differentiable in an open neighbourhood of $s$. Since $m \geq n - o(n^{1/2})$, for $n$ large enough $m$ is certainly at least $(1 - \epsilon)n$.

Writing $\Lambda'(s) - \Lambda'(t) = O(\Lambda''(t)(s - t))$ and letting $b = n - m$ yields that $(s - t) = O(b/m)$.

By a Taylor approximation of $\Lambda'$ around $t$, we also have that

$$\Lambda'(s) - \Lambda'(t) = (s-t)\Lambda''(t) + O((s-t)^2),$$

so since $\Lambda'(s) = (n/m)\Lambda'(t)$,

$$\begin{aligned}(s-t) &= \frac{\Lambda'(s) - \Lambda'(t) + O((s-t)^2)}{\Lambda''(t)} = \frac{(n/m)\Lambda'(t) - \Lambda'(t)}{\Lambda''(t)} + O((s-t)^2) \\ &= \frac{b\Lambda'(t)}{m\Lambda''(t)} + O((s-t)^2) = \frac{b\Lambda'(t)}{m\Lambda''(t)} + O\left(\frac{b^2}{m^2}\right). \qquad \text{(A.2)}\end{aligned}$$

Furthermore, a Taylor expansion of $f$ around $t$ gives

$$f(s) - f(t) = (s-t)f'(t) + O((s-t)^2) = (s-t)t\Lambda''(t) + O\left(\frac{b^2}{m^2}\right),$$

which combined with the approximation of $(s-t)$ from (A.2) and the fact that $f(t) = t\Lambda'(t) - \Lambda(t)$ gives

$$\begin{aligned}f(s) &= f(t) + \frac{b\Lambda'(t)f'(t)}{m\Lambda''(t)} + O\left(\frac{b^2}{m^2}\right) = f(t) + \frac{bt\Lambda'(t)}{m} + O\left(\frac{b^2}{m^2}\right) \\ &= \frac{nf(t)}{m} + \frac{b}{m}\Lambda(t) + O\left(\frac{b^2}{m^2}\right). \qquad \text{(A.3)}\end{aligned}$$

It follows that $e^{-mf(s)} = e^{-b\Lambda(t)} \cdot e^{-nf(t)} \cdot e^{O(b^2/m)}$. By this fact and by Theorem 82, we have

$$\begin{aligned}\mathbf{P}\{S_m \geq n\Lambda'(t)\} &= \mathbf{P}\{S_m \geq m\Lambda'(s)\} = (1+o(1))\frac{e^{-mf(s)}}{s\sqrt{2\pi m\Lambda''(s)}} \\ &= (1+o(1))\frac{e^{-b\Lambda(t)} \cdot e^{-nf(t)} \cdot e^{O(b^2/m)}}{s\sqrt{2\pi m\Lambda''(s)}}.\end{aligned}$$

When $b = o(\sqrt{n})$, the term $e^{O(b^2/m)}$ is $1 + o(1)$ and furthermore, $s\sqrt{\Lambda''(s)m} = (1 +$

$o(1))t\sqrt{\Lambda''(t)n}$. Therefore,

$$\mathbf{P}\left\{S_m \geq n\Lambda'(t)\right\} = (1+o(1))\frac{e^{-b(\Lambda(t))}e^{-mf(t)}}{t\sqrt{2\pi n\Lambda''(t)}}.$$

When $b \leq \sqrt{n}/g(n)$, $e^{O(b^2/m)} = e^{O(1/g(n)^2)}$, which yields the second claim of the lemma. $\square$

Combining the theorem and the first corollary, we obtain the bound

**Corollary 85.** *When $c$ and $c'$ are $o(\sqrt{n})$, for $n$ large enough,*

$$\mathbf{P}\left\{S_{n+c'} \geq n\Lambda'(t) + c\right\} = (1+o(1))\frac{e^{c'\Lambda'(t)-ct}e^{-nf(t)}}{t\sqrt{\Lambda''(t)2\pi n}},$$

*and given any $g(n)$ tending to infinity with $n$, the convergence $o(1)$ is uniform in $c, c' \leq \sqrt{n}/g(n)$.*

We omit a formal proof of Corolary 85 as it proceeds exactly as the proof of Corollary 83.

# Appendix B

# Ratio limit theorems

In Section 2.4, we stated the following theorem. Recall that $X$ is a *lattice random variable with period* $r$ if there is $r > 0$ such that $rX$ is integer, and let $f$ be the density function of a $\mathcal{N}(0,1)$ random variable.

**Theorem 86.** *Suppose $S_n$ is a sum of independent, identically distributed random variables distributed as $X$ with $\mathbf{E}X = 0$, and there is a sequence of constants $a_n$ such that $S_n/a_n$ converges to a $\mathcal{N}(0,1)$ random variable. If $X$ is non-lattice let $B$ be any bounded set; then for any $h \in B$ and $x \in R$*

$$\mathbf{P}\{|S_n - x| \leq h/2\} = \frac{h\Phi(x/a_n)}{a_n} + o(a_n^{-1}).$$

*Furthermore, if $X$ is a lattice random variable with period $r$, then for any $x \in \{n/r \mid n \in \mathbb{Z}\}$,*

$$\mathbf{P}\{S_n = x\} = \frac{\Phi(x/a_n)}{a_n} + o(a_n^{-1}).$$

*In both cases, $a_n o(a_n^{-1}) \to 0$ as $n \to \infty$ uniformly over all $x \in \mathbb{R}$ and $h \in B$.*

This theorem is in fact a consequence of a much more general result of Stone (1965b); it is known as a *ratio limit theorem*, as it gives control over the ratio

$$\frac{\mathbf{P}\{|S_n - x| \le h/2\}}{\mathbf{P}\{|S_n - y| \le h/2\}}$$

for certain $x$ and $y$ – in particular, if both are $O(a_n)$, then this ratio is $\Theta(\Phi(x/a_n)/\Phi(y/a_n)) = \Theta(e^{y^2-x^2})$.

The behavior of such ratios is closely linked to the *spread* of the sum $S_n$. The spread of a sum of (not necessarily identically distributed) random variables is generally measured by the *concentration function* $Q(S_n, h) = \sup\{\mathbf{P}\{x \le S_n \le x+h\} \; : \; x \in \mathbb{R}\}$, which measures how much of its mass $S_n$ puts in any interval of length $h$. If $\lim_{n\to\infty} Q(S_n, h) = 0$ then $S_n$ is called "essentially divergent"; of course, for iid sums of random variables; this is equivalent to the requirement that $\mathbf{Var}\{X_1\} > 0$. We used a result on the spread of sums of iid random variables (Theorem 9) in our development of the generalized ballot theorem of Chapter 2.4. Theorem 9 is implied by Theorem 86, which is fundamentally more powerful, as it supplies not just an upper bound for the probability that $S_n$ is in a certain set, but gives precise asymptotics for this probability.

Much more is known about such ratios than is contained in the above theorem. The paper by Stone (1965b) in fact proves a result such as Theorem 86 in the case that $X$ is in the range of attraction of *any* stable distribution with a density. In particular, this might give us hope that we could extend our analysis to random variables $X$ with $\mathbf{E}\{X^{1+\alpha}\} < \infty$ for some $0 < \alpha < 1$. (Unfortunately, as we noted in Section 2.6, the doubling argument that was a key element of the proof does not hold in this case; more work is still needed to derive a ballot theorem for such random variables.) Stone's work was one step in a series of increasingly general results, the most general of which is found in Stone (1966) and applies to random walks on locally compact commutative groups; other results on the subject are found in

(Rvaceva, 1962; Gnedenko and Kolmogorov, 1954; Bretagnolle and Dacunha-Castelle, 1964; Stone, 1965a).

We note that ballot theorems themselves provide a sort of ratio limit theorem: if

$$\mathbf{P}\left\{S_i > 0 \;\forall\; 0 < i < n \mid x \leq S_n \leq x+h\right\} = \Theta(x/n)$$

then letting $T^-$ be the time of entry of $S$ into the interval $(-\infty, 0)$, we have

$$\frac{\mathbf{P}\left\{x \leq S_n \leq x+h, T^- > n\right\}}{\mathbf{P}\left\{y \leq S_n \leq y+h, T^- > n\right\}} = \Theta\left(\frac{\mathbf{P}\left\{y \leq S_n \leq y+h\right\}}{\mathbf{P}\left\{x \leq S_n \leq x+h\right\}}\right).$$

If $S_n/a_n$ converges to a normal distribution, for example, and $x, y = O(a_n)$ then Theorem 86 tells us this ratio is $\Theta(e^{y^2-x^2})$.

In fact, the existence of ratio limit theorems has been investigated for *many* events other than $\{x \leq S_n \leq x+h\}$. The number of entries to a set up to time $n$, the event that $S_n$ is the first partial sum in a set, the probability of hitting one set before another, and many related quantities, have all been studied from this perspective. Some of the most powerful results appear in (Kesten and Spitzer, 1963a,b; Port and Stone, 1967; Ornstein, 1969a,b; Levitan, 1971). Kesten (1972) provides a rather comprehensive and easily digestible survey of results related to the spread of sums of independent random variables and ratio limit theorems. We note that Theorems 30 through 33 seem to be a first step towards a ratio limit theorem for ratios of the form

$$\frac{\mathbf{P}\left\{x \leq S_n \leq x+A, T_{-\epsilon\sqrt{n}} > n\right\}}{\mathbf{P}\left\{y \leq S_n \leq y+A, T_{-\epsilon\sqrt{n}} > n\right\}},$$

where $T_{-\epsilon\sqrt{n}}$ is the first time $t$ that $S_t \leq -\epsilon\sqrt{n}$. Ratio limit theorems of roughly this form have been established for certain classes of random walks and Markov chains (Kesten and Spitzer, 1963b; van Doorn and Schrijner, 1995; Kesten, 1995), but usually with $x$ and $y$ fixed (not varying with $n$) and with the stopping time $T_{-\epsilon\sqrt{n}}$ replaced by $T_r$ with $r$ fixed.

Isaac (1983) has shown a close connection between the existence of ratio limit theorems and an *asymptotic independence* property for random walks. (A random walk on $\mathbb{Z}$, say, is asymptotically independent if for all fixed $i, j, k$, $\mathbf{P}\{S_i = j \mid S_n = k\} \rightarrow \mathbf{P}\{S_i = j\}$ as $n \rightarrow \infty$.) Kesten (1995) has investigated the same connection for more general Markov chains. In Section 2.1, we discussed how the early steps of a random walk conditioned on the event $\{S_n =\}0$ should not be too different from an unconditioned random walk; asymptotic independence is one way to formalize what we might mean by "not too different". We strongly believe that the appearance of this connection in the ratio limit theorem setting is no accident, and that a better understanding of it will lead immediately to progress on generalizing the ballot theorems of Chapter 1.

# Bibliography

L. Addario-Berry, N. Broutin, and B. Reed. The diameter of the minimum spanning tree of the complete graph. *Random Structures and Algorithms*, submitted.

D. Aldous. A random tree model associated with random graphs. *Random Structures and Algorithms*, 4:383–402, 1990.

D. Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Annals of Probability*, 25 (2):812–854, 1997.

Noga Alon and Joel H. Spencer. *The Probabilistic Method.* John Wiley & Sons, Inc, second edition, 2000.

Erik Sparre Andersen. Fluctuations of sums of random variables. *Mathematica Scandinavica*, 1:263–285, 1953.

Erik Sparre Andersen. Fluctuations of sums of random variables ii. *Mathematica Scandinavica*, 2:195–223, 1954.

Désiré André. Solution directe du probleme resolu par m. bertrand. *Comptes Rendus de l'Academie des Sciences*, 105:436–437, 1887.

K. B. Athreya. Large deviation rates for branching processes–i. single type case. *The Annals of Applied Probability*, 4(3):779–790, August 1994.

K. B. Athreya and P. E. Ney. *Branching Processes.* Springer, Berlin, 1972.

Markus Bachmann. Limit theorems for the minimal position in a branching random walk with independent logconcave displacements. *Advances in Applied Probability*, 32(1):159–176, 2000.

Markus Bachmann. *Limit theorems for the minimal position in a branching random walk with independent logconcave displacements*. PhD thesis, Purdue University, 1998.

R.R. Badahur and R. Ranga Rao. On deviations of the sample mean. *Annals of Mathematical Statistics*, 31(4):1015–1027, December 1960.

N. Balakrishnan. *Advances in Combinatorial Methods and Applications to Probability and Statistics*. Birkhäuser, Boston, MA, first edition, 1997.

Émile Barbier. Generalisation du probleme resolu par m. j. bertrand. *Comptes Rendus de l'Academie des Sciences*, 105:407, 1887.

Itai Benjamini, Gil Kalai, and Oded Schramm. First passage percolation has sublinear distance variance. *Annals of Probability*, 31(4), 2003.

J. Bertrand. Observations. *Comptes Rendus de l'Academie des Sciences*, 105:437–439, 1887a.

J. Bertrand. Sur un paradox analogue au problème de saint-pétersburg. *Comptes Rendus de l'Academie des Sciences*, 105:831–834, 1887b.

J. Bertrand. Solution d'un probleme. *Comptes Rendus de l'Academie des Sciences*, 105:369, 1887c.

Biggins. Growth rates in the branching random walk. *Zeitschrift für Warscheinlichkeitstheorie und verwandte Gebeite*, 48:17–34, 1979.

J. D. Biggins. *Asymptotic properties of the branching random walk*. PhD thesis, University of Oxford, 1976a.

J. D. Biggins. The first- and last-birth problems for a multitype age-dependent branching process. *Advances in Applied Probability*, 8(3):446–459, 1976b.

J. D. Biggins. Chernoff's theorem in the branching random walk. *Journal of Applied Probability*, 14:630–636, 1977.

J. D. Biggins. How fast does a general branching random walk spread? In *Classical and Modern Branching Processes*, volume 84 of *IMA Volumes in Mathematics and its Applications*, pages 19–40. Springer Verlag, 1997.

B. Bollobás. *Random Graphs.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, second edition, 2001.

B. Bollobás. The evolution of random graphs. *Transactions of the American Mathematical Society*, 286(1):257–274, 1984.

B. Bollobás, Y. Kohayakawa, and T. Łuczak. The evolution of random subgraphs of the cube. *Random Structures and Algorithms*, 3:55–90, 1992.

O. Borůvka. O jistém problému minimálním. *Práce Mor. Př írodověd. Spol. v Brně (Acta Societ. Scient. Natur. Moravicae)*, 3:37–58, 1926.

Maury Bramson. Maximum displacement of branching brownian motion. *Communications on Pure and Applied Mathematics*, 31:531–581, 1978a.

Maury Bramson. Minimal displacement of branching random walk. *Zeitschrift für Warscheinlichkeitstheorie und verwandte Gebeite*, 55:89–108, 1978b.

Maury Bramson. Location of the travelling wave for the kolmogorov equation. *Probability Theory and Related Fields*, 73:481–515, 1986.

J. Bretagnolle and D. Dacunha-Castelle. Convergence de la nieme convoluée d'une loi de probabilité. *Comptes Rendus de l'Academie des Sciences*, 258:4910–4913, 1964.

N. Broutin and L. Devroye. Large deviations for the weighted height of an extended class of trees. *Algorithmica*, to appear, 2005.

N. Broutin, L. Devroye, and E. McLeish. Weighed height of random trees. in preparation, 2006.

N. Broutin, L. Devroye, E. McLeish, and M. de la Salle. The height of increasing trees. *Random Structures and Algorithms*, submitted.

B. Chauvin and M. Drmota. The random multisection problem, travelling waves, and the distribution of the height of m-ary search trees. to appear in Algorithmica, 2005.

H. Chernoff. A measure of the asymptotic efficiency for tests of a hypothesis based on the sum of observables. *Ann. Math. Statist.*, 2:493–509, 1952.

F. Chung and L. Lu. The diameter of random sparse graphs. *Advances in Applied Mathematics*, 26:257–279, 2001.

K.L. Chung and P. Erdös. On the application of the borel-cantelli lemma. *Transactions of the American Mathematical Society*, 72(1), 1952.

T.H. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein. *Introduction to algorithms.* MIT Press, Cambridge, MA, second edition, 2001.

F.M. Dekking and B. Host. Limit distributions for minimal displacement of branching random walks. *Zeitschrift für Warscheinlichkeitstheorie und verwandte Gebeite*, 90(3): 403–426, 1991.

A. Dembo and O. Zeitouni. *Large Deviation Techniques and Applications.* Jones and Bartlett, 1992.

L. Devroye. A note on the height of binary search trees. *Journal of the ACM*, 33:489–498, 1986.

L. Devroye and B. Reed. On the variance of the height of random binary search trees. *SIAM Journal on Computing*, 24:1157–1162, 1995.

M. Drmota. An analytic approach to the height of binary search trees II. *Journal of the ACM*, 50:333–374, 2003.

R. Durrett. Maxima of branching random walks. *Zeitschrift für Warscheinlichkeitstheorie und verwandte Gebeite*, 62:165–170, 1983.

Aryeh Dvoretzky and Theodore Motzkin. A problem of arrangements. *Duke Mathematical Journal*, 14:305–313, 1947.

Meyer Dwass. A fluctuation theorem for cyclic random variables. *The Annals of Mathematical Statistics*, 33(4):1450–1454, December 1962.

Meyer Dwass. The total progeny in a branching process and a related random walk. *Journal of Applied Probability*, 6(3):682–686, 1969.

P. Erdős and A. Rényi. On random graphs I. *Publ. Math. Debrecen*, 6:290–297, 1959.

P. Erdős and A. Rényi. On the evolution of random graphs. *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, 5:17–61, 1960.

P. Erdős and A. Rényi. On the strength of connectedness of a random graph. *Acta Math. Acad. Sci. Hung*, 12:261–267, 1961.

P. Erdös. Some remarks on the theory of graphs. *Bulletin of the American Mathematical Society*, 53:292–294, 1947.

P. Erdös, S. Suen, and P. Winkler. On the size of a maximal random graph. *Random Structures and Algorithms*, 6:309–318, 1995.

William Feller. *An Introduction to Probability Theory and Its Applications, Volume 1*, volume 1. John Wiley & Sons, Inc, third edition, 1968.

P. Flajolet and A. Odlyzko. The average height of binary trees and other simple trees. *Journal of Computer and System Sciences*, 25(171–213), 1982.

A. M. Frieze. On the value of a random minimum spanning tree problem. *Discrete Appl. Math.*, 10:47–56, 1985.

A. M. Frieze and C. McDiarmid. Algorithmic theory of random graphs. *Random Structures and Algorithms*, 10(1-2):5–42, 1997.

A. M. Frieze and C. J. H. McDiarmid. On random minimum length spanning trees. *Combinatorica*, 9:363–374, 1989.

I.I. Gikhman and A.V. Skorokhod. *Introduction to the Theory of Random Processes.* Saunders Mathematics Books. W.B. Saunders Company, 1969.

B.V. Gnedenko and A.N. Kolmogorov. *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley series in statistics. Addison-Wesley Publishing Company, 1954.

R. L. Graham and P. Hell. On the history of the minimum spanning tree problem. *Annals of the History of Computing*, 7:43–57, 1985.

Philip S. Griffin and Terry R. McConnell. On the position of a random walk at the time of first exit from a sphere. *The Annals of Probability*, 20(2):825–854, April 1992.

Philip S. Griffin and Terry R. McConnell. Gambler's ruin and the first exit position of random walk from large spheres. *The Annals of Probability*, 22(3):1429–1472, July 1994.

G.R. Grimmett and D.R. Stirzaker. *Probability and Random Processes.* Clarendon, Oxford, second edition, 1992.

Howard D. Grossman. Fun with lattice-points. *Duke Mathematical Journal*, 14:305–313, 1950.

A. Hald. *A History of Probability and Statistics and Their Applications before 1750.* John Wiley & Sons, Inc, New York, NY, 1990.

J.M. Hammersley. Postulates for subadditive processes. *The Annals of Probability*, 2(4): 652–680, 1974.

T. E. Harris. *The Theory of Branching Processes.* Springer, Berlin, 1963.

Richard Isaac. Asymptotic independence and individual ratio limit properties. *Zeitschrift für Warscheinlichkeitstheorie und verwandte Gebeite*, 62:201–214, 1983.

G.I. Ivchenko. The strength of connectivity of a random graph. *Theory of Probability and its Applications*, 18(2):396–403, 1973a.

G.I. Ivchenko. On the asymptotic behavior of degrees of vertices in a random graph. *Theory of Probability and its Applications*, 18(1):188–195, 1973b.

S. Janson. One, two and three times log n/n for paths in a complete graph with random weights. *Combinatorics, Probability and Computing*, 8(1999):347–361, 1999.

S. Janson, D.E. Knuth, T. Luczak, and B. Pittel. The birth of the giant component. *Random Structures nd Algorithms*, 4(3):233–359, 1993.

S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs.* Wiley, New York, 2000.

V. Jarňik. O jistém problému minimálnim. *Práce Mor. Přírodověd. Spol. v Brně (Acta Societ. Scient. Natur. Moravicae)*, 6:57–63, 1930.

Kurt Johansson. Transversal fluctuations for increasing subsequences on the plane. *Zeitschrift für Warscheinlichkeitstheorie und verwandte Gebeite*, 116(4):445–456, 2000.

Olav Kallenberg. *Foundations of Modern Probability.* Probability and Its Applications. Springer Verlag, second edition, 2003.

Olav Kallenberg. Ballot theorems and sojourn laws for stationary processes. *The Annals of Probability*, 27(4):2011–2019, October 1999.

Harry Kesten. Sums of independent random variables–without moment conditions. *Annals of Mathematical Statistics*, 43(3):701–732, June 1972.

Harry Kesten. Frank spitzer's work on random walks and brownian motion. *Annals of Probability*, 21(2):593–607, April 1993a.

Harry Kesten. On the speed of convergence in first passage percolation. *Annals of Applied Probability*, 3(2):On the speed of convergence in first passage percolation, 1993b.

Harry Kesten. A ratio limit theorem for (sub) markov chains on 1, 2,... with bounded jumps. *Advanced Applied Probability*, 27(3):652–691, 1995.

Harry Kesten and R.A. Maller. Infinite limits and infinite limit points of random walks and trimmed sums. *The Annals of Probability*, 22(3):1473–1513, 1994.

Harry Kesten and Frank Spitzer. Ratio theorems for random walks i. *Journal d'analyse Mathématique*, 11:285–322, 1963a.

Harry Kesten and Frank Spitzer. Ratio theorems for random walks 2. *Journal d'analyse Mathématique*, 11:323–379, 1963b.

Jeong Han Kim. A poisson cloning model for random graph. in preparation, 2006.

J.F.C. Kingman. The first birth problem for an age-dependent branching process. *The Annals of Probability*, 3(5):790–801, 1975.

Takis Konstantopoulos. Ballot theorems revisited. *Statistics & Probability Letters*, 24(4): 331–338, September 1995.

J. B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. American Math. Society*, 2:48–50, 1956.

Michael L. Levitan. A generalized doeblin ratio limit theorem. *The Annals of Mathematical Statistics*, 42(3):904–911, 1971.

T. Łuczak. Component behavior near the critical point of the random graph process. *Random Structures and Algorithms*, 1(3):287–310, 1990.

T. Łuczak. Cycles in a random graph near the critical point. *Random Structures and Algorithms*, 2:421–440, 1991.

T. Łuczak. The phase transition in a random graph. In *Combinatorics, Paul Erdős is Eighty*, volume 2, pages 399–422. Bolyai Society, 1993.

T. Łuczak. The number of trees with large diameter. *Journal of the Australian Mathematical Society A*, 58:298–311, 1995.

T. Łuczak. Random trees and random graphs. *Random Structures and Algorithms*, 13(3-4): 485–500, 1998.

T. Łuczak, B. Pittel, and J.C. Weirman. The structure of a random graph at the point of the phase transition. *Trans. AMS*, 341(2):721–748, 1994.

C. McDiarmid. Minimal positions in a branching random walk. *Advances in Applied Probability*, 5(1):128–139, 1995.

Sri Gopal Mohanty. An urn problem related to the ballot problem. *The American Mathematical Monthly*, 73(5):526–528, 1966.

J.W. Moon. *London Mathematical Society Lecture Notes 13*, chapter The distance between nodes in recursive trees, pages 125–132. Cambridge University Press, 1974.

Asaf Nachmias and Yuval Peres. The critical random graph, with martingales. submitted, December 2005.

J. Nešetřil. A few remarks on the history of the MST-Problem. *Archivum Mathematicum Brno*, 33:15–22, 1997.

C. M. Newman and M. Piza. Divergence of shape fluctuations in two dimensions. *Annals of Probability*, 23(3):977–1005, 1995.

Donald S. Ornstein. Random walks i. *Transactions of the American Mathematical Society*, 138:1–43, 1969a.

Donald S. Ornstein. Random walks ii. *Transactions of the American Mathematical Society*, 138(45-60), 1969b.

Robin Pemantle and Yuval Peres. *Probability and Phase Transition*, chapter Planar first-passage percolation times are not tight., pages 261–264. Kluwer Academic Publishers, 1994.

Yuval Peres. personal communication, 2006.

V.V. Petrov. *Sums of Independent Random Variables*, volume 82 of *A Series of Modern Surveys in Mathematics*. Springer Verlag, 1975.

B. Pittel. On the largest component of the random graph at a nearcritical stage. *Journal of Combinatorial Theory, Series B*, 82:237–269, 2001.

Sidney C. Port and Charles J. Stone. Hitting time and hitting places for non-lattice recurrent random walks. *Journal of Mathematics and Mechanics*, 17:35–57, 1967.

R.C. Prim. Shortest connection networks and some generalizations. *Bell System Techn. J.*, 36:1389–1401, 1957.

F.P. Ramsey. On a problem of formal logic. *Proceedings of the London Mathematical Society*, 39(2):264–286, 1929.

B. Reed. The height of a random binary search tree. *Journal of the ACM*, 50:306–332, 2003.

A. Rényi and G. Szekeres. On the height of trees. *Journal of the Australian Mathematical Society*, 7:497–507, 1967.

Émile Rouché. Sur la durée du jeu. *Comptes Rendus de l'Academie des Sciences*, 106: 253–256, 1888a.

Émile Rouché. Sur un problème relatif à la durée du jeu. *Comptes Rendus de l'Academie des Sciences*, 106:47–49, 1888b.

A. Ruciński and N. Wormald. Random graph processes with degree restrictions. *Combinatorics, Probability and Computing*, 1:169–180, 1992.

A. Ruciński and N. Wormald. Random graph processes with maximum degree 2. *Annals of Applied Probability*, 7:183–199, 1997.

E.L. Rvaceva. *Selected Translations in Mathematical Statistics and Probability*, chapter On domains of attraction of multi-dimensional distributions, pages 183–205. American Mathematical Society, 1962.

R. T. Smythe and H. M. Mahmoud. A survey of recursive trees. *Theor. Probability and Math. Statist.*, 51:1–27, 1995.

Frank Spitzer. A combinatorial lemma and its applications to probability theory. *Transactions of the American Mathematical Society*, 82(2):323–339, July 1956.

V.E. Stepanov. On the probability of connectedness of a random graph $g_m(t)$. *Theory of Probability and its Applications*, 15(1):55–67, 1970a.

V.E. Stepanov. Phase transitions in random graphs. *Theory of Probability and its Applications*, 15(2):187–203, 1970b.

Charles Stone. A local limit theorem for nonlattice multi-dimenstional distribution functions. *The Annals of Mathematical Statistics*, 36(2):546–551, April 1965a.

Charles Stone. Ratio limit theorems for random walks on groups. *Transactions of the American Mathematical Society*, 125(1):86–100, 1966.

Charles J. Stone. On local and ratio limit theorems. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pages 217–224, 1965b.

Lajos Takács. Ballot problems. *Zeitschrift für Warscheinlichkeitstheorie und verwandte Gebeite*, 1:154–158, 1962a.

Lajos Takács. Ballots, queues, and random graphs. *Journal of Applied Probability*, 26(1): 103–112, 1989.

Lajos Takács. A generalization of the ballot problem and its application in the theory of queues. *Journal of the American Statistical Association*, 57(298):327–337, 1962b.

Lajos Takács. The time dependence of a single-server queue with poisson input and general service times. *The Annals of Mathematical Statistics*, 33(4):1340–1348, December 1962c.

Lajos Takács. The distribution of majority times in a ballot. *Zeitschrift für Warscheinlichkeitstheorie und verwandte Gebeite*, 2(2):118–121, January 1963.

Lajos Takács. Combinatorial methods in the theory of dams. *Journal of Applied Probability*, 1(1):69–76, 1964a.

Lajos Takács. Fluctuations in the ratio of scores in counting a ballot. *Journal of Applied Probability*, 1(2):393–396, 1964b.

Lajos Takács. A combinatorial theorem for stochastic processes. *Bulletin of the American Mathematical Society*, 71:649–650, 1965a.

Lajos Takács. On the distribution of the supremum for stochastic processes with interchangeable increments. *Transactions of the American Mathematical Society*, 119(3):367–379, September 1965b.

Lajos Takács. *Combinatorial Methods in the Theory of Stochastic Processes.* John Wiley & Sons, Inc, New York, NY, first edition, 1967.

Lajos Takács. On the distribution of the maximum of sums of mutually independent and identically distributed random variables. *Advances in Applied Probability*, 2(2):344–354, 1970a.

Lajos Takács. On the distribution of the supremum for stochastic processes. *Annales de l'Institut Henri Poincaré B*, 6(3):237–247, 1970b.

M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. IHES*, 81:73–205, 1995.

J.C. Tanner. A derivation of the borel distribution. *Biometrika*, 48(1-2):222–224, June 1961.

R. van der Hofstad, G. Hooghiemstra, and P. van Mieghem. First passage percolation on the random graph. *Probability in the Engineering and Informational Sciences*, 15:225–237, 2001.

R. van der Hofstad, G. Hooghiemstra, and P. van Mieghem. The flooding time in random graphs. *Extremes*, 5:111–129, 2002.

R. van der Hofstad, G. Hooghiemstra, and P. van Mieghem. Size and weight of shortest path trees with exponential link weights. to appear, 2005a.

R. van der Hofstad, G. Hooghiemstra, and P. van Mieghem. The weight of the shortest path tree. to appear, 2005b.

Eric van Doorn and Pauline Schrijner. Ratio limits and limiting conditional distributions for discrete-time birth-death processes. *Journal of Mathematical Analysis and Applications*, 190:263–284, 1995.

V. Vazirani. *Approximation Algorithms.* Springer Verlag, New York, 2001.

N. Wormald. *Lectures on Approximation and Randomized Algorithms*, chapter The differential equation method for random graph processes and greedy algorithms, pages 73–155. PWN, Warsaw, 1999.

Sandy L. Zabell. Rates of convergence for conditional expectations. *The Annals of Probability*, 8(5):928–941, 1980.

O. Zeitouni and J.-D. Deuschel. On increasing subsequences of i.i.d. samples. *Combinatorics, Probability and Computing*, 8(247-263), 1999.