Three Problems in Discrete Probability

Erik Slivken

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Christopher Hoffman, Chair

Sara Billey

Ioana Dumitriu

Program Authorized to Offer Degree:
Mathematics

University of Washington

Abstract

Three Problems in Discrete Probability

Erik Slivken

Chair of the Supervisory Committee:
Advisor Christopher Hoffman
Mathematics

In this thesis we present three problems. The first problem is to find a good description of the number of fixed points of a 231-avoiding permutation. We use a bijection from Dyck paths to 231-avoiding permutations that allows us to compute the scaled distribution of the number of fixed points of a 231-avoiding permutation chosen uniformly at random. We also show a strong connection with a these permutations and Brownian excursion.

The second problem is an extension of work found in [31], where the authors study bootstrap percolation on the Hamming torus. We give a thorough description of the behavior of this model for finite lattices of all dimensions when the percolation threshold is 2.

Lastly we present a problem on jigsaw percolation, developed in [34] and [35] as a model for collaborative problem solving. This process considers a pair of graphs on a shared set of vertices and forms clusters of vertices based on the edges of the two underlying graphs. We consider the process where both graphs are Erdös-Rényi random graphs.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS

First and foremost, I would like to express sincere appreciation to my advisor, Christopher Hoffman, whose patience, guidance, and brilliance have inspired and shaped the way I view mathematics. He has made me a better mathematician. The support of committee members Sara Billey and Ioana Dumitriu have contributed greatly to my understanding of mathematics, and more importantly, what it takes to be a successful mathematician. Their mentorship has been invaluable.

I would also like to give special thanks to Douglas Rizzolo, whose countless hours of mentorship have been essential to my growth. Working with Douglas has vastly improved my writing of mathematics.

Fellow graduate students Toby Johnson, Lindsay Erickson, Joshua Tokle, Greg Drugan, Louis Fan, James Pfieffer, Joel Barnes, Matt Junge, Cris Negron, and many others deserve a paragraph of their own for their support along the way. The same must be said for friends Alex Alden, Dani Fumia, Mindy Au, and Aaron Miller and others. They made Seattle a fun place to be.

My family has been so supportive of their strange mathematician brother/son and I appreciate all the love they have given me over the years. My sister was my first math teacher, and even though she made me pretend to be a dog named circus, it was my springboard to a love of math from a young age.

Lastly, I wish to thank my girlfriend Sarah Stepanski for all her support and love. It was tough, but we made it.

Chapter 1

PATTERN AVOIDING PERMUTATIONS

1.1 Introduction

One hundred years ago Percy MacMahon initiated the study of pattern avoiding permutations with his study of "lattice permutations" [13]. MacMahon showed that every **321**-avoiding permutation can be decomposed into two increasing subsequences and that the number of **321**-avoiding permutations of length n is given by the nth Catalan number

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

The modern study of pattern avoiding permutations began with Donald Knuth who showed their importance in computer science. Knuth proved that the **231**-avoiding permutations are precisely those that can be sorted by a stack [12]. He also showed that the number of **231**-avoiding permutations is also equal to C_n. Further connections between pattern avoiding permutations and sorting algorithms in computer science were explored by Tarjan, Pratt and others [19, 1]. We say a permutation $\sigma \in S_n$ avoids the pattern **231** if for every subsequence satisfying $1 \leq i_1 < i_2 < i_3 \leq n$, if $\sigma(i_1) < \sigma(i_2)$ implies $\sigma(i_1) < \sigma(i_3)$. Similar definitions apply for all m and $\pi \in S_m$. We denote the set of permutations of length n that avoid the pattern $\pi \in \mathcal{S}_m$ by $\mathcal{S}_n(\pi)$.

From there the study of pattern avoiding permutations has gone in many directions. Connections between pattern avoiding permutations have been established with numerous other areas of mathematics and to many applications [11]. Enumeration of the number of different pattern avoiding permutation in $\mathcal{S}_n(\pi)$ for various permutations π has been a major avenue of study [8, 15].

Recently a more geometric view of pattern avoiding permutations has come into fashion. This line of work, initiated by Madras and Pehlivan [14] and Miner and Pak [16], aims to determine what a random pattern avoiding permutation looks like. These papers make

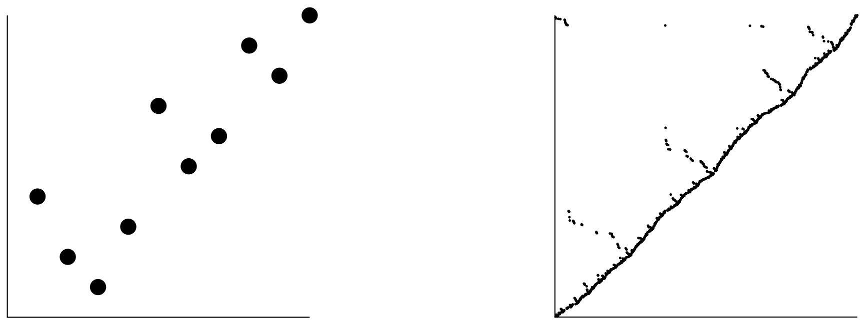

Figure 1.1: σ_{10} and σ_{1000} in $S_{10}(231)$ and $S_{1000}(231)$ respectively.

extensive studies of the distributions of permutations that avoid a pattern of length 3. For a uniformly chosen **231**-avoiding (or **321**-avoiding) permutation $\sigma \in S_n$ they determined the asymptotics of $\mathbb{P}(\sigma(i) = j)$, for i and j of the form $i(n) = an + cn^{\alpha}$ and $j(n) = bn + cn^{\alpha}$.

We extend this approach and show fundamental connections between the shape of a random permutation avoiding a pattern of length 3 and Brownian excursion. Brownian excursion is the process $\{\mathbb{e}_t\}_{0 \leq t \leq 1}$ which is Brownian motion conditioned to be 0 at 0 and 1 and positive in the interior [17].

One of the classical problems in probabilistic combinatorics is to show that the distribution of the number of fixed points in a uniformly chosen random permutation is converging in distribution to a Poisson(1) random variable. This classical result has been expanded to include a thorough examination of the cycle structure of a random permutation [18]. Moving in a different direction Elizalde initiated the study of the expected number of fixed points in pattern avoiding permutations [2, 3, 4, 5].

Instead of the local statistics studied in [14] and [16] we start by taking a functional approach. We prove that (in some appropriate senses) a randomly chosen pattern avoiding permutation converges to Brownian excursion. Then we exploit this convergence to Brownian excursion to determine the distribution of the number and location of fixed points of a random pattern avoiding permutation.

Notation

Throughout this chapter we use the following definition of a Dyck path.

Definition 1.1.1. *A* **Dyck path**, γ, *of length* $2n$ *is a piecewise linear function defined by the linear interpolation of points* $\{(x, \gamma(x))\}_{x=0}^{2n}$ *that satisfy the following conditions:*

- $\gamma(0) = \gamma(2n) = 0$
- $\gamma(x) \geq 0$ *for all* $x \in (0, 2n)$, *and*
- $|\gamma(x+1) - \gamma(x)| = 1$ *for all integers* $x \in \{0, 1, \ldots, 2n-1\}$.

We dente the set of such paths by $Dyck^{2n}$.

Definition 1.1.2. *An* **excursion** *in a Dyck Path starting at* x *with height* h *and length* l *is a path interval* $\gamma([x, x+l])$ *such that*

- $\gamma(x) = \gamma(x+l) = h - 1$
- $\gamma(x+1) = \gamma(x+l-1) = h$ *and*
- $l = \min\{j \geq 1 : \ \gamma(x+j) = h-1\}$.

Note that there are n excursions in a Dyck Path of length $2n$ as there is one excursion that begins with every upstep. Based on this correspondence we say the ith excursion is the one that begins with the ith upstep.

Definition 1.1.3. *For a Dyck path* $\gamma \in Dyck^{2n}$, *define the following:*

- $exc(i) :=$ *the* i*th excursion.*
- $v_i :=$ *the position after the* i*th up step, or one more than the start of* $exc(i)$.
- $h_i := \gamma(v_i) =$ *the height of the path after the start of* $exc(i)$.

Figure 1.2: A Dyck path in $\mathcal{D}_{10}$ with $v_6 = 8$, $h_6 = 4$, and $l_6 = 8$.

- l_i := *the length of the same excursion.*

Figure 1.2 illustrates these definitions for a particular γ. We change our notation slightly when we are dealing with a random path.

Definition 1.1.4. *For a Dyck path $\Gamma^n \in Dyck^{2n}$, chosen at random, we define the associated random variables as follows*

- $Exc^n(i)$:= *the ith excursion.*

- V_i^n := *the position after the ith up step, or 1 + the start of $Exc^n(i)$.*

- $H_i^n := \Gamma^n(V_i^n)$ = *the height of the path after the start of $Exc^n(i)$.*

- L_i^n := *the length of the same excursion.*

Using these definitions we show to define a bijection from Dyck^{2n} to **231**-avoiding permutations which will be useful in proving our results.

1.1.1 Convergence to Brownian Excursion

We start with a Dyck path γ of length $2n$. It is well known that the scaling limit of Dyck paths are Brownian excursion [10]. Using the bijection in Proposition 1.2.1, for each Dyck path γ we call its corresponding **231**-avoiding permutation σ_γ. A particular rescaling of the linear interpolation of the local maxima of $(i - \sigma_\gamma)$ closely approximates the rescaling of γ.

For a permutation $\sigma \in S_n$ the process, $(\sigma(i) - i, i \in [n])$, is referred to as the exceedance process. The mirror to this process, $(i - \sigma(i), i \in [n])$, is referred to as the anti-exceedance process. We also define the subset of local minima, $I = I(\sigma) \subset [n-1]$ as the set of points

$$\{i \in [n-1] : \sigma(i) < \sigma(i+1).\}$$

Restricting the anti-exceedance process to I gives the following.

Definition 1.1.5 (Maximal Anti-Exceedance Process)**.** *Fix $\sigma \in S_n$ and $I = I(\sigma)$. We define the point process*

$$M = \{i - \sigma(i), i \in I\}.$$

We extend M to all real values in $[0, n]$ by letting $M(0) = M(n) = 0$ with linear interpolation for $t \notin I \cup \{0, n\}$.

We state our theorem on convergence in terms of the maximal anti-exceedance process. For a random path $\Gamma^n \in \text{Dyck}^{2n}$ we have a corresponding random permutation $\sigma_{\Gamma^n} \in S_n(231)$. We denote the set of local minima of σ_{Γ^n} by I^n, and we denote the corresponding anti-exceedance process by M^n.

Theorem 1.1.1. *Let $\Gamma^n \in Dyck^{2n}$ be chosen uniformly at random. There exists $r > 0$ such that*

$$\mathbb{P}\left(\sup_{t\in[0,1]} \frac{1}{(2n)^{1/2}} |\Gamma^n(2nt) - M^n(nt)| > n^{-0.05}\right) < e^{-n^r}$$

*1.1.2 Fixed points of **231**-avoiding permutations*

We count the number of fixed points in an interval by

$$\theta_{[an,bn]}(\sigma) = |\{i \in [an, bn] : \ \sigma(i) = i\}|$$

where $0 \leq a < b \leq 1$. Miner and Pak proved that the expected number of fixed points in an interval of the form $[an, bn]$ is of order $n^{1/4}$. We prove that a typical **231**-avoiding permutation has on the order of $n^{1/4}$ fixed points. Moreover it allows us to calculate the distribution of

$$\frac{1}{n^{1/4}}\theta_{[an,bn]}(\sigma_{\Gamma^n})$$

where Γ^n is chosen uniformly at random from Dyck^{2n}.

Theorem 1.1.2. *Fix $0 < a < b < 1$, and $\epsilon > 0$. For any n sufficiently large and $\Gamma^n \in Dyck^{2n}$ chosen uniformly at random,*

$$\mathbb{P}\left(\left|\frac{1}{n^{1/4}}\theta_{[an,bn]}(\sigma_{\Gamma^n}) - \frac{1}{2^{7/4}\pi^{1/2}}\int_a^b \left(\frac{(2n)^{1/2}}{\Gamma^n(2nt)}\right)^{3/2} dt\right| > \epsilon\right) = o(1).$$

1.1.3 "Almost fixed points"

Perhaps the most interesting result in [16] is a phase transition it shows in

$$\mathbb{P}\left(\sigma(i) = \left\lfloor i - \left(\frac{i(n-i)}{n}\right)^{\alpha}\right\rfloor\right)$$

that occurs at $\alpha = 3/8$. In particular they show that

$$\mathbb{P}\left(\sigma(i) = \left\lfloor i - \left(\frac{i(n-i)}{n}\right)^{\alpha}\right\rfloor\right) \sim \begin{cases} Cn^{-3/4} & \text{if } \alpha \in (0, 3/8) \\ Cn^{3/2-2\alpha} & \text{if } \alpha \in [3/8, .5) \end{cases} \tag{1.1}$$

This result is particularly intriguing because it is not clear what is driving the phase transition. Miner and Pak say that their results on **231**-avoiding permutations "are extremely unusual, and have yet to be explained even on a qualitative level" [16]. We use a generalization of Theorem 1.1.2 to give an explanation of these results.

First we show that the difference is not the number of "almost" fixed points on a typical path. To make this precise we define

$$\theta^{K,\alpha}_{[an,bn]}(\sigma) = \left|\left\{i :\ \sigma(i) = i - \left\lfloor K\left(\frac{i(n-i)}{n}\right)^{\alpha}\right\rfloor\right\} \cap [an, bn]\right|.$$

Then we follow the proof of Theorem 1.1.2 very closely to show

Corollary 1.1.3. *Fix $0 < a < b < 1$, $K \in \mathbb{R}$, $\alpha \in [0, .5)$ and $\epsilon > 0$. Let $\Gamma^n \in Dyck^{2n}$ be chosen uniformly at random. For any n sufficiently large*

$$\mathbb{P}\left(\left|\frac{1}{n^{1/4}}\theta^{K,\alpha}_{[an,bn]}(\sigma_{\Gamma^n}) - \frac{1}{2^{7/4}\pi^{1/2}}\int_a^b \left(\frac{(2n)^{1/2}}{\Gamma^n(2nt)}\right)^{3/2} dt\right| > \epsilon\right) = o(1)$$

Thus for all K and α the distribution of the number of "almost" fixed points is the same as the distribution of the number of fixed points and we do not see the same phase transition that Miner and Pak observed.

But there is no inconsistency between our results and [16] in the regime $K > 0$ and $\alpha \in [3/8, .5)$. This is because a small number of permutations drive the probability that Miner and Pak calculate in (1.1). This is missed by our convergence in distribution. This small number of permutations are the ones σ_γ whose corresponding Dyck paths γ have height $\gamma(i) = \lfloor K(i(n-i)/n)^\alpha \rfloor$ for some $i \in [2an, 2bn]$.

As the density of these permutations becomes vanishingly small as $n \to \infty$, these permutations do not affect the limiting distribution of $n^{-1/4}\theta^{K,\alpha}_{[an,bn]}(\sigma)$ that we calculate. But these are the permutations that dominate the probabilities that Miner and Pak calculate.

1.2 A Bijection between Dyck Paths and 231-avoiding permutations

The total number of Dyck paths from 0 to $2n$ is given by C_n, the nth Catalan number. The number of **231**-avoiding permutations in S_n is also given by the nth Catalan number. Hence there is a bijection between the two sets. We now define a particular bijection that uses geometric properties of the path. Although we suspect this bijection does exist in the literature we are not sure where it does. For the sake of completeness we include a proof that it is a bijection here. For our purposes the most important geometric aspect of a Dyck path is an excursion.

Proposition 1.2.1. *For $\gamma \in Dyck^{2n}$ define σ_γ pointwise by*

$$\sigma_\gamma(i) = i + l_i/2 - h_i.$$

Then $\sigma \in S_n(231)$. Moreover, $(\gamma \mapsto \sigma_\gamma)$ is a bijection from $Dyck^{2n} \to S_n(231)$.

The proof of Proposition 1.2.1 will follow from the following lemmas.

Lemma 1.2.2. *For any Dyck path $\gamma \in Dyck^{2n}$, $\sigma_\gamma : [n] \to [n]$.*

Proof. Let u_i and d_i denote the number of up-steps and the number of down-steps, respectively, up until the beginning of the ith excursion and let $v_i = u_i + d_i$ denote the total number of steps until the beginning of the same excursion. Each up-step is the beginning of an excursion so $u_i = i - 1$. Moreover d_i is determined by h_i since $h_i = u_i - d_i$. For any path $\gamma \in \text{Dyck}^{2n}$, $\gamma(x) \leq x$. Therefore $0 \leq h_i = \gamma(v_i) \leq i$, hence $0 \leq i - h_i$. Moreover, $l_i/2$ counts the number of up-steps in the excursion. Only $n - i$ up-steps remain after the first i have occured so $l_i/2 - 1 + i \leq n$. Combining these inequalities gives:

$$1 \leq i - h_i + 1 \leq \sigma(i) \leq i + l_i/2 - 1 \leq n.$$

Hence σ maps $[n]$ into $[n]$. □

Lemma 1.2.3. *For any Dyck path γ and any $i < j$ either*

$$Exc(j) \subset Exc(i) \quad \textit{or} \quad Exc(i) \cap Exc(j) = \emptyset.$$

Proof. This follows from the definition of an excursion. □

Lemma 1.2.4. *For any Dyck path γ and any $i < j$ if*

$$Exc(j) \subset Exc(i) \quad \textit{then} \quad \sigma_\gamma(j) < \sigma_\gamma(i)$$

and if

$$Exc(j) \cap Exc(i) = \emptyset \quad \textit{then} \quad \sigma_\gamma(i) < \sigma_\gamma(j).$$

Proof. Let $1 \leq i < j \leq n$. By the previous lemma the jth excursion begins either before or after the ith excursion ends. The two cases to consider are when $j - i < l_i/2$ or $j - i \geq l_i/2$.

We first consider when $j - i < l_i/2$. For j in this region we have $h_j > h_i$ and $l_j < l_i$. Moreover $l_j/2 - l_i/2 < i - j$. Therefore

$$\sigma(j) - \sigma(i) = j - i + l_j/2 - l_i/2 - (h_j - h_i) \leq h_i - h_j < 0. \tag{1.2}$$

Now we consider when $j - i \geq l_i/2$. Since the path must return below h_i at the end of $Exc(i)$ then it needs at least $\max(0, h_j - h_i)$ up-steps after the the ith excursion ends to be at height h_j. Therefore $j - i \geq l_i/2 + \max(0, h_j - h_i)$. This gives

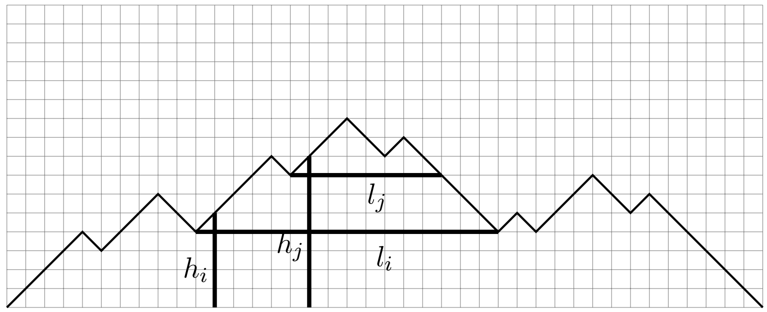

Figure 1.3: The jth excursion occurs during the ith excursion

$$\sigma(j) - \sigma(i) = j - i + l_j/2 - l_i/2 - (h_j - h_i) \geq 1 + l_j/2 \geq 1. \tag{1.3}$$

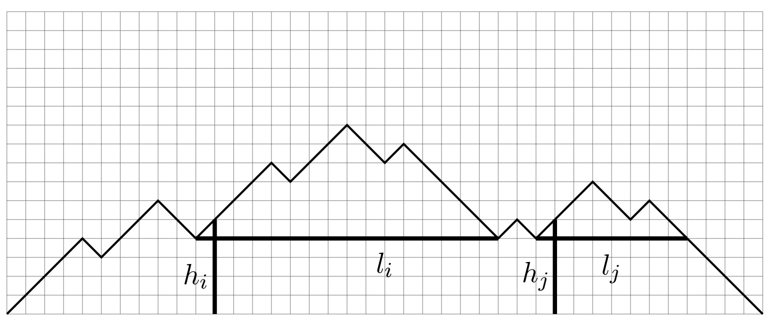

Figure 1.4: The jth excursion occurs after the ith excursion

In either case $\sigma(j) - \sigma(i) \neq 0$ so σ is in fact a bijection from $[n]$ to $[n]$.

□

Now we show that it is **231**-avoiding.

Proof of Proposition 1.2.1. If $\sigma \notin S_n(231)$, then there exists $i < j < k$ such that $\sigma(k) < \sigma(i) < \sigma(j)$. Note that $\sigma(k) < \sigma(i)$ implies the kth up-step occurs before the end of the

ith excursion. By Equations 1.2 and 1.3 the kth up-step occurs before the end of the ith excursion. Therefore the jth up-step also occurs before the end of the ith excursion which implies $\sigma(j) < \sigma(i)$, contradicting our assumption that $\sigma \notin S_n(231)$. Therefore we can safely conclude that σ must be **231**-avoiding.

All that remains is to show that $\sigma_\gamma \neq \sigma_{\gamma'}$ if $\gamma \neq \gamma'$. The quantity $\sigma_\gamma(i) - i$ attains a local minimum exactly when $l_i(\gamma) = 2$ and $\sigma(i) - i = -h_i + 1$. But $l_i = 2$ implies that $2i - h_i$ is a local maximum of the Dyck path. Hence there is one-to-one correspondence with local minima of $\sigma_\gamma(i) - i$ and local maxima of γ. A Dyck path is uniquely defined by the height and location of the local maxima. Hence the map from $\sigma_\gamma \to \gamma$ is well-defined. Therefore the map from Dyck^{2n} to $S_n(231)$ given by $(\gamma \to \sigma_\gamma)$ is a bijection. □

1.3 Convergence to Brownian Excursion

We develop a notion of convergence from permutations to Dyck paths similar to the work in [9]. As in the introduction, for $\Gamma^n \in \text{Dyck}^{2n}$ chosen uniformly at random, we let I^n denote the indices of the local minima of σ_{Γ^n} and M^n the linear interpolation of the restriction of this process to I^n.

Let $\hat{i}(x) = \max_{i \in I^n}\{i : i \leq x\}$. By the triangle inequality we have

$$\mathbb{P}\left(\sup_{t\in[0,1]} |\Gamma^n(2nt) - M^n(nt)| > n^{0.45}\right) \tag{1.4}$$

$$\leq \mathbb{P}\left(\sup_{t} \left|\Gamma^n(2nt) - \Gamma^n\left(2\hat{i}(nt)\right)\right| > n^{0.4}\right) \tag{1.5}$$

$$+ \mathbb{P}\left(\sup_{t} \left|\Gamma^n\left(2\hat{i}(nt)\right) - M^n\left(\hat{i}(nt)\right)\right| > n^{0.4}\right) \tag{1.6}$$

$$+ \mathbb{P}\left(\sup_{t} \left|M^n\left(\hat{i}(nt)\right) - M^n(nt)\right| > n^{0.4}\right) \tag{1.7}$$

We use lattice counting methods as found in [6]. Let $LP_{(a,b)}^{(c,d)}$ denote the total number of nonnegative lattice paths from (a, b) to (c, d) with steps of either $(1, 1)$ or $(1, -1)$. In terms of counting Dyck paths $LP_{(0,0)}^{(2n,0)} = C_n$, the nth Catalan number. More generally we have

$$LP_{(a,b)}^{(c,d)} = \binom{(c-a)}{\frac{1}{2}\left((c-a) + (d-b)\right)} - \binom{(c-a)}{\frac{1}{2}\left((c-a) + (d-b)\right) + b + 1}.$$

The following lemma will give control over the size of $|\Gamma^n(x)-\Gamma^n(y)|$ if we can bound $|x-y|$.

Lemma 1.3.1. *Let Γ^n be chosen uniformly from Dyck^{2n}. Let $\mathcal{C}$ denote the event that for any $0 \leq x < y \leq 2n$ such that $|x-y| < n^{0.6}$ and $|\Gamma^n(x)-\Gamma^n(y)| > n^{0.4}$. For sufficiently large n there exists $r > 0$ such that*

$$\mathbb{P}(\mathcal{C}) \leq \frac{1}{9}e^{-n^r}.$$

Proof. Let $h = \Gamma^n(\lfloor x \rfloor)$ and $h' = \Gamma^n(\lfloor y \rfloor)$. Then $h, h' \in \mathbb{N}$ and $|h-h'| \geq |\Gamma^n(x)-\Gamma^n(y)|-2$. Using the formula for counting lattice paths we have

$$\mathbb{P}\left(h-h'=m \mid \Gamma^n(\lfloor y \rfloor)=h'\right) = LP_{(0,0)}^{(\lfloor x \rfloor, h'+m)} LP_{(\lfloor x \rfloor, h'+m)}^{(\lfloor y \rfloor, h')} \left(LP_{(0,0)}^{(\lfloor y \rfloor, h')}\right)^{-1}.$$

By ▷ IX.1 on page 615 of [7] we see

$$\begin{aligned}
&\mathbb{P}(h-h'=m \mid \Gamma^n(\lfloor y \rfloor)=h') \\
&\qquad \leq \lfloor y \rfloor \binom{\lfloor x \rfloor}{(\lfloor x \rfloor - h' - m)/2} \binom{\lfloor y \rfloor - \lfloor x \rfloor}{(\lfloor y \rfloor - \lfloor x \rfloor - m)/2} \binom{\lfloor y \rfloor}{(\lfloor y \rfloor - h')/2}^{-1} \\
&\qquad \leq C\lfloor y \rfloor^2 \exp\left(-\frac{(h'+m)^2}{2\lfloor x \rfloor} - \frac{(m)^2}{2(\lfloor y \rfloor - \lfloor x \rfloor)} + \frac{(h')^2}{2\lfloor y \rfloor}\right) \\
&\qquad \leq Cn^2 \exp\left(-\frac{m^2}{2(\lfloor x \rfloor - \lfloor y \rfloor)}\right).
\end{aligned}$$

If $|y-x| < n^{0.6}$ and $|m| > n^{0.4}$ then the above expression is bounded by $n^2\exp(-n^{0.2}/2)$. The height of Γ^n is bounded by n so there are at most n possibilities for m and n possiblilities for h'. Moreover there are at most $4n^2$ possible pairs $(\lfloor x \rfloor, \lfloor y \rfloor)$, so for all sufficiently large n the union bound gives

$$\begin{aligned}
&\mathbb{P}\left(|\Gamma^n(x)-\Gamma^n(y)| > n^{0.4} \big| |x-y| < n^{0.6}\right) \\
&\quad \leq \sum_{0 \leq j < k \leq 2n, |j-i| < n^{0.6}} \sum_{0 \leq h' \leq n} \sum_{n^{0.4} \leq |m| \leq n} Cn^2 \exp\left(-\frac{m^2}{2(\lfloor x \rfloor - \lfloor y \rfloor)}\right) \\
&\quad \leq 4Cn^6 \exp(-n^{-0.2}/2) \\
&\quad \leq \exp(-n^{0.1}).
\end{aligned}$$

Any $r \in (0, 0.1)$ will suffice.

□

Now we show that $|\hat{i}(x) - x|$ will typically be small enough so that we may apply Lemma 1.3.1. Consider the event

$$\mathcal{C} : \left\{ \sup_{x \in [0,n]} \left(x - \hat{i}(x) \right) > n^{0.1} \right\}.$$

Lemma 1.3.2. *Let Γ^n be chosen uniformly from Dyck^{2n}. For n sufficiently large there exists $r > 0$ such that*

$$\mathbb{P}(\mathcal{C}) \leq \frac{1}{9} e^{-n^r}.$$

Proof. Based on our bijection $(nt) - \hat{i}(nt) > n^{0.1}$ if the corresponding Dyck path has a run of $n^{0.1}$ increases. The probability that a Dyck path has a run of $n^{0.1}$ increases beginning with the ith upstep is given by

$$LP^{(2n,0)}_{(a+n^{0.1}, b+n^{0.1})} \left(LP^{(2n,0)}_{(a,b)} \right)^{-1} \leq (3/4)^{n^{0.1}-2}.$$

There are at most n runs of increases so by the union bound every run of increases is less than $n^{0.1}$ with probability at least $1 - n(3/4)^{n^{0.1}-2}$. Then we may conclude for fixed $r > 0$ and every n sufficiently large

$$\mathbb{P}(\mathcal{C}) \leq n(3/4)^{\lfloor n^{0.1} \rfloor - 1} \leq e^{-n^r}$$

for some $r > 0$.

□

Lemma 1.3.3. *Let $\Gamma^n \in$ Dyck^{2n} be chosen uniformly at random. There exists constant $r > 0$ such that for n sufficiently large*

$$\mathbb{P}\left(\sup_{t \in [0,1]} |\Gamma^n(2nt) - M^n(nt)| > \sqrt{2} n^{0.45} \right) \leq e^{-n^r}.$$

Proof. We prove this lemma by showing there exists a constant $r > 0$ such that for n sufficiently large lines 1.5, 1.6, and 1.7 are each bounded by $\frac{1}{3} e^{-n^r}$.

From ▷ IX.1 on page 615 of [7] there exists a constant $s_1 > 0$ such that

$$\mathbb{P}\left(\sup_t \Gamma^n(2nt) > n^{0.51}\right) < \frac{1}{9}e^{-n^{s_1}}.$$

Let s_2 and s_3 denote the constants from Lemmas 1.3.2 and 1.3.1 respectively. Finally let $r = \min(s_1, s_2, s_3)$.

By Lemma 1.3.2

$$\begin{aligned}
&\mathbb{P}\left(\sup_t \left|\Gamma^n(2nt) - \Gamma^n\left(2\hat{i}(nt)\right)\right| > n^{0.4}\right) \\
&\quad \leq \mathbb{P}\left(\sup_t \left|\Gamma^n(2nt) - \Gamma^n\left(2\hat{i}(nt)\right)\right| > n^{0.4} \middle| \mathcal{C}^c\right) + \mathbb{P}\left(\mathcal{C}\right) \\
&\quad \leq \mathbb{P}\left(\sup_t \left|\Gamma^n(2nt) - \Gamma^n\left(2\hat{i}(nt)\right)\right| > n^{0.4} \middle| \mathcal{C}^c\right) + \frac{1}{9}e^{-n^r}.
\end{aligned}$$

By Lemma 1.3.1 there exists $s_2 > 0$ such that

$$\mathbb{P}\left(\sup_t \left|\Gamma^n(2nt) - \Gamma^n\left(2\hat{i}(nt)\right)\right| > n^{0.4} \middle| \mathcal{C}^c\right) < \frac{1}{9}e^{-n^r}$$

giving

$$\mathbb{P}\left(\sup_t \left|\Gamma^n(2nt) - \Gamma^n\left(2\hat{i}(nt)\right)\right| > n^{0.4}\right) < \frac{1}{3}e^{-n^r}.$$

This handles line 1.5.

By definition, $H_i^n = \Gamma^n(2i - H_i^n)$. Also $i \in I^n$ so

$$M^n\left(\hat{i}(nt)\right) = H^n_{\hat{i}(nt)} = \Gamma^n\left(2n\hat{i}(nt) - H^n_{\hat{i}(nt)}\right).$$

By Lemma 1.3.1, conditioned on $H^n_{\hat{(nt)}} < n^{0.51}$

$$\mathbb{P}\left(\sup_t \left|\Gamma^n\left(2\hat{i}(nt)\right) - \Gamma^n\left(2n\hat{i}(nt) - H^n_{\hat{i}(nt)}\right)\right| > n^{0.4} \middle| \{\sup_t H^n_{\hat{i}(nt)} < n^{0.51}\}\right) < \frac{1}{9}e^{-n^r}.$$

Then we have

$$\mathbb{P}\left(\sup_t \left|\Gamma^n\left(2\hat{i}(nt)\right) - \Gamma^n\left(2\hat{i}(nt) - H^n_{\hat{i}(nt)}\right)\right| > n^{0.4}\right) < \frac{1}{9}e^{-n^r} + \frac{1}{9}e^{-n^r}$$

$$\frac{1}{3}e^{-n^r}$$

giving the proper bound for line 1.6

Lastly we consider the quantities

$$\hat{j}(nt) := \inf_{j \in I^n} \{j : j \geq nt\}$$

and

$$\hat{k}(nt) := \left\{ k \in I^n : \hat{i}(nt) \leq k \leq \hat{j}(nt) \text{ and } \left| M^n(k) - M^n\left(\hat{i}(nt)\right)\right| \text{ is maximal } \right\}.$$

By definition

$$\left|M^n(nt) - M^n\left(\hat{i}(nt)\right)\right| \leq \left|M^n\left(\hat{k}(nt)\right) - M^n\left(\hat{i}(nt)\right)\right|.$$

Moreover, $\hat{k}(nt)$ and $\hat{i}(nt)$ are in I^n so

$$\mathbb{P}\left(\sup_{t\in[0,1]}\left|M^n\left(\hat{k}(nt)\right) - M^n\left(\hat{i}(nt)\right)\right| > n^{0.4}\right)$$

$$= \mathbb{P}\left(\sup_{t\in[0,1]}\left|\Gamma^n\left(2\hat{k}(nt) - H^n_{\hat{k}(nt)}\right) - \Gamma^n\left(2\hat{i}(nt) - H^n_{\hat{i}(nt)}\right)\right| > n^{0.4}\right).$$

By Lemma 1.3.2, $\sup_{t\in[0,1]}\left|2\hat{k}(nt) - 2\hat{i}(nt)\right| > 2n^{0.1}$ with probability at most $\frac{1}{9}e^{-n^r}$. With probability $\frac{1}{9}e^{-n^{s_3}}$, $H^n_{\hat{i}(nt)}$ and $H^n_{\hat{k}(nt)}$ are greater than $n^{0.51}$. Let B denote the union of these two events

$$\{\sup_{t\in[0,1]}\left|2\hat{k}(nt) - H^n_{\hat{k}(nt)} - 2\hat{i}(nt) + H^n_{\hat{i}(nt)}\right| > n^{0.6}\}\bigcup\{\sup_{t\in[0,1]}\left|2\hat{k}(nt) - 2\hat{i}(nt)\right| > 2n^{0.1}\}.$$

By the union bound we have

$$\mathbb{P}(B) < \frac{1}{9}e^{-n^r} + \frac{1}{9}e^{-n^r}.$$

Conditioned on B^c we may apply Lemma 1.3.1 to show that

$$\mathbb{P}\left(\left|\Gamma^n\left(2\hat{k}(nt) - H^n_{\hat{k}(nt)}\right) - \Gamma^n\left(2\hat{i}(nt) - H^n_{\hat{i}(nt)}\right)\right| > n^{0.4}\middle|B^c\right)$$

$$\leq \frac{1}{9}e^{-n^r}.$$

Removing the conditioning we have

$$\mathbb{P}\left(\left|\Gamma^n\left(2\hat{k}(nt) - H^n_{\hat{k}(nt)}\right) - \Gamma^n\left(2\hat{i}(nt) - H^n_{\hat{i}(nt)}\right)\right| > n^{0.4}\right) \leq \frac{1}{9}e^{-n^r} + \frac{1}{9}e^{-n^r} + \frac{1}{9}e^{-n^r}.$$

The union bound gives

$$\begin{aligned}\mathbb{P}&\left(\left|\Gamma^n\left(2\hat{k}(nt)-H^n_{\hat{k}(nt)}\right)-\Gamma^n\left(2\hat{i}(nt)-H^n_{\hat{i}(nt)}\right)\right|>n^{0.4}\right)\\&\leq\frac{1}{3}e^{-n^r}+13e^{-n^r}+13e^{-n^r}\\&\leq\frac{1}{3}e^{-n^r},\end{aligned}$$

properly bounding line 1.7.

Combining the bounds for lines 1.5 through 1.7 gives the desired bound for the lemma.

□

Scaling the expression in Lemma 1.3.3 gives Theorem 1.1.1.

1.4 Fixed Points for 231-avoiding permutations

For a 231-avoiding permutation $\sigma \in S_n(231)$, let $\theta_I(\sigma)$ denote the number of fixed points of σ contained in the subset $I \subset [n]$. Based on our bijection from Section 1.2, for a $\gamma \in \mathrm{Dyck}^{2n}$ and $\sigma = \sigma_\gamma$ and $\sigma(i) = i$ precisely when $l_i/2 = h_i$.

Theorem 1.4.1. *Fix $0 < a < b < 1$ and $\epsilon > 0$. Let Γ^n be chosen uniformly at random from Dyck^{2n}. Then*

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\frac{1}{n^{1/4}}\theta_{[an,bn]}(\sigma_{\Gamma^n})-\frac{1}{2\pi^{1/2}}\int_a^b\left(\frac{n^{1/2}}{\Gamma^n(2nt)}\right)^{3/2}dt\right|>\epsilon\right)=0.$$

We set up the notation necessary to prove Theorem 6.1. We break the interval $[an, bn]$ up into subintervals of size about $n^{0.9}$. In each of these intervals we will estimate the expected number of fixed points using the height of Dyck path at the start of the interval. Then we will bound the variance to show that with high probability the number of fixed points is close to the expected value.

Label the intervals $I_k = [a_k, b_k)$ for $k \in [0, \cdots, K-1]$, where $K = \lfloor (b-a)n^{0.1} \rfloor$ and

$$a_k = \lfloor an + (k/K)(bn - an) \rfloor \text{ and } b_k = a_{k+1}.$$

Denote a sequence of heights $\alpha = \{\alpha_k^n\}_{k=0}^{K-1}$ and define

$$\Omega^n(\alpha) := \bigcap_{k=0}^{K-1} \left\{\gamma \in \mathrm{Dyck}^{2n} | \gamma(v_{a_k}) = \alpha_k^n\right\}$$

where v_{a_k} is the total number of steps to and including the a_kth up step. Note that $\Omega^n(\alpha) \cap \Omega^n(\alpha') = \emptyset$ if $\alpha \neq \alpha'$. Let $\mathcal{A}$ denote the collection of all α.

Definition 1.4.1 (A Proper Subset of Dyck^{2n})**.** *We say a sequence of heights $\alpha = \{\alpha_k^n\}_{k=0}^K$ is proper if the following are satisfied:*

- $n^{0.499} < \alpha_k^n < n^{0.501}$ *for all* $0 \leq k \leq K$ *and*
- $|\alpha_k^n - \alpha_{k+1}^n| < n^{0.451}$ *for* $0 \leq k < K$*.*

We say $\Omega^n(\alpha)$ is proper if α is proper.

Definition 1.4.2. *Recalling Definition 1.1.3, we define the random variables for a random path $\Gamma^n \in$ Dyck^{2n} :*

- $V_i^n :=$ *number of steps up to and including the ith up step.*
- $H_i^n := \Gamma^n(V_i^n)$.
- $L_i^n :=$ *the length of the ith excursion.*

Let $\mathcal{B}_n$ denote the collection of proper $\alpha \in \mathcal{A}$. Most $\Gamma^n \in$ Dyck^{2n} will be in some proper $\Omega^n(\alpha)$.

Lemma 1.4.2. *For n sufficiently large, and Γ^n be chosen uniformly at random from Dyck^{2n},*

$$\mathbb{P}\left(\Gamma^n \in \bigcup_{\alpha \in \mathcal{B}_n} \Omega^n(\alpha)\right) > 1 - o(1).$$

Moreover

$$\mathbb{P}\left(\bigcap_{i \in [an,bn]} \left\{n^{0.49} < H_i^n < n^{0.51}\right\} \middle| \Omega^n(\alpha)\right) > 1 - e^{-n^{0.0001}}$$

for all proper $\Omega^n(\alpha)$.

Proof. The first statement follows from Lemmas 1.5.10 and 1.5.11. The second statement follows by applying Lemma 1.5.12 to the intervals I_k for $0 \leq k < K$. $\square$

For a fixed sequence of heights α, let $\hat{k}(x) = \sup_k\{2a_k - \alpha_k^n \leq x\}$. We define the following function $\rho_\alpha : [2an, 2bn] \to [0, n]$

$$\rho_\alpha(x) := \alpha^n_{\hat{k}(x)}.$$

For $\gamma \in \Omega^n(\alpha)$, $v_{a_k} = 2a_k - \alpha_k^n$ and $\gamma(v_{a_k}) = \rho_\alpha(v_{a_k}) = \alpha_k^n$. For most $\gamma \in \Omega^n(\alpha)$, γ will be close to ρ_α.

Lemma 1.4.3. *Fix* $0 < a < b < 1$, *and* $\epsilon > 0$. *For all* n *sufficiently large,*

$$\max_{\alpha \in \mathcal{B}_n} \left\{ \mathbb{P}\left(\sup_{t \in [a,b]} \left| \left(\frac{n^{1/2}}{\rho_\alpha(2nt)}\right)^{3/2} - \left(\frac{n^{1/2}}{\Gamma^n(2nt)}\right)^{3/2} \right| > n^{-0.01} \middle| \Omega^n(\alpha) \right) \right\} < e^{-n^{0.001}}.$$

Proof. Let $\hat{k} = \hat{k}(2nt)$. By definition $v_{a_{\hat{k}}} < 2nt < v_{a_{\hat{k}+1}}$ so

$$|2nt - v_{a_{\hat{k}}}| < |2a_{\hat{k}} - 2a_{\hat{k}+1} - \alpha^n_{\hat{k}} + \alpha^n_{\hat{k}+1}| < 2n^{0.9} + n^{0.451} < 3n^{0.9}.$$

We can expand the proof of Lemma 1.4.2 to include deviations bounds for all $i \in (a_k, a_{k+1})$ for $0 \leq k < K$. In particular we have for $t < 3$,

$$\mathbb{P}(|\Gamma^n(v_{a_k} + tn^{0.9}) - \Gamma^n(v_{a_k})| > n^{0.46} | \Omega^n(\alpha)) < e^{-n^{0.001}}.$$

By Lemma 1.5.12, $\Gamma^n(2nt) > n^{0.49}$ with probability $1 - e^{-0.001}$, so with probability at least $1 - e^{-n^{0.001}}$

$$\begin{aligned}
\left| \left(\frac{n^{1/2}}{\rho_\alpha(2nt)}\right)^{3/2} - \left(\frac{n^{1/2}}{\Gamma^n(2nt)}\right)^{3/2} \right| &< \left| \left(\frac{n^{1/2}}{\Gamma^n(2nt) - n^{0.46}}\right)^{3/2} - \left(\frac{n^{1/2}}{\Gamma^n(2nt)}\right)^{3/2} \right| \\
&< \left(\frac{n^{1/2}}{\Gamma^n(2nt)}\right)^{3/2} \left|1 - \frac{1}{1 - n^{0.46}/\Gamma^n(2nt)}\right|^{3/2} \\
&< n^{0.015} n^{-0.03} \\
&< n^{-0.01}.
\end{aligned}$$

□

Lemma 1.4.4. *Fix $0 < a < b < 1$. For all n sufficiently large,*

$$\max_{\alpha\in\mathcal{B}_n}\left|n^{-1/4}\mathbb{E}\left[\theta_{[an,bn]}\middle|\Omega^n(\alpha)\right] - \frac{1}{2\pi^{1/2}}\int_a^b\left(\frac{n^{1/2}}{\rho_\alpha(2nt)}\right)^{3/2}dt\right| < n^{-0.001}.$$

Lemma 1.4.5. *Fix $0 < a < b < 1$. For all n sufficiently large,*

$$\max_{\alpha\in\mathcal{B}_n}\mathbf{Var}\left[\theta_{[an,bn]}\middle|\Omega^n(\alpha)\right] < n^{0.48}.$$

Because these bounds are uniform over all proper $\Omega^n(\alpha)$ we will drop the α where no confusion should arise. We delay the proofs of these two lemmas until after the proof of Theorem 1.4.1 as they are long and somewhat technical. Many more technical calculations are pushed to Section 1.5 to make the proofs of Lemmas 1.4.4 and 1.4.5 more readable.

Proof of Theorem 1.4.1. Fix a proper Ω^n. For all n sufficiently large and Γ^n chosen uniformly from Ω^n, by Lemma 1.4.5 and Chebyshev's inequality

$$\mathbb{P}\left(\left|\theta_{[an,bn]}(\sigma_{\Gamma^n}) - \mathbb{E}\left[\theta_{[an,bn]}|\Omega^n\right]\right| > n^{0.005}n^{0.24}\middle|\Omega^n\right) < n^{-0.01}.$$

With Lemma 1.4.4 we have

$$\mathbb{P}\left(\left|n^{-1/4}\theta_{[an,bn]} - \frac{1}{2\pi^{1/2}}\int_a^b\left(\frac{n^{1/2}}{\rho(2nt)}\right)^{3/2}dt\right| > n^{-0.005} + n^{-0.001}\middle|\Omega^n\right) < n^{-0.01}.$$

Combined with Lemma 1.4.3

$$\mathbb{P}\left(\left|n^{-1/4}\theta_{[an,bn]}(\sigma_\Gamma^n) - \frac{1}{2\pi^{1/2}}\int_a^b\left(\frac{n^{1/2}}{\Gamma^n(2nt)}\right)^{3/2}dt\right| > 2n^{-0.001}\middle|\Omega^n\right) = \Delta(\Omega^n) \qquad (1.8)$$

where $\Delta(\Omega^n) = o(1)$ uniformly for all proper Ω^n.

Now consider $\Gamma^n \in \text{Dyck}^{2n}$ chosen uniformly at random.

$$\mathbb{P}\left(\left|n^{-1/4}\theta_{[an,bn]}(\sigma_\Gamma^n) - \frac{1}{2\pi^{1/2}}\int_a^b \left(\frac{n^{1/2}}{\Gamma^n(2nt)}\right)^{3/2} dt\right| > 2n^{-0.001}\right)$$
$$\leq \mathbb{P}\left(\Gamma^n \notin \cup_{\alpha\in\mathcal{B}}\Omega^n(\alpha)\right) + \sum_{\alpha\in\mathcal{B}} \Delta(\Omega^n(\alpha))\mathbb{P}(\Gamma^n \in \Omega^n(\alpha))$$
$$\leq o(1)$$

by Lemma 1.4.2.

□

1.4.1 Proof of Lemma 1.4.4

For $i \in [an, bn]$ we have that $\theta_i := \theta_i(\sigma_{\Gamma^n})$ is a 0-1 valued random variable where

$$\mathbb{P}(\theta_i = 1) = \mathbb{P}(L_i^n/2 = H_i^n).$$

Let $I_k = I_k^{int} \cup I_k^{out}$ where I_k^{out} consists of the $2n^{0.6}$ values both directly after a_k or directly before a_{k+1} and I_k^{int} is the rest of I_k.

Lemma 1.4.6. *Fix* $0 < a < b < 1$. *For all proper* Ω^n *and for each* k, *and* $i \in I_k^{int}$.

$$\mathbb{E}[\theta_i|\Omega^n] = \frac{1}{2\pi^{1/2}(\alpha_k^n)^{3/2}}(1+\Delta),$$

where $\Delta = \Delta(i, k, \Omega^n) = o(n^{-0.01})$ *is uniformly bounded for all choices* i, k *and proper* Ω^n.

Proof. For each k the and each i in I_k^{int} the conditions for Lemma 1.5.8 are satisfied since Ω^n is proper. Therefore

$$\mathbb{E}[\theta_i|\Omega^n] = \frac{1}{2\pi^{1/2}(\alpha_k^n)^{3/2}}(1+\Delta)$$

as desired.

□

For I_k^{out} we look at $\mathbb{E}[\theta_{I_k^{out}}|\Omega^n]$ as a whole rather than computing $\mathbb{E}[\theta_i|\Omega^n]$ for each individual i.

Lemma 1.4.7. *Fix* $0 < a < b < 1$ *and proper* Ω^n. *For all* $\gamma \in \Omega^n$

$$\sum_k \theta_{I_k^{out}} \le 5n^{0.21}.$$

Proof. For each k, I_k^{out} consists of two intervals of length $2n^{0.6}$ which can be covered by less than $5n^{0.6}/n^{049}$ subintervals of length $n^{0.49}$. As Ω^n is proper, then $h_i > n^{0.49}$ for $\imath \in [an, bn]$. Then by Lemma 1.5.13 each of the subintervals has at most one fixed point. Then $\theta_{I_k^{out}} \le 5n^{0.11}$ for each $0 \le k < K < n^{0.1}$. Adding them up proves the lemma.

□

Lemma 1.4.8. *For fixed* $0 < a < b < 1$ *and proper* Ω^n,

$$\mathbb{E}[\theta_{[an,bn]} | \Omega^n] = (1 + \Delta) \sum_{j=\lfloor an \rfloor}^{\lfloor bn \rfloor} \frac{1}{2\pi^{1/2} \rho(V_j^n)^{3/2}},$$

where $\Delta = o(n^{-0.01})$ *is uniformly bounded for all proper* Ω^n

Proof. By linearity of expectation:

$$\mathbb{E}[\theta_{[an,bn]} | \Omega^n] = \sum_{k=0}^{K-1} \sum_{i' \in I_k} \mathbb{E}[\theta_{i'} | \Omega^n] = \sum_{k=0}^{K-1} \sum_{i=0}^{|I_k|-1} \mathbb{E}[\theta_{a_k+i} | \Omega^n]$$

For each k, and $a_k + i \in I_k^{int}$, we can apply Lemma 1.4.6 to conclude

$$\mathbb{E}\left[\theta_{a_k+i} \middle| \Omega^n\right] = \frac{1}{2\pi^{1/2} (\alpha_k^n)^{3/2}} (1 + \Delta(i, k, \Omega^n))$$

where $\Delta(i, k, \Omega^n) = o(n^{-0.01})$ is uniformly bounded for all i, k and proper Ω^n.

By Lemma 1.5.12 we know the paths are high enough to apply Lemma 1.4.7 to show that $\mathbb{E}[\sum_k \theta_{I_k^{out}}] < n^{0.22}$. On the other hand $\alpha_k^n < n^{0.51}$ implies

$$\begin{aligned}
\sum_k \mathbb{E}[\theta_{I_k^{int}} | \Omega^n] &= \sum_k \sum_{i \in I_k^{int}} \frac{1}{2\pi^{1/2} (\alpha_k^n)^{3/2}} (1 + \Delta(i, k, \Omega^n)) \\
&> \sum_k \sum_{i \in I_k^{int}} n^{-0.765} \\
&> n^{0.23}
\end{aligned}$$

so the contribution from $\sum_k \mathbb{E}[\theta_{I_k^{out}}|\Omega^n]$ is dominated by $\sum_k \mathbb{E}[\theta_{I_k^{int}}|\Omega^n]$. Then by Lemma 1.4.6 and 1.4.7

$$\begin{aligned}\mathbb{E}[\theta_{[an,bn]}|\Omega^n] &= \mathbb{E}[\theta_{I_k^{out}}|\Omega^n] + \sum_k \mathbb{E}[\theta_{I_k^{int}}|\Omega^n] \\ &= 5n^{0.21} + \sum_k \frac{|I_k^{int}|}{2\pi^{1/2}(\alpha_k^n)^{3/2}}(1+\Delta(k,\Omega^n)) \\ &= (1+\Delta(\Omega^n))\sum_k \frac{|I_k^{int}|}{2\pi^{1/2}(\alpha_k^n)^{3/2}}\end{aligned}$$

where $\Delta(k,\Omega^n) = o(n^{-0.01})$ is uniformly bounded over k and proper Ω^n and $\Delta(\Omega^n) = o(n^{0.01})$ is uniformly bounded over Ω^n.

For each k, $|I_k| = (1+O(n^{-0.3}))|I_k^{int}|$ by the definitions of I_k and I_k^{int}. Then the above expression becomes

$$\begin{aligned}\mathbb{E}[\theta_{[an,bn]}|\Omega^n] =&(1+\Delta'(\Omega^n))\sum_k \frac{|I_k|}{2\pi^{1/2}(\alpha_k^n)^{3/2}} \\ =&(1+\Delta'(\Omega^n))\sum_k \sum_{j\in I_k} \frac{1}{2\pi^{1/2}(\alpha_k^n)^{3/2}}.\end{aligned}$$

with $\Delta'(\Omega^n) = o(n^{-0.01})$ uniformly bounded over all proper Ω^n. For $j \in I_k$, $\rho(V_j^n) = \alpha_k^n$, finishing the proof.

□

Proof of Lemma 1.4.4. By Lemma 1.4.8 we can write the conditional expectation of $\theta_{[an,bn]}$ as

$$\mathbb{E}[\theta_{[an,bn]}|\Omega^n] = (1+\Delta)\sum_{i=\lfloor an\rfloor}^{\lfloor bn\rfloor} \frac{1}{2\pi^{1/2}\rho(V_i^n)^{3/2}}$$

where $\Delta = o(n^{-0.01})$ is uniformly bounded over all proper Ω^n. Converting the sum into an integral we have

$$\mathbb{E}[\theta_{[an,bn]}|\Omega^n] = (1+\Delta)\int_{an}^{bn} \frac{1}{2\pi^{1/2}\rho(V_{\lfloor u\rfloor}^n)^{3/2}}du.$$

The change of variables $nt = u$ gives

$$\mathbb{E}[\theta_{[an,bn]}|\Omega^n] = (1+\Delta)\int_{a}^{b} \frac{1}{2\pi^{1/2}\rho(V_{\lfloor nt\rfloor}^n)^{3/2}}ndt.$$

By properness of Ω^n, $|\rho(V^n_{\lfloor nt \rfloor}) - \rho(2nt)| < n^{0.451}$ and $\rho(2nt) > n^{0.499}$ for $t \in [a, b]$ so

$$\frac{1}{\rho(V^n_{nt})^{3/2}} = (1 + o(n^{-0.001}))\frac{1}{\rho(2nt)^{3/2}}.$$

Scaling by $n^{1/4}$ completes the proof.

□

1.4.2 Proof of Lemma 1.4.5

Now that we have the conditional expectation $\mathbb{E}[\theta^{int}_{[an,bn]}|\Omega^n]$, we will bound the conditional variance, $\mathbf{Var}[\theta_{[an,bn]}|\Omega^n]$.

Our basic variance equation is

$$\mathbf{Var}[\theta_{[an,bn]}|\Omega^n] = \sum_{i,j} \mathbb{E}[\theta_i \theta_j|\Omega^n] - \mathbb{E}[\theta_i|\Omega^n]\mathbb{E}[\theta_j|\Omega^n].$$

The key to bounding the conditional variance for a proper Ω^n is understanding $\mathbb{E}[\theta_i\theta_j|\Omega^n]$ for various ranges of i and j. We cover $[an, bn]^2$ with $\cup_{l=1}^5 B_l$ where each B_l is defined as follows:

- $B_1 = \cup_k \cup_{k'} I^{out}_k \times I^{out}_{k'}$,
- $B_2 = \cup_k \cup_{k'} \{I^{int}_k \times I^{out}_{k'}\} \bigcup \{I^{int}_{k'} \times I^{out}_k\}$,
- $B_3 = \cup_k \cup_{k' \neq k} I^{int}_k \times I^{int}_{k'}$,
- $B_4 = \cup_k I^{int}_k \times \{j \in I^{int}_k s.t. |j - i| \leq 2n^{0.6}\}$,
- $B_5 = \cup_k I^{int}_k \times \{j \in I^{int}_k s.t. |j - i| > 2n^{0.6}\}$.

For each B_l we will show that

$$\sum_{(i,j)\in B_l} \mathbb{E}[\theta_i\theta_j|\Omega^n] - \mathbb{E}[\theta_i|\Omega^n]\mathbb{E}[\theta_j|\Omega^n] = o(n^{0.48}).$$

Hence the total variance is $o(n^{0.48})$.

Consider the property

$$P = \bigcap_{i \in [an,bn]} \{n^{0.49} < H_i^n < n^{0.51}\}.$$

The following allows us to consider only paths that satisfy this property when computing the variance.

Lemma 1.4.9. *Fix $0 < a < b < 1$. For n sufficiently large and proper Ω^n*

$$\sum_{(i,j) \in [an,bn]^2} \mathbb{E}[\theta_i \theta_j | \Omega^n] = (1 + \Delta) \sum_{(i,j) \in [an,bn]^2} \mathbb{E}[\theta_i \theta_j \mathbf{1}_P | \Omega^n]$$

with $\Delta = o(1)$ uniformly bounded for all proper Ω^n.

Proof. Since $\theta_i \theta_j \mathbf{1}_{PC} < \mathbf{1}_{PC}$, by Lemma 1.5.12 we have

$$\mathbb{E}[\theta_i \theta_j \mathbf{1}_{PC} | \Omega^n] \leq \mathbb{E}[\mathbf{1}_{PC} | \Omega^n] \leq n^3 \exp^{-n^{0.001}}.$$

Noting that $\theta_i \theta_j = \theta_i \theta_j \mathbf{1}_P + \theta_i \theta_j \mathbf{1}_{PC}$ and taking expectation gives

$$\mathbb{E}[\theta_i \theta_j | \Omega^n] \leq \mathbb{E}[\theta_i \theta_j \mathbf{1}_P | \Omega^n] + \mathbb{E}[\mathbf{1}_{PC} | \Omega^n] \leq \mathbb{E}[\theta_i \theta_j \mathbf{1}_P | \Omega^n] + n^3 \exp(-n^{0.001}).$$

Summing over i and j completes the proof.

□

For the following series of lemmas we will assume the following standard hypotheses.

- Fix $0 < a < b < 1$.
- $\Omega^n \subset \text{Dyck}^{2n}$ is proper.
- Let n be large enough such that $n^5 e^{-n^{0.0001}} < n^{-0.47}$ and for every proper Ω^n we have that $n^2 \mathbb{P}(P^C | \Omega^n) < n^7 e^{-n^{0.001}} < e^{-n^{0.0001}}$.

Lemma 1.4.10. *Assuming the standard hypotheses,*

$$\sum_{(i,j) \in B_1} \mathbb{E}[\theta_i \theta_j | \Omega^n] < n^{0.47}.$$

Proof. The union bound gives

$$\sum_{(i,j)\in B_1} \mathbb{E}[\theta_i\theta_j|\Omega^n] \le \mathbb{P}(P^C|\Omega^n) + \sum_{(i,j)\in B_1} \mathbb{E}[\theta_i\theta_j \mathbf{1}_P|\Omega^n].$$

By Lemma 1.5.12, $\mathbb{P}(P^C|\Omega^n) = n^3 e^{-n^{0.08}}$. For fixed k' we may use Lemma 1.5.13 to show $\sum_{j\in I_{k'}^{out}} \mathbb{E}[\theta_j|*] < 2n^{0.6-0.49}$ no matter the conditions given by $*$. In particular we have

$$\begin{aligned}\sum_{j\in I_{k'}^{out}} \mathbb{E}[\theta_i\theta_j|\Omega^n] &\le \mathbb{P}(P^C|\Omega^n) + \sum_{j\in I_{k'}^{out}} \mathbb{E}[\theta_j|\Omega^n, P, \theta_i = 1]\mathbb{E}[\theta_i|\Omega^n, P] \\ &\le \mathbb{E}[\theta_i|\Omega^n]4n^{0.6-0.49}.\end{aligned}$$

Then

$$\begin{aligned}\sum_k \sum_{k'} \sum_{i\in I_k^{out}} \sum_{j\in I_{k'}^{out}} \mathbb{E}[\theta_i\theta_j \mathbf{1}_P|\Omega^n] &\le \sum_k \sum_{k'} \sum_{i\in I_k^{out}} \mathbb{E}[\theta_i|\Omega^n]2n^{0.6-0.49} \\ &\le \sum_k \sum_{k'} 4n^{0.22} \\ &\le n^{0.42}\end{aligned}$$

□

Lemma 1.4.11. *Assuming the standard hypotheses,*

$$\sum_{(i,j)\in B_2} \mathbb{E}[\theta_i\theta_j|\Omega^n] < 6n^{0.47}.$$

Proof. This follows the proof of Lemma 1.4.10 closely. By Lemma 1.5.13

$$\sum_k \sum_{k'} \sum_{i\in I_k^{out}} \sum_{j\in I_{k'}^{out}} \mathbb{E}[\theta_i\theta_j|\Omega^n] \le n^2\mathbb{P}(P^C) + \sum_k \sum_{k'} \sum_{i\in I_k^{out}} \mathbb{E}[\theta_i|\Omega^n]2n^{0.6-0.49}.$$

For each k and each $i \in I_k^{out}$, Lemma 1.5.8 and the properness of Ω^n imply that $\mathbb{E}[\theta_i|\Omega^n] < (n^{-0.499})^{3/2} < n^{-0.74}$. Then

$$\sum_k \sum_{k'} \sum_{i \in I_k^{out}} 3n^{0.11}\mathbb{E}[\theta_i|\Omega^n] \le 3n^{0.1}n^{0.1}n^{0.9}n^{0.11}n^{-0.74} < 3n^{0.47}.$$

Changing the roles of i and j and doubling the upper bounded completes the proof. □

Lemma 1.4.12. *Assuming the standard hypotheses,*

$$\sum_{i,j \in B_3} \mathbb{E}[\theta_i\theta_j|\Omega^n] - \mathbb{E}[\theta_i|\Omega^n]\mathbb{E}[\theta_j|\Omega^n] < n^{0.47}.$$

Proof. The flavor of this proof is somewhat different from the previous lemmas. Without loss of generality we may assume that $k < k'$.

If $\theta_i \mathbf{1}_P = 1$, then the corresponding ith excursion will end before the $a_{k'}$th excursion begins as

$$L_i^n/2 = H_i^n < n^{0.51} < 2n^{0.6}$$

and $a_{k'} > i + |I_k^{out}|$. Therefore, for $i \in I_k$ and $j \in I_{k'}$,

$$\mathbb{E}[\theta_i\theta_j|\Omega^n, P] = \mathbb{E}[\theta_i|\Omega^n, P]\mathbb{E}[\theta_j|\Omega^n, P]$$

and

$$\begin{aligned}
&\sum_{(i,j)\in B_3} (\mathbb{E}[\theta_i\theta_j|\Omega^n] - \mathbb{E}[\theta_i|\Omega^n]\mathbb{E}[\theta_j|\Omega^n]) \\
&\quad \le n^2\mathbb{P}(P^C|\Omega^n) + 2\sum_k \sum_{k'>k} \sum_{i\in I_k^{int}} \sum_{j \in I_{k'}} (\mathbb{E}[\theta_i\theta_j|\Omega^n, P] - \mathbb{E}[\theta_i|\Omega^n, P]\mathbb{E}[\theta_j|\Omega^n, P]) \\
&\quad \le n^5 \exp(-n^{0.001}) \\
&\quad < n^{0.47}
\end{aligned}$$

□

Lemma 1.4.13. *Assuming the standard hypotheses,*

$$\sum_{(i,j)\in B_4} \mathbb{E}[\theta_i\theta_j|\Omega^n] < n^{0.47}.$$

Proof. By Lemma 1.5.13

$$\sum_{|i-j|<2n^{0.6}} \mathbb{E}[\theta_i\theta_j|\Omega^n] \leq n^2\mathbb{P}(P^C) + 5n^{0.6-0.49}\mathbb{E}[\theta_i \mathbf{1}_P|\Omega^n].$$

For each k, $|I_k^{int}| \leq n^{0.9}$ so

$$\sum_k \sum_{i\in I_k^{int}} 5n^{0.11}n^{-0.73} \leq 5n^{0.1+0.9+0.11-0.735} < n^{0.47}.$$

□

The last possibility is the one which requires the most care.

Lemma 1.4.14. *Assuming the standard hypotheses,*

$$\sum_{(i,j)\in B_5} \mathbb{E}[\theta_i\theta_j\mathbf{1}_P|\Omega^n] < n^{0.47}.$$

Proof. We proceed in a manner similar to Lemmas 1.5.6 and 1.5.7. For $(i,j) \in B_5$, with $i \in I_k^{int}$.

$$\mathbb{E}[\theta_i\theta_j\mathbf{1}_P|\Omega^n] = \sum_{n^{0.49}<h<n^{0.51}} \sum_{n^{0.49}<h'<n^{0.51}} G(i,j,h,h',a_k,a_{k+1},\alpha_k^n,\alpha_{k+1}^n).$$

where

$$G(i,j,h,h',a_k,a_{k+1},\alpha_k^n,\alpha_{k+1}^n)$$
$$= C_{h-1}C_{h'-1}\left|\mathcal{E}_{v_{a_k},\alpha_k^n}^{2i-h,h}\right|\left|\mathcal{E}_{2i-h+2h-1,h-1}^{2j-h',h'}\right|\left|\mathcal{E}_{2j-h'+2h'-1,h'-1}^{v_{a_{k+1}},\alpha_{k+1}^n}\right|\left|\mathcal{E}_{v_{a_k},\alpha_k^n}^{v_{a_{k+1}},\alpha_{k+1}^n}\right|^{-1}.$$

For fixed i, j and k there are two cases to consider for values h and h'. One where we can use Lemma 1.5.2 for each each section of the path, and one where we bound G by $e^{-n^{0.001}}$ using Lemma 1.5.1.

Define the set of pairs of heights $D_{i,j,k}$ such that for $(h,h') \in D_{i,j,k}$, Lemma 1.5.2 is valid for each of the path sections. For $(h,h') \notin D_{i,j,k}$ the contribution to $\mathbb{E}[\theta_i\theta_j\mathbf{1}_P|\Omega^n]$ is bounded by $e^{-n^{0.001}}$. Otherwise

$$\mathbb{E}[\theta_i\theta_j\mathbf{1}_P|\Omega^n] \leq \sum_{(h,h')\in D_{i,j,k}} \sqrt{\frac{n^{0.9}}{i(j-i)(n^{0.9}-j)h^3h'^3}}F(i,j,h,h',\alpha_k^n,\alpha_{k+1}^n)$$

where

$$F(i,j,h,h',\alpha_k^n,\alpha_{k+1}^n) = \exp\left(-\frac{(h-\alpha_k^n)^2}{4i} - \frac{(h'-h)^2}{4(j-i)} - \frac{(\alpha_{k+1}^n-h')^2}{4(n^{0.9}-j)} + \frac{(\alpha_{k+1}^n-\alpha_k^n)^2}{4n^{0.9}}\right).$$

For h and $h' \in (n^{0.49}, n^{0.51})$ we may replace $\frac{1}{(hh')^{3/2}}$ with $n^{-1.47}$. By Lemma 1.5.15 there exist a large constant C such that

$$\sum_{h,h'} F(i,j,h,h',\alpha_k^n,\alpha_{k+1}^n) < C\sqrt{\frac{i(j-i)(n^{0.9}-j)}{n^{0.9}}}$$

uniformly over all choices of $(i,j) \in B_5$ and α_k^n and α_{k+1}^n from a proper sequence of heights

This gives

$$\sum_{(i,j)\in B_5} \mathbb{E}[\theta_i\theta_j\mathbf{1}_P|\Omega^n] \leq \sum_{(i,j)\in B_5} Cn^{-1.47} \leq Cn^{1.9}n^{-1.47} < n^{0.47}.$$

□

Proof of Lemma 1.4.5. Lemmas 1.4.9, 1.4.10, 1.4.11, 1.4.12, 1.4.13, and 1.4.14 com show that the variance $\mathbf{Var}[\theta_{[an,bn]}|\Omega^n] < 8n^{0.47}$. □

The following section contains various technical lemmas that will be used throughout the paper. The statements of the lemmas are similar to results found elsewhere, but modified for use in this paper.

1.5 Technical Lemmas

We begin with a useful Lemma that will help count non-negative lattice paths between points. Let $\mathcal{A}_n$ denote the set of points $\{i,m\} \in \mathbb{Z}^2$ such that $0 < n^{0.6} < i < n$ and $|m| < i^{0.6}$.

Lemma 1.5.1. *For* $\{i,m\} \in \mathcal{A}_n$.

$$\binom{2i-m}{i} = \frac{(1+\Delta(i,m))4^i}{2^m\sqrt{\pi i}} e^{-\frac{m^2}{4i}}$$

where $\Delta(i,m) = o(n^{-0.1})$ *is bounded uniformly over* i *and* m *in* $\mathcal{A}_n$.

For $i > n^{0.6}$ *and* $|m| > i^{0.6}$

$$\binom{2i-m}{i} \leq \frac{4^i}{2^m} e^{-n^{0.1}}.$$

Proof. This first equality follows from ▷ IX.1 on page 615 of Flajolet and Sedgewick [7]. For the second equality we let $m = i^{0.6} + r$ or $m = -i^{0.6} - r$ for some $r > 0$.

$$\binom{2i-m}{i} = \binom{2i-i^{0.6}}{i} \prod_{k=0}^{r-1} \frac{i - i^{0.6} - k}{2i - i^{0.6} - k}$$

$$\leq \frac{4^i}{2^{i^{0.6}+r}} e^{-i^{0.2}/4} \prod_{k=0}^{r-1} \frac{2i - 2i^{0.6} - 2k}{2i - i^{0.6} - k}.$$

$$\leq \frac{4^i}{2^m} e^{-n^{0.12}/4}.$$

A similar computation holds for $m = -i^{0.6} - r$.

□

Consider a lattice path starting at (v_0, h_0). Recall Definition 1.1.3. We may extend those definitions to general lattice paths with a slight modification. The definitions v_i and h_i remain the same, the position and the height after the ith up step from the start of the path. For l_i we do not necessarily have an excursion. If the path never returns below h_i at some time later than v_i then we say that $l_i = \infty$.

Lemma 1.5.2. *Suppose* $\{i, m\} \in \mathcal{A}_n$ *and* $h = h_0 + m$. *The number of lattice paths from* (v_0, h_0) *to* (v_i, h) *is given by*

$$\binom{2(i-1)-(m-1)}{i-1} = \frac{4^i}{2^{m+1}\sqrt{\pi i}} \exp\left(-\frac{m^2}{4i}\right)(1+\Delta(i,m))$$

where $\Delta(i,m)$ *as defined in Lemma 1.5.1.*

Proof. Let i and d denote the number of up and down steps respectively in a lattice path up to and including the ith up step. We denote the total number of steps by $v_i = i+d$. The change in height for that path is given by $h - h_0 = m = i - d$. Then $v_i = 2i - m$ counts the total number of steps. The second to last position of the path is $(x+2i-m-1, h_0+m-1)$ since the v_ith step is assumed to be an up step. Therefore the total number of lattices paths from (v_0, h_0) to $(v_0+2i-m-1, h_0+m-1)$ is counted by Lemma 1.5.1, giving the equation found in Lemma 1.5.2

□

Now that we can accurately count the number of lattice paths from one point to another we can count the number of non-negative paths between two points. For a pair of points (v_0, h_0) and (v_i, h_i) we let $\mathcal{E}_{v_0,h_0}^{v_i,h_i}$ denote the set of non-negative lattice paths ending with an up step between the two points.

Lemma 1.5.3. *For $i, m \in \mathcal{A}_n$, let $h_0 > n^{0.499}$, $v_i = v_0 + 2i - m$, and $h_i = h_0 + m$. Then*

$$\left|\mathcal{E}_{v_0,h_0}^{v_i,h_i}\right| = \frac{4^i}{2^{m+1}\sqrt{\pi i}} \exp\left(-\frac{m^2}{4i}\right)(1 + \Delta'(i,m))$$

where $\Delta'(i,m) = o(n^{-0.1})$ is uniformly bounded over $i, m \in A_n$.

Proof. We count using standard ballot counting arguments.

$$\left|\mathcal{E}_{v_0,h_0}^{v_0+2i-m,h_0+m}\right| = \binom{2i-m-1}{i-1} - \binom{2i-m-1}{i+h_0}.$$

For $h_0 > n^{0.49}$,

$$\binom{2i-m-1}{i+h_0} < \binom{2i-m-1}{i-1} \exp(-(h_0)^2/2i).$$

Moreover $(h_0)^2/2i > n^{0.07}$, so

$$\left|\mathcal{E}_{v_0,h_0}^{v_i,h_i}\right| = \binom{2i-m-1}{i-1}(1 + O(\exp(-n^{0.06}))) = \frac{4^i}{2^{m+1}\sqrt{\pi i}} \exp\left(-\frac{m^2}{4i}\right)(1 + \Delta'(i,m))$$

as desired.

□

For paths chosen uniformly from $\mathcal{E}_{v_0,h_0}^{v_j,h_j}$ for $0 < i < j$ we would like to know for various values of i and h how many of these path go through the point (v_i, h) after the ith up step. Given $\Gamma^n \in \mathcal{E}_{v_0,h_0}^{v_j,h_j}$ chosen uniformly at random what is the probability that $H_i^n = h$?

Lemma 1.5.4. *Fix v_0, h_0, h_j and h_j. Let X^n be chosen uniformly from $\mathcal{E}_{v_0,h_0}^{v_j,h_j}$. For $h > 0$ and $0 < i < j$,*

$$\mathbb{P}(H_i^n = h) = \left|\mathcal{E}_{v_0,h_0}^{2i-(h-h_0),h}\right| \left|\mathcal{E}_{2i-(h-h_0),h}^{v_j,h_j}\right| \left|\mathcal{E}_{v_0,h_0}^{v_j,h_j}\right|^{-1}.$$

Proof. Any path in $\mathcal{E}_{v_0,h_0}^{v_j,h_j}$ can be decomposed uniquely into a concatenation of two paths, one in $\mathcal{E}_{v_0,h_0}^{v_i,h_i}$ and the other in $\mathcal{E}_{v_i,h_i}^{v_j,h_j}$ for some appropriate values of v_i and h_i that satisfy $v_i = v_0 + 2i - (h_i - h_0)$. If $h_i = h$, then $v_i = v_0 + 2i - (h - h_0)$. The set $\left\{X^n \in \mathcal{E}_{v_0,h_0}^{v_j,h_j} | h_i = h\right\}$ is in bijection with $\mathcal{E}_{v_0,h_0}^{v_i,h} \times \mathcal{E}_{v_i,h}^{v_j,h_j}$. Then

$$\begin{aligned}\mathbb{P}(H_i^n = h) &= \left|\left\{\Gamma^n \in \mathcal{E}_{v_0,h_0}^{v_j,h_j} | H_i^n = h\right\}\right| \left|\mathcal{E}_{v_0,h_0}^{v_j,h_j}\right|^{-1} \\ &= \left|\mathcal{E}_{v_0,h_0}^{2i-(h-h_0),h}\right| \left|\mathcal{E}_{2i-(h-h_0),h}^{v_j,h_j}\right| \left|\mathcal{E}_{v_0,h_0}^{v_j,h_j}\right|^{-1}\end{aligned}$$

as desired. □

Lemma 1.5.5. *Fix v_0, h_0, v_j, with h_j such that $v_j - v_0 = 2j - (h_j - h_0)$ as above. Also fix $i, h > 0$ and $i < j - h$. For X^n chosen uniformly from $\mathcal{E}_{v_0,h_0}^{v_j,h_j}$,*

$$\mathbb{P}(L_i^n/2 = h | H_i^n = h) = C_{h-1} \left|\mathcal{E}_{2i-(h-h_0)+2h-1,h-1}^{v_j,h_j}\right| \left|\mathcal{E}_{2i-(h-h_0),h}^{v_j,h_j}\right|^{-1}.$$

Proof. Let $X^n \in \mathcal{E}_{v_0,h_0}^{v_j,h_j}$ also satisfy $H_i^n = h$. From Lemma 1.5.4 there are precisely

$$|\mathcal{E}_{v_0,h_0}^{2i-(h-h_0),h}||\mathcal{E}_{2i-(h-h_0),h}^{v_j,h_j}|$$

such paths. Each of these paths that satisfies $L_i^n/2 = h$ has a unique decomposition into three parts:

- A path $X_1^n \in \mathcal{E}_{v_0,h_0}^{2i-(h-h_0),h}$,

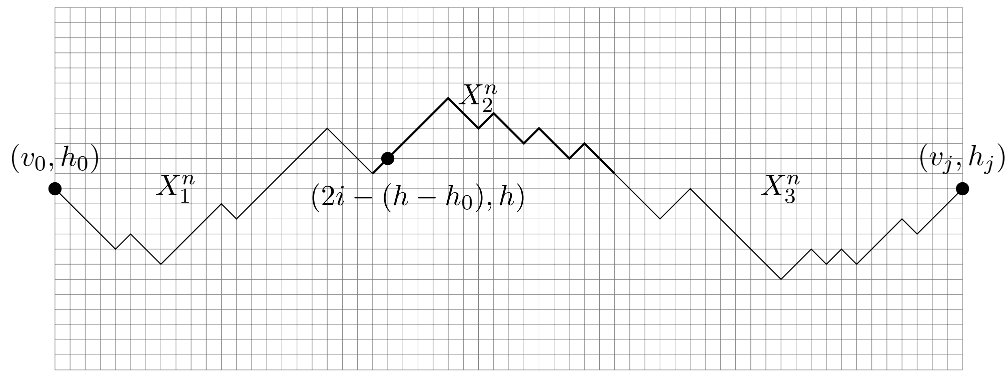

Figure 1.5: Decomposition of X^n into X_1^n, X_2^n, and X_3^n.

- an excursion X_2^n from $(2i-(h-h_0)_i, h)$ to $(2i-(h-h_0)+2h-2, h)$, staying above $h-1$,
- a single down step from $(2i-(h-h_0)+2h-2, h)$ to $(2i-(h-h_0)+2h-1, h-1)$,
- and a path $X_3^n \in \mathcal{E}^{v_j,h_j}_{2i-(h-h_0)+2h-1,h-1}$.

The choice of X_1^n, X_2^n, and X_3^n uniquely determines X^n. There are

$$\left|\mathcal{E}^{2i-(h-h_0),h}_{v_0,h_0}\right|, C_{h-1}, \text{ and } \left|\mathcal{E}^{v_j,h_j}_{2i-(h-h_0)+2h-1,h-1}\right|$$

such choices for X_1^n, X_2^n, and X_3^n respectively. Therefore

$$\begin{aligned}\mathbb{P}(L_i^n/2 = h | H_i^n = h) &= \left|\mathcal{E}^{2i-(h-h_0),h}_{v_0,h_0}\right| C_{h-1} \left|\mathcal{E}^{v_j,h_j}_{2i-(h-h_0)+2h-1,h-1}\right| \left(\left|\mathcal{E}^{2i-(h-h_0),h}_{v_0,h_0}\right| \left|\mathcal{E}^{v_j,h_j}_{2i-(h-h_0),h}\right|\right)^{-1} \\ &= C_{h-1} \left|\mathcal{E}^{v_j,h_j}_{2i-(h-h_0)+2h-1,h-1}\right| \left|\mathcal{E}^{v_j,h_j}_{2i-(h-h_0),h}\right|^{-1}\end{aligned}$$

□

Lemma 1.5.6. *For $X^n \in \mathcal{E}^{v_j,h_j}_{v_0,h_0}$ chosen uniformly at random and $0 < i < j$,*

$$\mathbb{P}(L_i^n/2 = H_i^n) = \sum_{h=\max(0,h_j-(j-i))}^{h_0+i} C_{h-1} \left|\mathcal{E}^{2i-(h-h_0),h}_{v_0,h_0}\right| \left|\mathcal{E}^{v_j,h_j}_{2i-(h-h_0)+2h-1,h-1}\right| \left|\mathcal{E}^{v_j,h_j}_{v_0,h_0}\right|^{-1}.$$

Proof. If $L_i^n/2 = H_i^n$ then there is some $h \in \mathbb{N}$ such that $\{H_i^n = h\} \cap \{L_i^n/2 = h\}$ occurs. Therefore

$$\{L_i^n/2 = H_i^n\} = \bigcup_h \{\{H_i^n = h\} \cap \{L_i^n/2 = h\}\}.$$

Luckily the event $\{H_i^n = h \cap L_i^n/2 = h\}$ is disjoint from $\{H_i^n = h' \cap L_i^n/2 = h'\}$ for $h \neq h'$. Then

$$\mathbb{P}\left(\bigcup_h \{H_i^n = h\} \cap \{L_i^n/2 = h\}\right) = \sum_h \mathbb{P}\left(\{H_i^n = h\} \cap \{L_i^n/2 = h\}\right).$$

If $h \notin (\max(0, H_j^n - (j-i), h_0 + i)$ then $\mathbb{P}(H_i^n = h) = 0$. Otherwise we have

$$\mathbb{P}\left(\{H_i^n = h\} \cap \{L_i^n/2 = h\}\right) = \mathbb{P}\left(L_i^n/2 = h | H_i^n = h\right) \mathbb{P}(H_i^n = h).$$

Combining Lemmas 1.5.4 and 1.5.5 provides the result.

□

Lemma 1.5.7. *Let $0 < i < j \leq O(n^{0.9})$ with $i > 2n^{0.6}$ and $j - i > 2n^{0.6}$, and let $h_0 \in (n^{0.499}, n^{0.501})$. Define m and m_j such that $h_j = h_0 + m_j$ and $h = h_0 + m$ where $|m_j| < \min(j^{0.6}, n^{0.451})$. Let $m_{max} = \min(i^{0.6}, m_j + (j-i)^{0.6}$ and $m_{min} = \max(-i^{0.6}, m_j - (j-i)^{0.6})$. For $m_{min} < m < m_{max}$*

$$\mathbb{P}(L_i^n/2 = h | H_i^n = h)\mathbb{P}(H_i^n = h) = \frac{\sqrt{j}}{4\pi\sqrt{i(j-i)(h_0)^3}} \exp\left(-\frac{(jm - im_j)^2}{4ij(j-i)}\right)(1 + \Delta)$$

where $\Delta = o(n^{-0.001})$ is bounded uniformly of values of i, j, m, m_j, h_0 that satisfy the above conditions.

For $m < m_{min}$ or $m > m_{max}$,

$$\mathbb{P}(L_i^n/2 = h | H_i^n = h)\mathbb{P}(H_i^n = h) < \exp(-n^{0.001}).$$

Proof. The summand in Lemma 1.5.6 is given by

$$\mathbb{P}(L_i^n/2 = h | H_i^n = h)\mathbb{P}(H_i^n = h) = C_{h-1} \left|\mathcal{E}_{v_0,h_0}^{2i-m,h}\right| \left|\mathcal{E}_{2i-m+2h-1,h-1}^{v_j,h_j}\right| \left|\mathcal{E}_{v_0,h_0}^{v_j,h_j}\right|^{-1}.$$

By Lemma 1.5.3 we can make the following substitutions:

$$
\begin{aligned}
C_{h-1} &= \frac{4^{h-1}}{\pi^{1/2}h^{3/2}}(1+\Delta_1(h))), \\
\mathcal{E}^{2i-m,h}_{v_0,h_0} &= \binom{2(i-1)-(m-1)}{i-1} = \frac{4^i}{2^{m+1}\pi^{1/2}i^{1/2}}e^{-m^2/4i}(1+\Delta_2(i,m)), \\
\mathcal{E}^{v_j,h_j}_{2i-m+2h-1,h-1} &= \binom{2(j-i-h)-(m_j-m)}{j-i-h} \\
&= \frac{4^{j-i-h}}{2^{m_j-m}\pi^{1/2}(j-i-h)^{1/2}}e^{-(m_j-m)^2/4(j-i-h)}(1+\Delta_2(j-i-h,mj-m)), \\
&= \frac{4^{j-i-h}}{2^{m_j-m}\pi^{1/2}(j-i)^{1/2}}e^{-(m_j-m)^2/4(j-i)}(1+\Delta_2(j-i-h,mj-m)), \\
\mathcal{E}^{v_j,h_j}_{v_0,h_0} &= \binom{2(j-1)-(m_j-1)}{j-1} = \frac{4^j}{2^{m_j+1}\pi^{1/2}j^{1/2}}e^{-m_j^2/4j}(1+\Delta_2(j,m_j))
\end{aligned}
$$

where both Δ_1 and Δ_2 are bounded uniformly by $n^{-0.01}$ over all parameters satisfying the conditions of the lemmas. Combining these equations together proves Lemma 1.5.7.

□

Let's consider the special case where $j \approx n^{0.9}$.

Lemma 1.5.8. *For X^n chosen uniformly from $\mathcal{E}^{v_j,h_j}_{v_0,h_0}$ with $i, j, h_0,$ and h_j satisfying*

- $j = n^{0.9}(1+\Delta')$,
- $n^{0.499} < h_0 < n^{0.501}$.
- $i \in (2n^{0.6}, n^{0.9} - 2n^{0.6})$,
- $h_j = h_0 + m_j$ *where* $|m_j| \le n^{0.451}$,

$$
\mathbb{P}(L^n_i/2 = H^n_i - 1) = \frac{1}{2\pi^{1/2}(h_0)^{3/2}}(1+\Delta),
$$

where $\Delta' \le n^{-0.1}$ and $\Delta = \Delta(i, h_0, h_j) = o(n^{-0.001})$ is uniformly bounded for all i, h_0, h_j in the ranges above.

Proof. Let $m_{min} = \max(-i^{0.6}, m_j - (j-i)^{0.6})$ and $m_{max} = \min(i^{0.6}, m_j + (j-i)^{0.6})$ and consider the inequality which follows from Lemma 1.5.6.

$$\sum_{m=m_{min}}^{m_{max}} \mathbb{P}(L_i^n/2 = h_0 + m | H_i^n = h_0 + m)\mathbb{P}(H_i^n = h_0 + m) \tag{1.9}$$
$$\leq \mathbb{P}(L_i^n/2 = H_i^n)$$
$$\leq ne^{-n^{0.001}} + \sum_{m=m_{min}}^{m_{max}} \mathbb{P}(L_i^n/2 = h_0 + m | H_i^n = h_0 + m)\mathbb{P}(H_i^n = h_0 + m).$$

Lemma 1.5.7 gives

$$\begin{aligned}
&\mathbb{P}(L_i^n/2 = H_i^n) \\
&= (1 + o(n^{-0.01})) \sum_{m=m_{min}}^{m_{max}} \frac{\sqrt{j}}{4\pi\sqrt{i(j-i)(h_0)^3}} \exp\left(-\frac{(jm - im_j)^2}{4ij(j-i)}\right)(1+\Delta). \\
&= (1 + o(n^{-0.001})) \int_{m_{min}}^{m_{max}} \frac{1}{4\pi\sqrt{i(1-i/j)(h_0)^3}} \exp\left(-\frac{(m - im_j/j)^2}{4i(1-i/j)}\right)(1+\Delta)dm.
\end{aligned}$$

By our definition $(m_{min} - \frac{i}{j}m_j) < -n^{0.01}$ and $(m_{max} - \frac{i}{j}m_j) > n^{0.01}$. Therefore the integral above is computed in the standard way, with

$$\int_{-t}^{t} \exp\left(-\frac{(m-m_0)^2}{4c}\right) dm = 2c^{1/2}\pi^{1/2} + \delta(t).$$

where $\delta(t)$ is an error function with exponential decay.

$$\mathbb{P}(L_i^n/2 = H_i^n) = \frac{1}{2\pi^{1/2}(h_0)^{3/2}}(1 + o(n^{-0.001})).$$

□

Corollary 1.5.9. *For any $k \in \mathbb{R}$ and $\alpha \in (0, .48)$ let Γ^n chosen uniformly from $\mathcal{E}_{v_0,h_0}^{v_j,h_j}$ with $i, j, h_0,$ and h_j satisfying*

- $j = n^{0.9}(1 + \Delta')$,
- $n^{0.499} < h_0 < n^{0.501}$.

- $i \in (2n^{0.6}, n^{0.9} - 2n^{0.6})$,
- $h_j = h_0 + m_j$ *where* $|m_j| \leq n^{0.451}$,

$$\mathbb{P}(L_i^n/2 = H_i^n - k(i(n-i)/n)^\alpha) = \frac{1}{2\pi^{1/2}(h_0)^{3/2}}(1+\Delta),$$

where $\Delta' < n^{-0.1}$ *and* $\Delta = \Delta(i, h_0, h_j) = o(n^{-0.001})$ *is uniformly bounded for all* i, h_0, h_j *in the ranges above.*

Proof. The proof goes exactly as in Lemma 1.5.8 with $L_i^n = H_i^n$ replaced by $L_i^n = H_i^n - k(i(n-i)/n)^\alpha$. The order of $k(i(n-i)/n)^\alpha$ is less than $n^{0.49}$ so it will not affect the approximation. □

Lemmas 1.5.11, 1.5.10, and 1.5.12 are used to show that a typical Dyck path is in a proper Ω.

Lemma 1.5.10. *Fix* $0 < a < b < 1$ *and let* $a_k = \lfloor an + nk/K \rfloor$ *where* $K = \lfloor (b-a)n^{0.1} \rfloor$. *For* $\Gamma^n \in Dyck^{2n}$ *chosen uniformly at random,*

$$\mathbb{P}\left(\bigcap_{k=0}^{K} \{n^{0.499} < \Gamma^n(V_{a_k}^n) < n^{0.501}\}\right) > 1 - o(1).$$

Proof. The key fact here is that the Brownian excursion, e_t, is positive for $t \in [a, b]$ when $0 < a < b < 1$ [10]. For some continuous function g such that $\lim_{x \to 0} g(x) = 0$,

$$\mathbb{P}\left(\inf_{t \in [a,b]} \{e_t\} > \epsilon\right) > 1 - g(\epsilon).$$

Convergence of scaled $\Gamma^n/(2n)^{1/2} \in \text{Dyck}^{2n}/(2n)^{1/2}$ to Brownian excursion says we can choose n large enough such that $\mathbb{P}(\inf_{t \in [a,b]}\{\Gamma^n(2nt)/(2n)\} > \epsilon) = 1 - 2g(\epsilon)$. Then

$$\mathbb{P}\left(\inf_{t \in [an, bn]} \Gamma^n(2nt) > \epsilon n^{1/2}/2\right) > 1 - g(\epsilon).$$

Letting $\epsilon = 2n^{-0.001}$ proves the lower bound holds for each $\Gamma_{V_{a_k}^n}^n$.

By standard ballot count analysis for $i \in [an, bn]$ and the approximation used in Lemma 1.5.1,

$$\mathbb{P}\left(\Gamma^n(V_i^n) = rn^{0.5}\right) = C_n^{-1} \frac{rn^{0.5}}{i} \binom{2i - rn^{0.5}}{i} \frac{rn^{0.5}}{n - i + rn^{0.5}} \binom{2n - 2i + rn^{0.5}}{n - i}$$
$$\leq \frac{r^2}{n(a(1-b))} e^{-r^2/4}.$$

Letting $r = n^{0.001}$ gives

$$\mathbb{P}\left(\Gamma^n(V_i^n) > n^{0.501}\right) \leq \int_{n^{0.001}}^{\infty} Cr^2 e^{-r^2/4} dr < e^{-n^{0.001}}$$

as long as n is large enough. Then by the union bound

$$\mathbb{P}\left(\bigcup \{\Gamma^n(V_i^n) > n^{0.501}\}\right) < ne^{-n^{0.001}}.$$

Therefore

$$\mathbb{P}\left(\bigcap_{k=0}^{K-1} \{n^{0.499} < \Gamma^n(V_{a_k}^n) < n^{0.501}\}\right) > 1 - g(n^{-0.001}) - ne^{-n^{0.001}} = 1 - o(1).$$

□

Lemma 1.5.11. *Fix $0 < a < b < 1$. For any n large enough and $\Gamma^n \in Dyck^{2n}$,*

$$\mathbb{P}\left(\bigcup_{k=0}^{K-1} \left\{|\Gamma^n(V_{a_k}^n) - \Gamma^n(V_{a_{k+1}}^n)| > n^{0.451}\right\} \middle| \bigcap_k \left\{n^{0.499} < \Gamma^n(V_{a_k}^n) < n^{0.501}\right\}\right) < e^{-n^{0.0001}}.$$

Proof.

$$\mathbb{P}\left(\Gamma^n(V_{a_k}^n) - \Gamma^n(V_{a_{k+1}}^n) = yn^{0.45} \middle| n^{0.499} < A_k^n < n^{0.501}\right)$$
$$\leq Z^{-1} \left| \mathcal{E}_{V_{a_k}^n, A_k^n}^{V_{a_k}^n + 2n^{0.9} - yn^{0.45}, A_k^n + yn^{0.45}} \right|$$

for some normalizing constant $Z \approx 4^{n^{0.9}}$

Since $A_k^n > n^{0.499}$ we may use Lemma 1.5.3 to show

$$Z^{-1} \left| \mathcal{E}_{V_{a_k}^n, A_k^n}^{V_{a_k}^n + 2n^{0.9} - yn^{0.45}, A_k^n + yn^{0.45}} \right| < 2e^{-|y|}.$$

Integrating this expression over $|y| > n^{0.001}$ shows

$$\mathbb{P}(|\Gamma^n(V_{a_k}^n) - \Gamma^n(V_{a_{k+1}}^n)| > n^{0.451}) < 4e^{-n^{0.001}}.$$

This bound is uniform for all $\Gamma^n(V^n_{a_k}) > n^{0.499}$. Therefore for large enough n

$$\mathbb{P}\left(\bigcup_{k=0}^{K-1}\left\{|\Gamma^n(V^n_{a_k}) - \Gamma^n(V^n_{a_{k+1}})| > n^{0.451} | n^{0.499} < \Gamma^n(V^n_{a_k}) < n^{0.501}\right\}\right) < 4ne^{-n^{0.001}}$$
$$< e^{-n^{0.0001}}.$$

□

Lemma 1.5.12. *Fix* $0 < j \leq 2n^{0.9}$, *and* h_0, h_j *both bounded between* $n^{0.499}$ *and* $n^{0.501}$. *For* $X^n \in \mathcal{E}^{v_j,h_j}_{v_0,h_0}$ *chosen uniformly at random*

$$\mathbb{P}\left(n^{0.49} < H^n_i < n^{0.51} \ \forall i \in (0,j)\right) \geq 1 - n^2 e^{-n^{0.08}}.$$

Proof. For $i > j/2$, if $h < n^{0.49}$ or $h > n^{0.51}$ then $|h_0 - h| > n^{0.498}$. We may appeal to Lemma 1.5.2 which gives

$$|\mathcal{E}^{V^n_i,h}_{v_0,h_0}| < \frac{4^i}{2^{h-h_0+1}} e^{-n^{0.08}}$$

and

$$|\mathcal{E}^{v_j,h_j}_{V^n_i,h}| < \frac{4^{j-i}}{2^{h_j-h+1}}.$$

Then we can conclude

$$\mathbb{P}(H^n_i < n^{0.49} \text{ or } H^n_i > n^{0.51}) < e^{-n^{0.08}}.$$

A similar bound can be found for $i < j/2$. Summing over $0 < h < n^{0.49}$ and $n^{0.51} < h < i$ gives

$$\mathbb{P}(H^n_i \notin (n^{0.49}, n^{0.51})) < ne^{-n^{0.08}}.$$

Now apply the union bound to complete the proof.

□

The following two lemmas are deterministic and allow us to bound the contribution to the total number of fixed points by small intervals.

Lemma 1.5.13. *Let $I \subset [an, bn]$ denote an interval of length at most n^α. For $\gamma \in Dyck^{2n}$, if $h_i > n^{0.49}$ for all $i \in I$, then*

$$\theta_I \leq n^{\alpha - 0.49}.$$

Proof. If the jth excursion is contained in ith excursion, then both $h_j > h_i$ and $l_j/2 < l_i/2$. If $\theta_i = 1$ then for $i < j < l_i/2$, $\theta_j = 0$.

For an interval I with $|I| < n^{0.49}$ suppose at least one fixed point exists. Let i^* denote the first excursion that satisfies $\theta_{i*} = 1$. Since i^* corresponds to a fixed point,

$$l_{i^*} = h_{i^*} > n^{0.49}.$$

Therefore, for all $j \in I$ such that $j > i^*$ the jth excursion is contained in the i^*th excursion and $\theta_j = 0$. Therefore either $\theta_I = 0$ if no such i^* exists, or $\theta_I = \theta_{i^*} = 1$. For an interval of size less than n^α, it can be covered by $n^{\alpha-0.49}$ intervals of size $n^{0.49}$ each of which has at most one fixed point so the total number of fixed points will be bounded by $n^{\alpha-0.49}$.

□

Corollary 1.5.14. *Let $I \subset [an, bn]$ denote an interval of length at most n^α. For $\gamma \in Dyck^{2n}$, if $h_i > n^{0.49}$ for all $i \in I$, then*

$$\theta_I^{K,\alpha} \leq 2n^{\alpha-0.49}.$$

Proof. The function $f(i) = (i(n-i)/n)^\alpha$ can change by at most 1 over any interval of length $n^{.49}$. This implies that over any interval I' of length $n^{.49}$ has $\theta_{I'}^{K,\alpha} \leq 2$. Then the result follows as in Lemma 1.5.13. □

Lemma 1.5.15. *There exists constant $C > 0$ such that for sufficiently large n and i, j, w that satisfy the following:*

- $2n^{0.6} < i < j < n^{0.9} - 2n^{0.6}$,
- $|i - j| > 2n^{0.6}$,

- *and* $w < n^{0.451}$,

$$\Psi(i,j,w,n) = \sum_{m'=m'_{min}}^{m'_{max}} \sum_{m=m_{min}}^{m_{max}} \exp\left(-\frac{1}{4}\left(\frac{m^2}{i} + \frac{(m-m')^2}{j-i} + \frac{(w-m')^2}{n^{0.9}-j} - \frac{w^2}{n^{0.9}}\right)\right)$$
$$< C\sqrt{\frac{i(j-i)(n^{0.9}-j)}{n^{0.9}}}.$$

Proof. Our goal will be to convert this double sum into a recognizable form. Since the summand is positive for all values of m and m' we may replace the bounds of the sum with $\pm\infty$.

$$\begin{aligned}\Psi(i,j,w,n) &\leq \sum_{m'=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \exp\left(-\frac{1}{4}\left(\frac{m^2}{i} + \frac{(m-m')^2}{j-i} + \frac{(w-m')^2}{n^{0.9}-j} - \frac{w^2}{n^{0.9}}\right)\right)\\ &\leq \sum_{m'=-\infty}^{\infty} \left[\exp\left(-\frac{1}{4}\left(\frac{(w-m')^2}{n^{0.9}-j} - \frac{w^2}{n^{0.9}}\right)\right) G(i,j,m')\right]\end{aligned}$$

where

$$G(i,j,m') = \sum_{m=-\infty}^{\infty} \exp\left(-\frac{1}{4}\left(\frac{m^2}{i} + \frac{(m-m')^2}{j-i}\right)\right).$$

With some algebra we see

$$\begin{aligned}G(i,j,m') =& \exp\left(-\frac{1}{4}\frac{m'^2}{j}\right) \sum_{m=-\infty}^{\infty} \exp\left(-\frac{1}{4}\frac{j(m-m'i/j)^2}{i(j-i)}\right)\\ \leq& C_1 \exp\left(-\frac{1}{4}\frac{m'^2}{j}\right)\sqrt{\frac{i(j-i)}{j}}\end{aligned}$$

for some positive constant C_1 that does not depend on i, j or m'. Inserting this into the upper bound for $\Psi(i,j,w,n)$ gives

$$\begin{aligned}\Psi(i,j,w,n) \leq& C_1\sqrt{\frac{i(j-i)}{j}} \sum_{m'=-\infty}^{\infty} \exp\left(-\frac{1}{4}\left(\frac{m'^2}{j} + \frac{(m'-w)^2}{n^{0.9}-j} - \frac{w^2}{n^{0.9}}\right)\right)\\ \leq& C_1\sqrt{\frac{i(j-i)}{j}} \sum_{m'=-\infty}^{\infty} \exp\left(-\frac{1}{4}\left(\frac{n^{0.9}(m'-wj/n^{0.9})^2}{j(n^{0.9}-j)}\right)\right)\\ \leq& C\sqrt{\frac{i(j-i)(n^{0.9}-j)}{n^{0.9}}}\end{aligned}$$

where $C > 0$ and does not depend on i, j, w, and n.

□

Chapter 2

BOOTSTRAP PERCOLATION

Bootstrap percolation first appeared in a paper by Chalupa et al [20] as a model for ferromagnetism. Adler [23] provides a wonderful introduction to the subject.

In general we take a graph $G = (V, E)$ with vertex set V and edge set E and a parameter θ which we call the *threshold*. Each vertex in the graph is initially in one of two possible states, either open or closed. At each subsequent step a vertex can become open if at least θ of its neighbors are open. Once open, a vertex remains open for all eternity.

We describe the increasing evolution of the configuration formally. Let $\omega_t \in \{0,1\}^V$ denote the configuration of the vertices at time $t \geq 0$. If a vertex v is open at step t we say $\omega_t(v) = 1$ and similarly if the vertex is closed at time t, $\omega_t(v) = 0$. For bootstrap percolation with threshold θ, ω_t evolves as follows for $t \geq 0$:

$$\omega_{t+1}(v) = \begin{cases} 1, & \omega_t(v) = 1 \\ 1, & \sum_{v' \sim v} \omega_t(v') \geq \theta \\ 0, & \text{otherwise} \end{cases} \tag{2.1}$$

where $v' \sim v$ if there is an edge in E connecting v and v'.

Many natural questions can be asked. Given some initial configuration, we can ask what the evolved configuration look like after some time. In particular we care about the steady state, ω_∞. Typically this is viewed probabilistically. Given a distribution on ω_0 what can we say about ω_∞.

The first major progress came from van Enter [22] and later Schonmann [21]. They showed a $0-1$ law for configurations on $V = Z^d$ with edges connecting each vertex to its $2d$ nearest neighbors. In there model, initially each vertex is independently open with probability p. For $\theta \leq d$, if $p > 0$ then the entire grid becomes open with probability 1. If $\theta > d$ then everything becomes completey open only if $p = 1$.

The next big step in the history of bootstrap percolation was to view the process on an increasing family of graphs $\mathcal{G} = \{G_n = (V_n, E_n)\}$ where the initial probability that a vertex is open is given by a function of n, $p = p(n)$. As each graph is finite, $f^n(p) := \mathbb{P}_p(\omega_\infty \equiv \mathbf{1})$ can be viewed as an increasing polynomial in p with $f^n(0) = 0$ and $f^n(1) = 1$. By continuity there is a critical value, p_c, such that $f^n(p_c) = 1/2$. Much work centers around finding bounds on p_c as a function of n.

Aizenman and Lebowitz [24] showed for the finite d dimensional grid, $[n]^d$, and threshold $\theta = 2$, there exists constants c_1, c_2 such that $c_1 < (\log n)^{d-1} p_c < c_2$. Moreoever, they show that the transition for $f^n(p)$ from 0 to 1 is sharp near p_c.

In a widely celebrated paper Holroyd [25] showed that

$$p_c \sim \pi^2/18 \log n$$

when $d = \theta = 2$. Later this result was expanded on by Holroyd, Ligget, and Romik [26] to $d = 2, \theta = k+1$ where the neighborhood of a vertex is the k closest vertices in each of the cardinal directions. They show $p_c \sim \pi^2/(3(k+2)(k+1)\log n)$ for this graph. These types of results have been extended to higher dimesnions by [27], random graphs [28], and more geometric settings [33]. It is a very active area of research.

Our graph of interest is the d-dimensional Hamming torus. The Hamming torus has the same vertex set as the lattice, $V = [n]^d$, but the edge set is modified to connect every vertex that can be connected with a straight path on the grid. In terms of the coordinates edge set is

$$E = \{(v, w) : v \text{ differs from } w \text{ in exactly one coordinate } \}.$$

Gravner et al. introduced the study of bootstrap percolation on the Hamming torus [31]. In their paper they focus on evolution thresholds greater than 2. They find threshold functions of the critical probability, $p_c(\theta, d)$, for the event $\mathcal{C} = \{\omega_\infty \equiv \mathbf{1}\}$, where $\mathbb{P}_{p_c(\theta,d)}(\mathcal{C}) = 1/2$. They also consider finer structure:

$$\mathcal{C}_i = \{\exists\ W \subset [n]^d \text{ with } \dim(W) = i \ s.t.\ \omega_\infty|_W = \mathbf{1}\},$$

and they find bounds for the critical exponent of threshold functions $p_c(\theta, i, d)$, where

$\mathbb{P}_{p_c(\theta,i,d)}(\mathcal{C}_i) = 1/2$. For $i = 0$ we slightly alter the definition to be

$$\mathcal{C} = \{\exists\ v\ s.t.\ \sum_{w \sim v} \omega(w) \geq \theta\}.$$

In many cases they were able to show the critical probability is of the form $p_c(\theta, i, d) = (1 + o(1))an^{\beta}$ for some $\beta < 0$. We call β the critical exponent. For $d = \theta = 3$ they showed the threshold function for $\mathcal{C}_0$, and $\mathcal{C}_1$ is $(1 + o(1))an^{-2}$ and the threshold function for $\mathcal{C}_2, \mathcal{C}_3$ is $(1+o(1))an^{-5/3}$, for some constant $a > 0$. They showed for $d \geq 3$, $\mathcal{C}_1$ and $\mathcal{C}_2$ have different critical exponents. For $d = 2$, and all values of θ, the critical exponent is the same for $\mathcal{C}_1$ and $\mathcal{C}_2$.

We consider the case $\theta = 2$ and $d > 2$. Let $j < \sqrt{d}$. We show that the equivalently defined threshold functions for $\mathcal{C}_2, \mathcal{C}_4, \cdots, \mathcal{C}_{2j}$ have distinct exponents. We will also show for $i < j$, the threshold functions for $\mathcal{C}_{2i-1}$ and $\mathcal{C}_{2i}$ have the same exponent. For $i > \sqrt{d}$, we show that $\mathcal{C}_i$ all have the same critical exponent. After we have determined the critical exponent for these events, we will give a precise description of the asymptotics of $p_c(\theta, i, d)$. Unlike the threshold functions for the grid $\mathbb{Z}^d$ found in Holroyd [25], $p_c(\theta, i, d)$ is not sharp. For this remainder of this chapter we drop the parameter θ as it will always be 2.

2.1 Statements

First we need a few definitions.

Definition 2.1.1. *A subset $V \subset [n]^d$ is a* **sublattice** *if there exists a set of indices $I(V)$ and constants $\{\alpha_l\}_{l \in I(V)}$ such that $v_l = \alpha_l$ for all $l \in I(V)$ iff $v \in V$. We say V has dimension i if $|I(V)| = d - i$.*

Definition 2.1.2. *For a set of open nodes, $S = \{v : \omega_t(v) = 1\}$ at some time t, we denote the open nodes of the evolved configuration by*

$$\langle S \rangle = \{v : \omega_\infty(v) = 1 \text{ where } \omega_0(v) = 1 \iff v \in S\}.$$

We say a sublattice V is **internally spanned** *if there exists a subset $S \subset V$ of open nodes at time 0, such that $V = \langle S \rangle$.*

Definition 2.1.3. *A sublattice V is* **maximal** *in $\langle S\rangle$ if no other sublattice in $\langle S\rangle$ contains V.*

Our results center around the following events:

- $\mathcal{I}_V = \{\exists\ S \subset V$ such that $\omega_0(v) = 1\ \forall\ v \in S$ and $\langle S\rangle = V\}$. In other words V is internally spanned.
- $\mathcal{I}_l\ = \{\exists\ V$ such that $\dim(V) = l$ and $\mathcal{I}_V$ occurs $\} = \bigcup_{\dim(V)=l} \mathcal{I}_V$.
- $\mathcal{C}_l\ = \{\exists\ V$ such that $\dim(V) = l$ and $\omega_\infty(v) = 1$ for all $v \in V\}$.

Note the slight difference in the definitions of $\mathcal{I}_i$ and $\mathcal{C}_i$. For $\mathcal{C}_i$ the only thing that matters is the final state ω_∞ where for $\mathcal{I}_i$ it is important how one gets to ω_∞.

Lastly we prove a statement about the random variable $D = \sup_{0\le l\le d}\{(\mathbf{1}_{\mathcal{I}_l})l\}$. This gives the dimension of the largest sublattice that is internally spanned.

Much of the work is in finding bounds for the threshold function for $\mathcal{I}_l$ denoted by $p_{\mathcal{I}}(l,d)$. Then we show that $p_c(l,d) = p_c(2,l,d)$ will have the same asymptotic behavior as $p_{\mathcal{I}}(l,d)$ when l is even. We also will show for odd dimension subspaces, $l = 2j-1$, the threshold function $p_c(2j-1,d) = p_c(2j,d)$.

Now we are in a position to state our main results. Fix d, j such that $j(j+1) < d$.

Theorem 2.1.1. *If $p = f(n)n^{-d/(j+1)-j}$, then*

$$\mathbb{P}_p(\mathcal{I}_{2j}), \mathbb{P}_p(\mathcal{C}_{2j}) \to \begin{cases} 0 & \text{if } f(n) \to 0 \\ 1 & \text{if } f(n) \to \infty \\ 1 - e^{-\lambda(j,d,a)} & \text{if } f(n) \to a \end{cases}$$

where $\lambda(j,d,a) = \binom{d}{2j}(2j)!2^{-j}a^{j+1}$.

This implies

$$p_c = (1 + o(1))a_{\frac{1}{2}}n^{-d/(j+1)-j}$$

where $1 - e^{-\lambda(j,d,a_{\frac{1}{2}})} = 1/2$.

The next result comes from the application of the Stein-Chen method [30]. For two non-negative integer valued random variables Y and Z the total variation is defined as

$$d_{TV} = \frac{1}{2}\sum_{k=0}^{\infty} |\mathbb{P}(Y = k) - \mathbb{P}(Z = k)|.$$

For the remaining results we will assume $p = an^{-d/(j+1)-j}$ and $\lambda(j, d, a) = \binom{d}{2j}(2j)!2^{-j}a^{j+1}$. To simplify the statement of results we let $\lambda = \lambda(j, d, a)$ where no confusion will arise.

Theorem 2.1.2. *For $j(j + 1) < d$, let Y denote the number of sublattices $V \subset [n]^d$ such that both $\dim(V) = 2j$ and $\mathcal{I}_V$ occurs, and let Z denote a Poisson(λ) random variable. Then*

$$\lim_{n\to\infty} d_{TV}(Y, Z) \to 0.$$

The precision given by Theorem 2.1.2 leads to the following result:

Theorem 2.1.3. *Let $j(j+1) = d > 6$, $c = \binom{d-2j+2}{2}$, and $\lambda' = \binom{d}{2j-2}(2j-2)!2^{-j+1}a^j$. Then*

$$\mathbb{P}_p(\mathcal{I}_{2j}) \to \sum_{k=1}^{\infty} \frac{e^{-\lambda'}}{k!}\lambda'^k(1 - e^{-ack}).$$

For $l \geq j$,

$$\mathbb{P}_p(\mathcal{I}_{2l}) \to \mathbb{P}_p(\mathcal{I}_{2j}).$$

Lastly we have a useful corollary for the random variable D. We assume p, λ, λ', and c are as above.

Corollary 2.1.4. *Largest Sublattices ($D = 2l$)*

If $(j + 1)(j + 2) < d$, then

$$\mathbb{P}_p(D = 2j - 2) \to e^{-\lambda}$$

$$\mathbb{P}_p(D = 2j) \to 1 - e^{-\lambda}.$$

If $j(j+1) < d < (j+1)(j+2)$, *then*

$$\mathbb{P}_p(D = 2j-2) \to e^{-\lambda}$$
$$\mathbb{P}_p(D = d) \to 1 - e^{-\lambda}.$$

If $j(j+1) = d > 6$, *then*

$$\mathbb{P}_p(D = 2j-4) \to e^{-\lambda'},$$
$$\mathbb{P}_p(D = 2j-2) \to \sum_{k=1}^{\infty} \frac{e^{-\lambda'}}{k!} \lambda'^k (e^{-ack}),$$
$$\mathbb{P}_p(D = d) \to 1 - \sum_{k=0}^{\infty} \frac{e^{-\lambda'}}{k!} \lambda'^k e^{-ack}.$$

In Section 3, we prove lemmas that describe the evolution ω_t when $\theta = 2$. In Section 4, we prove both upper and lower bounds for the critical exponent for the events $\mathcal{C}_{2j}$ and $\mathcal{I}_{2j}$. In Section 5 we use the Stein-Chen method [30] to describe precisely the asymptotics of $p_c(l, d)$. In Section 6 we combine everything to prove our theorems.

2.2 Growth for $\theta = 2$

We begin with the simplest case. Suppose $u \neq v$ are the only nodes which are initially open. Define the distance between to nodes as $\text{dis}(u, v) = \sum_{i=1}^{d} \mathbf{1}_{u_i \neq v_i}$, the number of coordinates where u and v differ. If $\text{dis}(u, v) > 2$ then no new nodes become open $\langle u, v \rangle = \{u, v\}$. If $\text{dis}(u, v) \leq 2$ then u and v must agree for all but at most 2 indices. Without loss of generality, we may choose a basis so that $u_i = v_i$ for $i > 2$. Suppose first that $u_2 = v_2$ as well (i.e. $\text{dis}(u, v) = 1$), the line $\{(t, u_2, \cdots), t \in [n]\}$ has two nodes initially on, and after one step every node in that line becomes open. Every node not on the line has at most one neighbor on the line, so growth stops.

If $\text{dis}(u, v) = 2$, then after one step the co-neighbors of u and v, $u' = (u_1, v_2, \cdots)$ and $v' = (v_1, u_2, \cdots)$, become open. The nodes u and u' are two different open neighbors for every closed node in the line $\{(u_1, s, \cdots) : s \in [n]\}$, so after two steps the entire line becomes open. The same is true for the lines containing both u and v', both v and u', and both v' and v. Once those lines are open every other node in the plane $\{(t, s, \cdots) : (t, s) \in [n]^2\}$ has a at least two (in fact four) open neighbors, so the entire plane becomes open.

Growth for higher dimension sublattices is a bit more involved. First we generalize the distance function to subsets S_1, S_2 as follows,

$$\mathrm{dis}(S_1, S_2) = \inf_{u \in S_1, v \in S_2} \mathrm{dis}(u, v).$$

For a $2j$-dimension sublattice, V_{2j}, to become open, most of the time there are $j+1$ distinct points $\{v_1, \cdots, v_{j+1}\} \subset V_{2j}$ such that $\langle v_1, \cdots, v_{j+1} \rangle = V_{2j}$ and that occurs as follows. Let $V_{2i} = \langle v_1, \cdots, v_{i+1} \rangle$. For $1 \le i \le j$, $\mathrm{dis}(v_{i+1}, V_{2i-2}) = 2$.

$$\begin{aligned} V_0 =& \{v_1\} \\ V_2 =& \langle v_1, v_2 \rangle \\ V_{2i} =& \langle v_{i+1}, V_{2i-2} \rangle = \langle v_1, \cdots, v_{i+1} \rangle \\ V_{2j} =& \langle v_{j+1}, V_{2j-2} \rangle = \langle v_1, \cdots, v_{j+1} \rangle. \end{aligned}$$

There exist exactly two points that are distance 1 away from both v_{j+1} and some point in V_{2j-2}. After one step, those nodes become open. Then the lines connecting those points with v_{j+1} and V_{2j-2} become open, followed by the planes containing those new lines. This continues until eventually the whole $2j$ dimension sublattice becomes open.

We will state and prove a few necessary lemmas. The key point in the next few lemmas is growth continues only if there are two sets of open nodes within distance 2 of each other.

Lemma 2.2.1. *For $S \subset [n]^d$, let $\overline{S}$ denote the smallest sublattice that contains S. If V is a sublattice and u is a node with dis$(V, u) \le 2$ then*

$$\langle V, u \rangle = \overline{\{V \cup u\}}.$$

Lemma 2.2.2. *If V, W are open sublattices and dis$(V, W) \leq 2$ then $\langle V, W\rangle = \overline{\{V \cup W\}}$.*

The next two lemmas give conditions for when and how a sublattice is internally spanned.

Lemma 2.2.3. *For an initial configuration of open nodes S, let V be a maximal sublattice in $\langle S\rangle$. Then V is internally spanned with $V = \langle S \cap V\rangle$.*

Lemma 2.2.4. *Let S be a set of open nodes in $[n]^d$ with $V \subset \langle S\rangle$ a maximal open sublattice. There exist disjoint subsets $S_1, S_2 \subset S$ and sublattices $V_1, V_2 \subset V$ such that $\langle S_1\rangle = V_1$, $\langle S_2\rangle = V_2$, and $\langle S_1 \cup S_2\rangle = V$.*

Proof of Lemma 2.2.1. (By induction) We have shown that the lemma holds if V has dimension 0 (is a single node). Suppose the lemma holds for all sublattices W with $\dim(W) < i$. Let V be a sublattice with $\dim(V) = i$ and let u be a node with $\text{dis}(V, u) \leq 2$. Without loss of generality we assume the last $d - i$ coordinates are fixed, i.e. $\{I(V)\} = [i + 1, d]$. Again without loss of generality we may also assume that

$$u \in \{(u_1, \cdots, u_d) : u_l = \alpha_l(V) \text{ for } l > i + 2\}.$$

Let V_k denote the sublattice of V that fixes the k^{th} coordinate to the value u_k. Then V_k has dimension $i - 1$ and $\text{dis}(V_k, u) \leq 2$. By the induction hypothesis, $\langle V_k, u\rangle = \overline{\{V_k, u\}}$. For $a = (a_1, \cdots, a_d) \in \overline{\{V, u\}}$, there are two neighbors

$$a_{u_1} = (u_1, a_2, \cdots, a_{i+2}, \alpha_{i+3}, \cdots, \alpha_d) \in \overline{\{V_1, u\}}$$

and

$$a_{u_2} = (a_1, u_2, \cdots, a_{i+2}, \alpha_{i+3}, \cdots, \alpha_d) \in \overline{\{V_2, u\}},$$

so a becomes open and we can conclude $\overline{\{V, u\}} \subseteq \langle V_1, V_2, u\rangle \subseteq \langle V, u\rangle$. Trivially $\langle V, u\rangle \subseteq \overline{\{V, u\}}$ so we have equality for the two sets. Moreover, if $u \notin V$ then $i + 1 \leq \dim(\overline{\{V, u\}}) \leq i + 2$.

□

Proof of Lemma 2.2.2. This is a natural extension of Lemma 2.2.1. Trivially we have $\langle V, W\rangle \subseteq \overline{\{V, W\}}$. Let $V^0 = V$. We define V^l based on V^{l-1}. Let W^{l-1} denote the subset of W that satisfies $0 < \text{dis}(V^{l-1}, u) \leq 2$ for every $u \in W^{l-1}$. For $l > 0$ if $W \cap (V^{l-1})^c$ is non-empty there exists a $w_l \in W^{l-1}$. We then define $V^l = \langle \{V^{l-1}, w_l\} \rangle$ for some choice of w_l. By the previous this is the sublattice $\overline{\{V^{l-1}, w_l\}}$. Its dimension is strictly greater than $\dim(V^{l-1})$. If $W \cap (V^{l-1})^c$ is empty then $V^l = V^{l-1}$. Since $\{V^l\}$ is an increasing sequence of sublattices bounded by $\overline{\{V, W\}}$ it must stabilize to some sublattice V^m in a finite number of steps. By definition $V \subseteq V^m$, and more importantly, $W \cap (V^m)^c = \emptyset$ so $W \subseteq V^m$. We also have that $V^m \subseteq \langle V, \cup_l w_l \rangle \subseteq \langle V, W \rangle$. Combining everything we get

$$\overline{\{V, W\}} \subseteq V^m \subseteq \langle V, W \rangle \subseteq \overline{\{V, W\}}$$

and the lemma holds.

$\square$

Proof of Lemma 2.2.3. Let $S_1 = S \cap V$ and $S_2 = \{S \backslash S_1\}$. If $\langle S_1 \rangle = V$ then we are done. Suppose that $\langle S_1 \rangle \neq V$. Since V eventually becomes open, there must be some node $u \in \langle S_2 \rangle$ such that $\text{dis}(\langle S_1 \rangle, \langle u \rangle) \leq 2$, otherwise evolution would stop and V could not be contained in $\langle S \rangle$.. In particular, there is a node $u \in \langle S_2 \rangle$ such that $u \notin V$ yet $\text{dis}(V, u) \leq 2$. By Lemma 2.2.2 the smallest sublattice that contains both u and V becomes open eventually. However V is maximal so no such u can exists and $\langle S_1 \rangle = V$.

$\square$

Proof of Lemma 2.2.4. V is maximal so we may assume $\langle S \rangle = V$. Consider the sequence of nested collections of sublattices contained in $\langle S \rangle$,

$$\{W_i^0\} \subset \{W_i^1\} \subset \cdots \subset \{W_i^k\} \subset V$$

where $S = \{W_i^0\}$ and $\{W_i^{k+1}\}$ is formed by finding two sublattices $W_{i_1}^k$ and $W_{i_2}^k$ within Hamming distance 2 of each other and setting $W_{i_1}^{k+1} = \langle W_{i_1}^k \cup W_{i_2}^k \rangle$ and reindexing the others appropriately. Since S is finite, eventually we will have two sublattices $W_{i_1}^k, W_{i_2}^k \neq V$ such that $\langle W_{i_1}^k \cup W_{i_2}^k \rangle = V$. Each $W_{i_l}^k$ had a unique set S_l such that $\langle S_l \rangle = W_{i_l}^k$ for $l = 1, 2$.

$\square$

2.2.1 Growth Hueristics

To get an idea of the exponent of $p_c(2j,d)$ we make a heuristic argument for the exponent of $p_{\mathcal{I}}(2j,d)$. For a $2j$-dimension sublattice V, we get an estimate on the probability $\mathbb{P}_p(\mathcal{I}_V)$. The probability for $\mathcal{I}_V$ only depends on the dimension of V. From Lemma 2.2.4 we know that V is internally spanned if there exist two internally spanned sublattices, V_1 and V_2, with $\langle V_1, V_2\rangle = V$. We assume that $\dim(V_1) = 0$ and $\dim(V_2) = 2j-2$. There are roughly n^{2j} choices for V_1 and roughly n^2 choices for V_2. We estimate the probability that at least one V_1 and one V_2 are internally spanned by assuming independence and using expectation. We let $M_i(p) = M_i$ denote the probability that a particular i-dimension sublattice is internally spanned. In particular, $\mathbb{P}_p(\mathcal{I}_V) = M_{2j}$. This gives

$$M_{2j} \approx \mathbb{P}_p(\exists\ V_1 \subset V\ s.t.\ \mathcal{I}_{V_1})\mathbb{P}_p(\exists\ V_2 \subset V\ s.t.\ \mathcal{I}_{V_2}) \approx n^{2j}pn^2M_{2j-2}.$$

We approximate the M_{2j-2} in the same manner and get the estimate:

$$M_{2j} \approx \prod_{i=0}^{j} n^{2j-2i}pn^2 = n^{j(j+1)+2j}p^{j+1}.$$

There are roughly n^{d-2j} choices for V so

$$\mathbb{P}_p(\mathcal{I}_{2j}) \approx n^{d-2j}M_{2j} \approx n^{d-2j+j(j+1)+2j}p^{j+1}.$$

Setting this equal to 1 and solving for p gives the appropriate estimate:

$$p_{\mathcal{I}}(2j,d) \approx n^{-d/(j+1)-j}.$$

The next few sections will show that this estimate is reasonably accurate (up to a constant factor). We will also show that $p_{\mathcal{I}}(2j,d)$ has the same asymptotics as $p_c(2j,d)$.

2.3 Critical Probability

To find the asymptotics of $p_c(2j,d)$, we will first prove upper and lower bounds for the exponent of $p_{\mathcal{I}}(2j,d)$. Since $\mathcal{I}_{2j} \subset \mathcal{C}_{2j}$ any upper bound for $p_{\mathcal{I}}(2j,d)$ will hold for $p_c(2j,d)$. With a little more work, we then prove the lower bound for the exponent of $p_{\mathcal{I}}(2j,d)$ will also be a lower bound for the exponent of $p_c(2j, d)$.

For odd dimension sublattices we will show that $\mathbb{P}_p(\mathcal{I}_{2j-1}) \leq \mathbb{P}_p(\mathcal{I}_{2j})$ hence $p_c(2j-1, d) = p_c(2j, d)$. This is apparent in the case of a line and a plane. For a line to be internally spanned, two nodes need to be co-linear, whereas for a plane to be internally spanned, two nodes only need to be co-planar.

First the upper bound.

2.3.1 Upper Bound

Proposition 2.3.1. *For any $f(n) \to \infty$, and for all $d > j(j+1)$, if $p = f(n)n^{-d/(j+1)-j}$, then*

$$\mathbb{P}_p(\mathcal{C}_{2j}), \mathbb{P}_p(\mathcal{I}_{2j}) \to 1.$$

This implies

$$p_c(2j-1, d) \leq p_c(2j, d) \leq p_{\mathcal{I}}(2j, d) < f(n)n^{-d/(j+1)-j}.$$

We will prove this proposition with the caveat that $f(n)$ does not grow too fast. Since $\mathbb{P}_p(\mathcal{I}_{2j})$ and $\mathbb{P}_p(\mathcal{C}_{2j})$ are increasing in p the proposition will still be true for faster growing $f(n)$.

Proof. First we define a sufficient event $E_{2j} \subset \mathcal{I}_{2j} \subset \mathcal{C}2j$. If we can show $\mathbb{P}_p(E_{2j}) \to 1$ for some value of p then we can conclude $\mathbb{P}_p(\mathcal{C}2j), \mathbb{P}_p(\mathcal{I}_{2j}) \to 1$ as well.

For a fixed set of constants $\alpha = \{\alpha_{2j+1}, \cdots, \alpha_d\}$, let V_α denote the sublattice given by

$$V_\alpha = \{v \in [n]^d : v_i = \alpha_i \text{ for } 2j+1 \leq i \leq d\}.$$

There are n^{d-2j} such sublattices. For $\alpha \neq \alpha'$, $V_\alpha \cap V_{\alpha'} = \emptyset$. Each event $\mathcal{I}_{V_\alpha}$ will depend only on the nodes in V_α so the events are independent. The nodes in each V_α are all *i.i.d.*

random $\{0,1\}$-variables so the events will all have the same probability $\mathbb{P}_p(\mathcal{I}_{V_\alpha}) = \mathbb{P}_p(\mathcal{I}_{V_{\alpha'}})$. We now define the sufficient event

$$E_{2j} = \bigcup_{\alpha} \mathcal{I}_{V_\alpha}.$$

We will show that $\mathbb{P}_p(E_{2j}) \to 1$ for sufficiently large p that satisfy the conditions of the proposition. Since $E_{2j} \subset \mathcal{I}_{2j} \subset \mathcal{C}_{2j}$ this implies $\mathbb{P}_p(\mathcal{I}_{2j}), \mathbb{P}_p(\mathcal{C}_{2j}) \to 1$ as well.

Lemma 2.3.2. *Let j, d and p be as defined in Proposition 2.3.1, and $2i \leq 2j$.*

$$M_{2i} \geq (2i)! 2^{-i-1} n^{i(i+3)} p^{i+1}(1 - o(1)). \tag{2.2}$$

Proof of Lemma 2.3.2. Let V be a sublattice with dimension $2i$. Suppose we have a collection of nodes $\alpha = \{v_1, \cdots, v_{i+1}\} \subset V$ such that $\langle \{v_1, \cdots, v_{i+1}\} \rangle = V$. The probability that only these nodes are open is exactly $p^{i+1}(1-p)^{n^{2i}-i-1}$. Let $\mathcal{L}_V$ be the set of all such collections. Then

$$M_{2i} = \mathbb{P}_p(\mathcal{I}_V) \geq \sum_{\mathcal{L}_V} p^{i+1}(1-p)^{n^{2i-i-1}} = |\mathcal{L}_V| p^{i+1}(1 - o(1)).$$

We call a collection, $\alpha = \{v_1, \cdots, v_{i+1}\}$ *perfect* in V if $\text{dis}(v_{l'}, v_l) = 2(l-1)$ for $l' < l$ and $v_1 < v_2$ in lexicographical ordering. For $i' \leq i$, $\alpha_{i'} = \{v_1, \cdots, v_{i'+1}\}$ is perfect in $\langle \alpha_{i'} \rangle = V'$ and $\dim(V') = 2i'$. Note that a non-trivial rearrangement of a perfect collection is not a perfect collection. This makes counting them easier.

Let $\mathcal{L}_V^* \subset \mathcal{L}_V$ denote the subset of perfect collections for V. We will compute a lower bound for $|\mathcal{L}_V^*|$ inductively. Suppose for any sublattice W with $\dim(W) = 2i - 2$ we have

$$|\mathcal{L}_W^*| \geq (2i-2)! 2^{-i} n^{(i-1)(i+2)}.$$

By induction we have,

$$
\begin{aligned}
|\mathcal{L}_V^*| = & \sum_{\dim(W)=2i-2} \sum_{\alpha' \in \mathcal{L}_W^*} \sum_{v \in V} \mathbf{1}_{\alpha' \cup v \text{ is perfect}} \\
\geq & \sum_{\dim(W)=2i-2} \sum_{\alpha' \in \mathcal{L}_W^*} (n-i-1)^{2i} \\
\geq & \sum_{\dim(W)=2i-2} (2i-2)!2^{-i} n^{(i-1)(i+2)} n^{2i}(1-o(1)) \\
\geq & \binom{2i}{2} n^2 (2i-2)!2^{-i} n^{(i-1)(i+2)}(1-o(1)) \\
= & (2i)!2^{-i-1} n^{i(i+3)}(1-o(1))
\end{aligned}
$$

Since this formula holds if V is a plane then it will hold for all V with dimension small enough that the approximation $(1-p)^{n^{2i}-i-1} = 1-o(1)$ is valid. This approximation is valid exactly when $d > i(i+1)$. Therefore

$$
M_{2i} = \mathbb{P}_p(\mathcal{I}_V) \geq |\mathcal{L}_V| p^{i+1} \geq |\mathcal{L}_V^*| p^{i+1} = (2i)!2^{-i-1} n^{i(i+3)} p^{i+1}(1-o(1)).
$$

□

Initially we will assume that $f(n)$ grows slowly enough that

$$
O(n^{j(j+3)} p^{j+1}) = O(n^{2j-d} f(n)^{j+1})
$$

is much less than 1.

The event E_{2j} is equivalent to the event that there exists an α such that $\mathcal{I}_{V_\alpha}$ occurs. There are n^{d-2j} choices for α. The events $\{\mathcal{I}_{V_\alpha}\}$ are independent, so the probability that no $\mathcal{I}_{V_\alpha}$ occurs is bounded by

$$
\left(1 - O(n^{2j-d} f(n)^{j+1})\right)^{n^{d-2j}} \leq e^{-O(f(n)^{j+1})} = o(1).
$$

Therefore the probability that $\mathcal{I}_{V_\alpha}$ occurs for some α, and hence E_{2j} occurs, is bounded below by $\mathbb{P}_p(E_{2j}) \geq 1 - o(1) \to 1$. Since E_{2j} is increasing in p we can remove the restriction on the growth of $f(n)$. Therefore for any $f(n) \to \infty$ we have $\mathbb{P}_p(E_{2j}) \to 1$.

We now have shown

$$\mathbb{P}_p(E_{2j}) \leq \mathbb{P}_p(\mathcal{I}_{2j}) \leq \mathbb{P}_p(\mathcal{C}_{2j}) \to 1$$

and we conclude

$$p_c(2j,d) \leq p_{\mathcal{I}}(2j,d) < f(n)n^{-d/(j+1)-j}.$$

□

2.3.2 Lower Bound

In this section we prove the lower bound for for the critical exponent of $p_{\mathcal{I}}(2j,d)$. Again we assume $j(j+1) < d$.

Proposition 2.3.3. *For any $f(n) \to 0$, if $p = f(n)n^{-d/(j+1)-j}$, then*

$$\mathbb{P}_p(\mathcal{I}_{2j}) \to 0.$$

This implies $p_{\mathcal{I}}(2j,d) > f(n)n^{-d/(j+1)-j}$.

Proof. Let $\mathcal{V}_{2j}$ denote the set of all sublattices of $[n]^d$ that have dimension $2j$. The union bound gives:

$$\mathbb{P}_p(\mathcal{I}_{2j}) \leq \sum_{\dim(V)=2j} \mathbb{P}_p(\mathcal{I}_V) \leq \binom{d}{2j} n^{d-2j} M_{2j}(p).$$

The majority of the proof is showing $M_{2j}(p) = O(f(n)^{j+1}n^{2j-d})$ when $p = f(n)n^{-d/(j+1)-j}$. Then $\mathbb{P}_p(\mathcal{I}_{2j}) = O(f(n)^{j+1}) \to 0$ which implies $p_{\mathcal{I}}(2j,d) > f(n)n^{-d/(j+1)-j}$.

First let's start with the simplest possibilities for V: a single node, a line, and a plane.

- For a single node u,

$$\mathbb{P}_p(\mathcal{I}_u) = p.$$

- For a single line l,

$$\mathbb{P}_p(\mathcal{I}_l) = \mathbb{P}\big(\mathrm{Bin}(n,p) \geq 2\big) = \binom{n}{2} p^2 (1-p)^{n-2} = O(n^2 p^2).$$

- For a single plane P,

$$\mathbb{P}_p(\mathcal{I}_P) \leq \mathbb{P}\big(\mathrm{Bin}(n^2,p) \geq 2\big) = (1+o(1))2^{-1} n^4 p^2.$$

Note that a plane is more likely to be internally spanned than as an internally spanned line requires at least two collinear points.

Keeping the conditions of Proposition 2.3.3 and $p = f(n)n^{-d/(j+1)-j}$, we have the following lemma:

Lemma 2.3.4. *For* $1 \leq i \leq j$,

$$M_{2i-1}(p) = O(n^{i(i+3)-2} p^{i+1}). \tag{2.3}$$

$$M_{2i}(p) = (1+o(1))(2i)!2^{-i-1} n^{i(i+3)} p^{i+1}. \tag{2.4}$$

Proof. (By induction)

We assume the lemma holds for all $0 \leq l \leq 2i-2$ and show by induction that the formulas hold for dimensions $2i$ and $2i-1$. Note the lemma holds for a point, a line and a plane, so our base case is covered.

First let's assume a sublattice V is internally spanned. By Lemma 2.2.4, there exists proper sublattices V_1, $V_2 \subset V$ both internally spanned by disjoint subsets S_1 and S_2 such that $V = \langle V_1, V_2 \rangle$. Let D_V denote the set of possible pairs of such sublattices of V with

$\dim(V_1) \leq \dim(V_2)$. For V with dimension $2i$ or $2i-1$, let D'_V denote the subset of D_V where $\dim(V_1) = 0$, and $\dim(V_2) = 2i-2$.

$\mathcal{I}_V$ can be expressed as a union over D_V of events of the form $\mathcal{I}_{V_1} \circ \mathcal{I}_{V_2}$, where $\circ$ denotes the events occur disjointly.

We will show the probability $\mathcal{I}_V$ occurs is almost entirely determined by the probability there exists a pair $(V_1, V_2) \in D'_V$ such that $\mathcal{I}_{V_1} \circ \mathcal{I}_{V_2}$ occurs. Let $E_{D'_V} = \{\exists\ (V_1, V_2) \in D'_V\ s.t.\ \mathcal{I}_{V_1} \circ \mathcal{I}_{V_2}\}$. We have the following lemma.

Lemma 2.3.5. $\mathbb{P}_p(\mathcal{I}_V) = \mathbb{P}_p(E_{D'_V})(1 + O(n^{-1}))$.

Suppose this lemma is true. If $\dim(V) = 2i-1$, then there are at most $O(n^{2i-1}n)$ pairs in D'_V. The union bound gives

$$\mathbb{P}_p(E_{D'_V}) \leq \sum_{D'_V} M_0 M_{2i-2} \leq O(n^{2i}) M_0 M_{2i-2} = O\left(n^{i(i+3)-2} p^{i+1}\right),$$

proving the first part of Lemma 2.3.4.

If $\dim(V) = 2i$, there are at most $\binom{2i}{2}(n^{2i}n^2)$ pairs in D'_V. The union bound gives

$$\mathbb{P}_p(E_{D'_V}) \leq \binom{2i}{2} n^{2i} n^2 M_0 M_{2i-2} = (1 + o(1))(2i)! 2^{-i-1} n^{i(i+3)} p^{i+1},$$

proving the second part of Lemma 2.3.4.

Proof of Lemma 2.3.5 (By induction). We may assume Lemma 2.3.4 is true for up to dimension $2i-2$. This will allow us to prove Lemma 2.3.5 for dimension $2i-1$ and $2i$ hence proving Lemma 2.3.4 for those values as well. We can then proceed inductively to prove the lemmas are true for all values up to $2j$.

The union bound gives us:

$$\mathbb{P}_p(\mathcal{I}_V) \leq \mathbb{P}_p\left(\bigcup_{D_V} \mathcal{I}_{V_1} \circ \mathcal{I}_{V_2}\right).$$

$\mathcal{I}_{V_k}$ are increasing events. By the BK-inequality [32]

$$\mathbb{P}_p(\mathcal{I}_{V_1} \circ \mathcal{I}_{V_2}) \leq \mathbb{P}_p(\mathcal{I}_{V_1})\mathbb{P}_p(\mathcal{I}_{V_2})$$

and

$$\mathbb{P}_p(\mathcal{I}_V) \leq \sum_{D_V} \mathbb{P}_p(\mathcal{I}_{V_1} \circ \mathcal{I}_{V_2}) \leq \sum_{D_V} \mathbb{P}_p(\mathcal{I}_{V_1})\mathbb{P}_p(\mathcal{I}_{V_2}).$$

For now we assume the lemma is true for $\dim(V) = 2i - 1$. Then we prove the lemma is true for $\dim(V) = 2i$ and let $D_V(a, b)$ denote the subset of D_V where $\dim(V_1) = a$ and $\dim(V_2) = b$ and $a \leq b < 2i$. $\langle V_1, V_2 \rangle$ has at most dimension $a + b + 2$ if it is in fact a subspace. Therefore if $a + b + 2 < 2i$, then $D_V(a, b)$ is empty. Otherwise $|D_V(a, b)|$ is at most $O(n^{2i-a}n^{2i-b})$. Assume $a + b + 2 = 2i + \delta$ for some $\delta > 0$.

$$\mathcal{M}(a, b) = \sum_{D_V(a,b)} M_a M_b = O\left(n^{4i-a-b} M_a M_b\right).$$

If $a = 2l+1$, then $n^{2i-(2l-1)}M_{2l-1} \leq n^{-1}n^{2i-2l}M_{2l}$ so we may assume that a (and b) are both even. Let $a = 2i_1$, and $b = 2i_2$, with $i_1 + i_2 + 1 = i + k$. We know the values of M_{2i_1} and M_{2i_2}. Therefore

$$\begin{aligned}
\mathcal{M}(a, b) \leq &O(n^{4i-2i_1-2i_2} M_{2i_1} M_{2i_2}) \\
\leq &O(n^{4i-2i_1-2i_2} n^{i_1(i_1+3)+i_2(i_2+3)} p^{i_1+1+i_2+1}) \\
\leq &O(n^{i(i+3)} p^{i+1} n^{k(k-1)-2i_1 i_2}).
\end{aligned}$$

If $i_1 > 0$, then $k(k - 1) - 2i_1 i_2 < -1$ and Lemma 2.3.5 is true for $\dim(V) = 2i$. The proof for $\dim(V) = 2i - 1$ follows a similar approach.

□

With Lemma 2.3.5 proved, we can conclude from Lemma 2.3.4

$$M_{2i} \leq (2i!)2^{-i-1}n^{i(i+3)}p^{i+1}(1+o(1)).$$

This with Lemma 2.3.2 gives

$$M_{2i} = (2i!)2^{-i-1}n^{i(i+3)}p^{i+1}(1 \pm o(1)).$$

□

Combining these results we get

$$\mathbb{P}_p(\mathcal{I}_{2j}) \leq O(n^{d-2j})\mathbb{P}_p(\mathcal{I}_V) \leq O(n^{d-2j})o(n^{2j-d}) = o(1).$$

This implies

$$p_{\mathcal{I}}(2j,d) > f(n)n^{-d/(j+1)-j}$$

as desired.

2.3.3 Bounds for $p_c(2j, d)$

In this section we will show the bounds for the exponent of $p_{\mathcal{I}}(2j,d)$ hold for the exponent of $p_c(2j,d)$. We know the upper bound holds from previous arguments. We need to show the lower bound for $p_{\mathcal{I}}(2j,d)$ from Proposition 2.3.3 also holds for $p_c(2j,d)$.

If $\mathcal{C}_{2j}$ occurs then there exists some sublattice with dimension greater than or equal to $2j$ that is internally spanned. The next lemma will show that for any dimension $b > 2j$, $\mathbb{P}_p(\mathcal{I}_b) \to 0$ if $\mathbb{P}_p(\mathcal{I}_{2j}) \to 0$. This implies that $\mathbb{P}_p(\mathcal{C}_{2j}) \to 0$ as well.

Lemma 2.3.6. *If $p = f(n)n^{-d/(j+1)-j}$ as in Proposition 2.3.3 then $\mathbb{P}_p(\mathcal{C}_{2j}) \to 0$. Therefore*

$$p_c(2j,d) > f(n)n^{-d/(j+1)-j} \text{ for large } n.$$

Proof of Lemma 2.3.6. If $\mathcal{C}_{2j}$ occurs, then there must be some b-dimension maximal sublattice V_b such that $\mathcal{I}_{V_b}$ occurs and $b \geq 2j$. We know $\mathcal{I}_{2j} \to 0$ by the choice of p, so we need to show that $\mathbb{P}_p(\mathcal{I}_b) \to 0$ for $b > 2j$. By Lemma 2.2.4, if $\mathcal{I}_{V_b}$ occurs for some V_b, there exist V_1 and $V_2 \subset V_b$ with $\dim(V_1) \leq \dim(V_2) < b$ such that $\mathcal{I}_{V_1} \circ \mathcal{I}_{V_2}$ occurs and $\langle V_1, V_2 \rangle = V_b$. We may assume $\dim(V_2) \leq 2j$. For simplicity we assume the following

$$\begin{aligned}\dim(V_1) =& 2j - 2a_1\\ \dim(V_2) =& 2j - 2a_2\\ \dim(V_b) =& b = 2j + 2i\end{aligned}$$

with $a_1 \geq a_2 \geq 0$, and $0 < 2k < d - 2j$.

Since $\langle V_1, V_2 \rangle = V_b$ we have the that $j - a_1 - a_2 + 1 = i + k$ for some $k > 0$. Let

$$E_{i,k,a_1,a_2} = \{\exists\ V_1, V_2 \subset V_b\ s.t.\ \langle V_1, V_2 \rangle = V_b\}$$

that satisfy all the requirements above. Using $j(j+1) < d$ and $j - i - k + 1 = a_1 + a_2$ we get the following bound:

$$\begin{aligned}\mathbb{P}_p(E_{k,a_1,a_2}) =& O(n^{4j+4i-2j+2a_1-2j+2a_2} M_{2j+2a_1} M_{2j+2a_2})\\ =& O\left(n^{4i+2a_1+2a_2} n^{(j+a_1)(j+a_1+3)+(j+a_2)(j+a_2+3)} p^{j-a_1+j-a_2+2}\right)\\ =& O\left(n^{-2d-2j(j+1)+6j+4i+5a_1+5a_2+2j^2+2ja_1+2ja_2+a_1^2+a_2^2+d(a_1+a_2)/(j+1)}\right)\\ \leq& O(n^{-d})\end{aligned}$$

There are only finitely many choices for b, a_1, and a_2 and only $O(n^{d-b})$ sublattices of dimension b. Therefore the probability there exists an internally spanned sublattice of dimension greater than $2j$ tends to zero.

□

Now we can conclude that $p_c(2j,d)$ is also bounded below $f(n)n^{-d(j+1)-j}$ for any $f(n) \to 0$.

□

2.3.4 $p_c(2j,d)$ for $j(j+1) > d$

Let $j' = \sup\{i : i(i+1) < d\}$. For $j(j+1) > d$ and $f(n) \to \infty$, suppose $p > n^{-d/(j'+1)-j'}$. For any $2j$-dimension sublattice, the expected number of nodes that are initially open is $n^{2j}p \geq n^{2j-2j'} \to \infty$ and hence

$$\mathbb{P}_p(\mathcal{I}_{2j}|\mathcal{I}_{2j-2}) \to 1.$$

Therefore $p_c(2j,d) \to p_c(2j-2,d)$. The case where $j(j+1) = d$ will require a little more work and will be presented in Section 6.

2.4 Poisson Approximation

We use the Stein-Chen method for approximation by a Poisson distribution. We will use the version found in [30]:

Theorem 2.4.1. *Let $X_1, \ldots, X_m$ be indicator variables with $\mathbb{P}(X_i = 1) = p_i$, $Y = \sum_{i=1}^n X_i$, and $\lambda = \mathbb{E}[Y] = \sum_i p_i$. For each $i \in [m]$, let $N_i \subset [m]$ where $i \in N_i$ and X_i is independent of $\{X_j : j \notin N_i\}$. If $p_{ij} := \mathbb{E}[X_i X_j]$ and $Z \sim Po(\lambda)$, then*

$$d_{TV}(Y,Z) \leq \sum_{i=1}^{m} \left(\sum_{j\in N_i} p_i p_j + \sum_{j\in N_i\setminus\{i\}} p_{ij} \right). \tag{2.5}$$

Let Γ denote the set of all $2j$-dimension sublattices in $[n]^d$. Each sublattice V has a dependency set Γ_V where $\mathcal{I}_W$ depends on $\mathcal{I}_V$ for $W \in \Gamma_V$. When $p = an^{-d/(j+1)-j}$ then each subspace $V \in \Gamma$ is internally spanned with probability $(2j)!2^{-j-1}a^{j+1}n^{2j-d}$. Although some dependency exists, if $j(j+1) < d$, we will show that the distribution of the number of sublattices with dimension $2j$ which are internally spanned approaches a Poisson distribution.

To fit our random variables with that of the theorem, we let $\mathbf{1}_V$ denote the indicator random variable for the event $\mathcal{I}_V$. For all $V, W \in \Gamma$,

$$p_V = p_W = M_{2j} = (2j)!2^{-j-1}a^{j+1}n^{2j-d}(1 \pm o(1)) \tag{2.6}$$

and

$$p_{VW} = \mathbb{E}[\mathbf{1}_V \mathbf{1}_W] = \mathbb{P}(\mathcal{I}_V \cap \mathcal{I}_W). \tag{2.7}$$

Let $Y = \sum_\Gamma \mathbf{1}_V$. Then

$$\lambda = \mathbb{E}[Y] = (1 + o(1)) \sum_\Gamma (2j)!2^{-j-1}a^{j+1}n^{2j-d} = \binom{d}{2j}(2j)!2^{-j-1}a^{j+1}. \tag{2.8}$$

Finally we let $Z \sim \mathrm{Po}(\lambda)$, a Poisson random variable with parameter λ.

Plugging everything into (2.5) we get

$$d_{TV}(Y, Z) \leq \sum_{V \in \Gamma} \left(\sum_{W \in \Gamma_V} p_V p_W + \sum_{W \in \Gamma_V \setminus \{V\}} p_{VW} \right). \tag{2.9}$$

Since

$$\sum_{W \in \Gamma_V} p_V p_W + \sum_{W \in \Gamma_V \setminus \{V\}} p_{VW} \tag{2.10}$$

does not depend on the choice of V, we can approximate the right-hand side of (2.9) by

$$|\Gamma||\Gamma_V|M_{2j}^2 + |\Gamma| \sum_{W \in \Gamma_V \setminus \{V\}} p_{VW} \tag{2.11}$$

The quantity p_{VW} depends on the dimension $\dim(V \cap W) = l$. We break up Γ_V into subsets Γ_V^l where $W \in \Gamma_V^l$ if $\dim(V \cap W = l)$.

For each l, $|\Gamma_V^l| = O(n^{2j-l})$ so $|\Gamma_V| = O(n^{2j})$. This gives the bound

$$|\Gamma||\Gamma_V|M_{2j}^2 = O(n^{4j-d}) \to 0$$

for the left half of (2.11). The remaining portion of the (2.11) requires a bit more work. We compute upper bounds for p_{VW} that depend on l.

As before $j(j+1) < d$ and $p = an^{-d/(j+1)-j}$. We state a slightly more general lemma in that we have $\dim(V) = \dim(W) = 2i \leq 2j$.

Lemma 2.4.2. *Let* $\dim(V \cap W) = 2i - 2k$ *with* $0 \leq k \leq i$. *Then,*

$$p_{VW} = \mathbb{P}_p(\mathcal{I}_V \cap \mathcal{I}_W) = O(n^{4ik-2k(k-1)+(i-k)(i-k+3)}p^{i+k+1}). \tag{2.12}$$

In particular, if $i = j$ *then for some* $\epsilon > 0$,

$$\mathbb{P}_p(\mathcal{I}_V \cap \mathcal{I}_W) = O(n^{2j-2k-d-\epsilon}). \tag{2.13}$$

This upper bound also holds for $\dim(V \cap W) = 2i - 2k + 1$ though we will always assume even dimension intersection for simplicity.

Proof of Lemma 2.4.2 (By induction). In this proof we will use induction on both i and k. Our base case of $i = k = 0$ is satisfied. To continue we state two useful sublemmas.

Sublemma 2.4.3. *Let* $V' \subset V$ *be sublattices that satisfy* $\dim(V) = 2i$ *and* $\dim(V') = 2i-2k$ *or* $2i - 2k + 1$. *Then*

$$\mathbb{P}(\mathcal{I}_V \big| \mathcal{I}_{V'}) = O(n^{2ki-k(k-1)}p^k). \tag{2.14}$$

Sublemma 2.4.4. *Let* V *and* W *be sublattices with non-trivial intersection. Then*

$$\mathbb{P}(\mathcal{I}_V \cap \mathcal{I}_W) = O\left(\mathbb{P}(\mathcal{I}_V \big| \mathcal{I}_{V\cap W})\mathbb{P}(\mathcal{I}_W \big| \mathcal{I}_{V\cap W})\mathbb{P}(\mathcal{I}_{V\cap W})\right). \tag{2.15}$$

Combining Lemma 2.3.4, Sublemma 2.4.3, where $V' = V \cap W$, and Sublemma 2.4.4 we get

$$\mathbb{P}(\mathcal{I}_V \cap \mathcal{I}_W) \quad = O\left(n^{4ki-2k(k-1)+(i-k)(i-k+3)}p^{i+k+1}\right). \tag{2.16}$$

When $i = j$ we get for some $\epsilon > 0$,

$$\mathbb{P}(\mathcal{I}_V \cap \mathcal{I}_W) = O(n^{j(j+3)}p^{j+1}(n^{2j-k-1}p)^k) \tag{2.17}$$

$$= O(n^{2j-d}n^{-k(k+1)}n^{-\epsilon}) \tag{2.18}$$

$$= O(n^{2j-2k-d-\epsilon}) \tag{2.19}$$

where we use the simplification $n^{2j}p \leq n^{-\epsilon}$. Assuming both Sublemmas 2.4.3 and 2.4.4 are true, this proves Lemma 2.4.2.

Proof of Sublemma 2.4.3 (By induction). Let $V' \subset V$ satisfy $\dim(V) = 2i$ and $\dim(V') = 2i - 2k$ where $0 \leq k \leq i$. We assume the lemma is true for all pairs of values (i', k') such that if $i' < i$, then $0 \leq k' \leq i'$ and if $i' = i$, $0 \leq k' < k$. The statement holds for all i if $k = 0$, covering our base case.

If $\mathcal{I}_V$ occurs, then from Lemma 2.2.4 there exists $(V_1, V_2) \in D_V$ such that $\mathcal{I}_{V_1} \circ \mathcal{I}_{V_2}$ occurs. Similar to the proofs of Lemma 2.3.4 and Lemma 2.3.5, the union bound for the conditional probability will be dominated by pairs $(V_1, V_2) \in D_V$ that satisfy $\dim(V_1) = 0$ and $\dim(V_2) = 2i - 2$. We denote this subset of D_V by D'_V.

$$\mathbb{P}(\mathcal{I}_V \big| \mathcal{I}_{V'}) \leq \sum_{(V_1,V_2)\in D_V} \mathbb{P}(\mathcal{I}_{V_1} \big| \mathcal{I}_{V'})\mathbb{P}(\mathcal{I}_{V_2} \big| \mathcal{I}_{V'}). \tag{2.20}$$

$$\leq O(1) \sum_{(v_1,V_2)\in D'_V} \mathbb{P}(\mathcal{I}_{v_1} \big| \mathcal{I}_{V'})\mathbb{P}(\mathcal{I}_{V_2} \big| \mathcal{I}_{V'}) \tag{2.21}$$

If $v_1 \notin V'$ then there are $O(1)$ choices of V_2 such that $V' \subset V_2$ and $O(n^2)$ choices such that $\dim(V_2 \cap V') < \dim(V')$. When $v_1 \in V'$ all $O(n^2)$ choices of V_2 have $\dim(V_2 \cap V') < \dim(V')$. There are less than n^{2i} choices of $v_1 \notin V'$ and n^{2i-2k} choices of $v_1 \in V'$. Let V_2^*, V_2^{**}, and V_2^{***} denote representatives from each of these choices of V_2. Expectation gives us the upper bound

$$\mathbb{P}(\mathcal{I}_V \big| \mathcal{I}_{V'}) = O(1)n^{2i}p\mathbb{P}(\mathcal{I}_{V_2^*}|\mathcal{I}_{V'}) \tag{2.22}$$

$$+ n^{2i}pO(n^2)\mathbb{P}(\mathcal{I}_{V_2^{**}} \big| \mathcal{I}_{V_2^{**}\cap V'}) \tag{2.23}$$

$$+ n^{2i-2k}O(n^2)\mathbb{P}(\mathcal{I}_{V_2^{***}} \big| \mathcal{I}_{V_2^{***}\cap V'}). \tag{2.24}$$

Here we apply the inductive hypothesis to each of these terms. The contribution from (2.23) and (2.24) will be negligible compared to the right-hand side of (2.22). This gives

$$\begin{aligned}\mathbb{P}(\mathcal{I}_V \big| \mathcal{I}_{V'}) &= O(n^{2i}pn^{2(i-1)(k-1)-(k-1)(k-2)}p^{k-1}) \\ &= O(n^{2ik-k(k-1)}p^k).\end{aligned}$$

□

Proof of Sublemma 2.4.4 (By induction). The direction of the induction is the reverse of Sublemma 2.4.3. We have $\dim(V) = 2i$, $\dim(W) = 2i'$, $\dim(V \cap W) = 2i - 2k$ and will assume the sublemma is true if either $i' < i$ or $k' > k$.

$$\begin{aligned}\mathbb{P}(\mathcal{I}_V \cap \mathcal{I}_W) \leq &\sum_{(W_1,W_2)\in D_W} \mathbb{P}(\mathcal{I}_V \cap \{\mathcal{I}_{W_1} \circ \mathcal{I}_{W_2}\}) \\ =&O(1) \sum_{(w_1,W_2)\in D'_W} \mathbb{P}(\mathcal{I}_V \cap \{\mathcal{I}_{w_1} \circ \mathcal{I}_{W_2}\}) \\ =&O(1) \sum_{(w_1,W_2)\in D'_W} \mathbb{P}(\mathcal{I}_V \big| \{\mathcal{I}_{w_1} \circ \mathcal{I}_{W_2}\})\mathbb{P}(\{\mathcal{I}_{w_1} \circ \mathcal{I}_{W_2}\})\end{aligned}$$

The terms where $w_1 \notin V$ and $V \cap W = V \cap W_2$ dominate this sum. There are at most n^{2i} such w_1 and $O(1)$ such W_2.

$$\begin{aligned}\mathbb{P}(\mathcal{I}_V \cap \mathcal{I}_W) =&O(1) \sum_{\substack{w_1 \notin V \\ V\cap W \subset W_2}} \mathbb{P}(\mathcal{I}_V \big| \mathcal{I}_{W_2})\mathbb{P}(\mathcal{I}_{w_1})\mathbb{P}(\mathcal{I}_{W_2}) \\ =&O(1)n^{2i}p\mathbb{P}(\mathcal{I}_V \cap \mathcal{I}_{W_2})\end{aligned}$$

for some choice of W_2. Since $\dim(W_2) < 2i$ we can apply now use the inductive hypothesis to get

$$\begin{aligned}\mathbb{P}(\mathcal{I}_V \cap \mathcal{I}_W) =& O(1)n^{2i}p\mathbb{P}(\mathcal{I}_V \big| \mathcal{I}_{V\cap W_2})\mathbb{P}(\mathcal{I}_{W_2}\big|\mathcal{I}_{V\cap W_2})\mathbb{P}(\mathcal{I}_{V\cap W_2})\\ =& O(1)\mathbb{P}(\mathcal{I}_V\big|\mathcal{I}_{V\cap W})\mathbb{P}(\mathcal{I}_W\big|\mathcal{I}_{V\cap W})\mathbb{P}(\mathcal{I}_{V\cap W})\end{aligned}$$

as desired.

□

A slightly modified argument will show the same upper bound holds when $\dim(V\cap W) = 2i-2k+1$. Also the base case where $\dim(V\cap W) = 0$ holds. We conclude that

$$|\Gamma_V^{2j-2k+1}|p_{VW} \leq |\Gamma_V^{2j-2k}|p_{VW} \leq O(n^{2k}n^{2j-2k-d-\epsilon}) \leq O(n^{2j-d-\epsilon}).$$

□

Plugging this into the Stein-Chen bound we conclude that $d_{TV}(Y,Z) \to 0$ and the number of internally spanned sublattices are approximately Poisson with parameter $\lambda = \binom{d}{2j}(2j)!2^{-j-1}a^{j+1}$. This proves Theorem 2.1.2.

2.5 *Proofs of Theorems*

The proof of Theorem 2.1.1 follows from Theorem 2.1.2 and Lemma 2.3.6. These combine to show $\mathbb{P}_p(\mathcal{C}_{2j}\backslash\mathcal{I}_{2j}) \to 0$.

For Theorem 2.1.3 we have to do a little work. If $d = j(j+1)$ then

$$d/(j+1) + j = d/j + j - 1 = 2j.$$

Let $p = an^{-2j}$, $\lambda' = \lambda_{2j-2} = \binom{d}{2j-2}(2j-2)!2^{-j+1}a^j$, and Y_{2j-2} denote the number of sublattices of dimension $2j-2$ that are internally spanned. By Theorem 2.1.2

$$d_{TV}(Y_{2j}, Z(\lambda)) \to 0,$$

where $Z(\lambda')$ is again a Poisson(λ') random variable. For each $k \geq 0$ we have

$$\mathbb{P}_p(Y_{2j-2} = k) \to \frac{e^{-\lambda'}}{k!}\lambda'^k. \tag{2.25}$$

For each of these k open sublattices, there are exactly $c = \binom{d-2j+2}{2}$ distinct sublattices with dimension $2j$. The number of nodes u with distance exactly 2 away from one of the open sublattices is $cn^{2j}(1 - o(1))$. Although it is possible for two open sublattices of dimension $2j - 2$ to exist in the same $2j$-dimension sublattice, from the this event has probability tending to zero. The probability that there exists some $2j$-dimension sublattice with two disjoint open $(2j - 2)$-dimension sublattices is bounded by

$$O(n^{d-2j}n^4n^{4j-4-2d}) = O(n^{j-j^2}).$$

This tends to zero if $j > 1$. When $j = 1$, $d = 2$ and we are dealing with a plane, which is well understood.

Otherwise, there are in total $ckn^{2j}(1 - o(1))$ that, if open, would lead to a sublattice of dimension $2j$ that is internally spanned. The probability that none of these are open is given by $(1 - p)^{ckn^{2j}(1-o(1))}$. Hence

$$\mathbb{P}(\mathcal{I}_{2j}) = \sum_{k=1}^{\infty} \mathbb{P}_p(Y_{2j-2} = k)\left(1 - (1 - an^{-2j})^{ckn^{2j}(1-o(1))}\right),$$

which for large n gives

$$\mathbb{P}_p(\mathcal{I}_{2j}) \to \sum_{k=1}^{\infty} \frac{e^{-\lambda}}{k!}\lambda^k(1 - e^{-ack}),$$

proving the theorem.

For Corollary 2.1.4 we look at the three cases. In each case we assume $p = an^{-d/(j+1)-j}$.

- $(j + 1)(j + 2) < d$. By Theorem 2.1.2 $\mathbb{P}_p(\mathcal{I}_{2j+2}) \to 0$, and $\mathbb{P}_p(\mathcal{I}_{2j-2}) \to 1$. Therefore the largest sublattice has either dimension $2j - 2$ with probability $e^{-\lambda}$ or dimension $2j$ with probability $1 - e^{-\lambda}$. In terms of the random variable D, we have

$$\mathbb{P}(D = 2j - 2) \to e^{-\lambda},$$
$$\mathbb{P}(D = 2j) \to 1 - e^{-\lambda}.$$

- $j(j+1) < d < (j+1)(j+2)$. Similar to the previous case, there are no internally spanned sublattices with dimension $2j$ with probability $e^{-\lambda}$ leaving the maximal sublattice to have dimension $2j-2$. However if there is an open $2j$-dimension sublattice, then the expected number of open nodes exactly distance 2 away is $O(n^{2j+2}p)$. Since this expectation tends to infinity, with probability tending to 1, an open $2j$-dimension sublattice will become an open $(2j+2)$-dimension sublattice. This will continue until all of $[n]^d$ is open. Hence with probability tending to $1 - e^{-\lambda}$, the maximal sublattice will be the entire space. This gives

$$\mathbb{P}_p(D = 2j - 2)) \to e^{-\lambda},$$
$$\mathbb{P}_p(D = d) \to 1 - e^{-\lambda}.$$

- $j(j+1) = d > 6$. With probability $e^{-\lambda'}$ no sublattice with dimension $2j-2$ or greater is open, and with probability tending to one there is a open sublattice with dimension $2j-4$. Therefore
$$\mathbb{P}_p(D = 2j - 4) \to e^{-\lambda'}.$$
For $2j-2$ to be the maximum dimension of open sublattices, there must be some positive number k of such sublattices open, but no neighbors distance 2 away from those sublattices can be open. Hence
$$\mathbb{P}_p(\mathcal{D}_{2j-2}) = \sum_{k=1}^{\infty} \frac{e^{-\lambda'}}{k!} \lambda'^k e^{-ack}.$$
If there exists an open $2j$-dimension sublattice, then by the arguments in the previous case the entire lattice, $[n]^d$ becomes open. The limiting value for $\mathbb{P}_p(\mathcal{I}_{2j})$ is given in Theorem 2.1.3. Altogether we have

$$\mathbb{P}_p(D = 2j - 4) \to e^{-\lambda'},$$
$$\mathbb{P}_p(D = 2j - 2) \to \sum_{k=1}^{\infty} \frac{e^{-\lambda'}}{k!} \lambda'^k e^{-ack},$$
$$\mathbb{P}_p(D = d) \to 1 - \sum_{k=0}^{\infty} \frac{e^{-\lambda'}}{k!} \lambda'^k e^{-ack}.$$

- $d = 6$ $(j = 2)$. We still have with probability $e^{-\lambda'}$ that no plane (dimension $2j - 2$) is open. The big difference is that if $k \geq 2$ planes are open, there is a non-trivial probability that two of the planes are exactly distance two from each other, in which case the entire space would become open and $D = d$. A plane embedded in a 6-dimension space is determined by the values of the 4 fixed coordinates. The other two coordinates we call the free coordinates as they take on all values in $[n]$. If the free coordinates of the two planes do not overlap, then the planes are exactly distance 2 apart. Let d_k denote the probability that for k distinct planes, at least two have free coordinates that do not overlap.

For a plane P, let $N(P)$ denote the set of at most cn^4 possible nodes, $u \in [n]^6$, such that $\dim(\langle u, P \rangle) = 4$. With probability tending to 1, the number of nodes in both $N(P_s) \cap N(P_t)$ is at most $o(n^4)$ for all $1 \leq s < t \leq k$. Hence the total number of nodes that cause at least one of k planes to evolve into an internally spanned 4-dimension sublattice is at least $ckn^4 - o(n^4)$. The number of nodes that determine d_k is only $O(n^2)$ so if we remove those we still have at least $ckn^4 - o(n^4)$ nodes remaining that would cause a 4-dimensional sublattice to be internally spanned. This occurs with probability at least $(1 - an^{-4})^{ckn^4 - o(n^4)} = e^{-ack}(1 - o(1))$. Therefore

$$\begin{aligned}
\mathbb{P}_p(D=0) &\to e^{-\lambda'}, \\
\mathbb{P}_p(D=2) &\to \sum_{k=1}^{\infty} \frac{e^{-\lambda'}}{k!} \lambda'^k (1-d_k) e^{-ack}, \\
\mathbb{P}_p(D=6) &\to 1 - e^{-\lambda'} - \sum_{k=1}^{\infty} \frac{e^{-\lambda'}}{k!} \lambda'^k (1-d_k) e^{-ack}.
\end{aligned}$$

Chapter 3

JIGSAW PERCOLATION

This section is an extension of the work done by Brummitt et. al. in [34]. They define a new kind of percolation on finite graphs called *jigsaw percolation*. It can be viewed as a model for collaborative networks working together on a problem, where each node represents a person with a unique part to play in solving the problem.

Jigsaw percolation is a process on a graph with two distinct edge sets $G = (V, E_{\text{puzzle}}, E_{\text{people}})$. The edges in E_{people} denote acquaintances. The edges in E_{puzzle} denote connections that need to be made in order for the problem to be solved. The process begins with everyone in disjoint clusters of size one. Initially if two people share a puzzle edge and are also acquaintances they merge together to form a larger cluster. For each subsequent step, clusters merge together if there exists at least one pair of nodes between the clusters connected by a people edge and at least one pair of nodes connected by a puzzle edge. If the process ends with all nodes in a single cluster we say the people graph *solves* the puzzle graph.

We will consider the case where (V, E_{people}) and (V, E_{puzzle}) are Erdös-Rényi random graphs with probabilities p_{ppl} and p_{puz} respectively. Our results will be stated in terms of the product of these probabilities, $p_{\text{eff}} = p_{\text{ppl}} p_{\text{puz}}$. For E_{people} to solve E_{puzzle} we need that both (V, E_{people}) and (V, E_{puzzle}) to be connected. Hence for simplicity we will assume for some $0 < \epsilon < 1$, both $p_{\text{ppl}} \geq p_{\text{puz}} \geq n^{\epsilon}/n$, guaranteeing connectivity with high probability for both graphs.

Definition 3.0.1. *Jigsaw Percolation*

We repeat the definition of jigsaw percolation in [34]. Jigsaw percolation proceeds on $(V, E_{\text{people}}, E_{\text{puzzle}})$ as follows. At every step $i \geq 0$, we have a partition, $\mathcal{C}_i$, of V into disjoint subsets, where if $A \in \mathcal{C}_i$ then the nodes of A will have merged by step i. At step i, let $\mathcal{E}_i$ denote the unordered pairs of clusters in $\mathcal{C}_i$ that are both people- and puzzle-adjacent. We get $\mathcal{C}_{i+1}$ by merging the connected components of $(\mathcal{C}_i, \mathcal{E}_i)$.

$$\begin{aligned} \mathcal{C}_0 &= \{\{v\} : v \in V\} \\ \mathcal{C}_1 &= \{A : A \text{ is a connected component of } (V, E_{\text{puzzle}} \cap E_{\text{people}})\} \\ \mathcal{C}_{i+1} &= \{\cup A \in UA : U \text{ is a connected component of } (\mathcal{C}_i, \mathcal{E}_i)\}\,. \end{aligned}$$

We are particularly interested in the event **Solve** $:= \{\mathcal{C}_\infty = \{V\}\}$, when the people graph solves the puzzle graph. We hope to find good bounds for the critical effective probability p^c_{eff}, where $\mathbb{P}_{p^c_{\text{eff}}}(\textbf{Solve}) = 1/2$.

The following proposition gives upper and lower bounds on p^c_{eff}.

Theorem 3.0.1.

1. *If* $p_{\mathit{eff}} \leq \frac{e^{-5}}{n \log n}$, *then* $\mathbb{P}_{p_{\mathit{eff}}}(\boldsymbol{Solve}) \to 0$.

2. *If* $p_{\mathit{eff}} \geq \frac{64 \operatorname{loglog} n}{n \log n}$, *then* $\mathbb{P}_{p_{\mathit{eff}}}(\boldsymbol{Solve}) \to 1$.

Combined we get the upper and lower bounds

$$\frac{e^{-5}}{n \log n} < p^c_{\mathit{eff}} < \frac{64 \operatorname{loglog} n}{n \log n}.$$

We'll proceed by first proving the lower bound for p_{eff}. The proof for the upper bound is a little more involved. We will sketch the arguments for the simpler case where $p_{\text{ppl}} = p_{\text{puz}} = (2c/n)^{1/2}$. The proof for $p_{\text{eff}} \geq 64 \operatorname{loglog} n / n \log n$ will follow with minor adjustments.

3.1 Lower Bound for p_{eff}

Let $S_i \in \mathcal{C}_i$ denote the largest cluster after i steps. We assume that $p_{\text{eff}} = e^{-5}/n \log n$. For a cluster of vertices, $A \subset V$, to have merged together at some point in the jigsaw percolation process, the induced graphs (A, E_{people}) and (A, E_{puzzle}) must both be connected. If both subgraphs are connected then both subgraphs must contain a spanning tree. We denote the existence of two such spanning trees on A by the event $\mathbf{DualSpan}_A$. Furthermore, we denote with $\mathbf{DualSpan}_k$ the existence of such a A where $|A| = k$. The following few lemmas will show that

$$\mathbf{Solve} \subset \left\{ \bigcup_{j=\log n}^{2\log n} \mathbf{DualSpan}_j \right\} \cup \{|S_1| \geq \log n\}$$

and the probability of the event on the right tends to 0 as n increases.

Proof of Theorem 3.0.1 (a), $p_{\text{eff}} \leq \frac{e^{-5}}{n \log n}$.

Lemma 3.1.1. *If $\exists\ k$ s.t. $|S_1| < k < n/2$ and for no $j \in [k, 2k]$ does $\boldsymbol{DualSpan}_j$ occur, then **Solve** does not occur.*

Proof. If **Solve** occurs, then at some point clusters of small size must merge together to form one giant cluster. Although many clusters may merge together in one step, we can break down each step into a series of substeps where exactly two clusters merge. The upper bound for the largest cluster at most doubles in size during each of these substeps. Eventually two clusters of size less than k must merge to form a cluster, A, with $|A| \in [k, 2k]$. Therefore if this does not happen, **Solve** cannot occur. □

Lemma 3.1.2. *If $p < e^{-5}/(n \log n)$, $\mathbb{P}(|S_1| < \log n) \to 1$.*

Proof. Two vertices merge after one step if they are connected by both a people edge and a puzzle edge, so the largest cluster after the first step can be understood by looking at Erdös-Rényi random graphs with edge probability $p = p_{\text{eff}}$. Let T_k denote the number of trees of size k in a Erdös-Rényi random graph with such an edge probability. From [29], the expectation of T_k is given by

$$\begin{aligned}\mathbb{E}T_k &= \binom{n}{k} k^{k-2} {p_{\text{eff}}}^{k-1} (1 - p_{\text{eff}})^{k(n-k)+\binom{k}{2}-k+1} \\ &\leq \binom{n}{k} k^{k-2} {p_{\text{eff}}}^{k-1}.\end{aligned}$$

Plugging in $k = \log n$ in the above expression shows that $\mathbb{E}T_k \to 0$ as $n \to \infty$. As T_k is the sum of positive random variables, with high probability we have that $\mathbb{P}(|S_1| < \log n) \to 1$.

□

Lemma 3.1.3. *Let* $|A| = k$*, then*

$$\mathbb{P}(\boldsymbol{DualSpan}_A) \leq k^{2k-4} {p_{\mathit{eff}}}^{k-1},$$

and

$$\mathbb{P}(\boldsymbol{DualSpan}_k) \leq \binom{n}{k} k^{2k-4} {p_{\mathit{eff}}}^{k-1}.$$

Proof. We use expectation to bound the probabilities in question. Fix a subset of vertices A of size k. Let T_{ppl} and T_{puz} denote the number of spanning trees in (A, E_{people}) and (A, E_{puzzle}) respectively. The event $\mathbf{DualSpan}_A$ is the intersection of the two independent events $\{T_{\text{ppl}} \geq 1\} \cap \{T_{\text{puz}} \geq 1\}$. As there are k^{k-2} possible spanning trees on A we have

$$\mathbb{P}(T_{\text{ppl}} \geq 1) \leq \mathbb{E}T_{\text{ppl}} = k^{k-2} p_{\text{ppl}}^{k-1}$$

and

$$\mathbb{P}(T_{\text{puz}} \geq 1) \leq \mathbb{E}T_{\text{puz}} = k^{k-2} p_{\text{puz}}^{k-1}.$$

Since the events are independent we may take the product to get

$$\mathbb{P}(\mathbf{DualSpan}_A) = \mathbb{P}(T_{\text{ppl}} \geq 1)\mathbb{P}(T_{\text{puz}} \geq 1) \leq k^{2k-4} {p_{\text{eff}}}^{k-1}$$

as desired. Since there are $\binom{n}{k}$ choices for A we get

$$\mathbb{P}(\mathbf{DualSpan}_k) \leq \binom{n}{k} k^{2k-4} {p_{\text{eff}}}^{k-1},$$

the expected number of subsets of vertices, A, such that $|A| = k$ and $\mathbf{DualSpan}_A$ occurs.

□

Lemma 3.1.4. $\mathbb{P}(\boldsymbol{Solve}) \leq \mathbb{P}(|S_1| \geq \log n) + \sum_{k=\log n}^{2\log n} \mathbb{P}(\boldsymbol{DualSpan}_k)$.

Proof. By Lemma 3.1.1, the event $\mathbf{DualSpan}_k$ occurs for some $k \in (\log n, 2\log n)$. Using the union bound with Lemmas 3.1.2 and 3.1.3 gives:

$$
\begin{aligned}
\mathbb{P}(\mathbf{Solve}) &\leq \sum_{k=\log n}^{2\log n} \mathbb{P}(\mathbf{DualSpan}_k) + \mathbb{P}(|S_1| \geq \log n) \\
&\leq \sum_{i=0}^{\log n} \frac{ne^{i+\log n}(i+\log n)^{i+\log n} \log n e^{-5i-5\log n}}{\sqrt{2\pi(i+\log n)}(\log n)^{(i+\log n)}(i+\log n)^4} + o(1) \\
&\leq \frac{ne^{2\log n}(2)^{2\log n} e^{-5\log n}}{\sqrt{2\pi \log n}} \\
&\leq e^{(5-.25-5)\log n} \to 0.
\end{aligned}
$$

□

The constant e^{-5} is in no way optimal. By Lemma 3.1.2 the largest cluster in $\mathcal{C}_1$ is bounded by $\log n$ with high probability, so $\mathbb{P}(\mathbf{Solve}) \to 0$ as desired.

□

3.2 *Upper Bound for* p_{eff}

The proof of part (b) is a bit more involved. We first show with high probability there exists a tree in $(V, E_{\text{puzzle}} \bigcap E_{\text{people}})$ that is large enough to act as a seed, growing until it merges with every other vertex.

Split V into two sets V_1 and V_2 with the same cardinality. We inductively define the following sets:

- $S_1 := \{$ vertices in the largest connected component of $(V_1, E_{\text{people}} \cap E_{\text{puzzle}})\}$.
  We call S_1 the seed.

- $R_1 := V_2$.

- $S_{i+1} := \{v \in R_i \ s.t. \ \exists \ u, w \in S_i$ with $(u, v) \in E_{\text{people}}, (w, v) \in E_{\text{puzzle}}\}$,
  the vertices that have both a people and puzzle edge connecting them to S_i.

- $R_{i+1} := \{v \in V_2 \setminus \bigcup_{j=1}^{i+1} S_j\}$, the vertices in V_2 not yet merged with the big cluster.

3.2.1 $p_{\mathit{eff}} = 2c/n$

First we analyze the simpler case where $p_{\text{ppl}} = p_{\text{puz}} = (2c/n)^{1/2}$, where $c > 0$. Classical random graph arguments show that the size of S_1 is at least $O(\log n)$ (and much greater if $2c \geq 1$). These vertices merge with roughly $O((\log n)^2)$ new vertices in R_1 which in turn merge with $O((\log n)^4)$ vertices in R_2. In general, at the $(i+1)$st step we merge $O((\log n)^{2^i})$ vertices in R_i with S_i as long as $(\log n)^{2^i} \ll n$. Once we have an S_{i_n} with over $n^{1/2}$ vertices we have with high probability that every remaining vertex in R_{i_n} merges with S_{i_n}.

The size of S_1 will be at least $4 \log n$ with high probability. The 4 is chosen for ease of exposition. For each $v \in R_1$ the probability that v is connected to $u \in S_1$ by a people edge is given by $p = (2c/n)^{1/2}$. The same probability holds for a puzzle edge. These edges exist independently.

The probability that v is connected to S_1 by both a people edge and a puzzle edge is bounded below by

$$\left(1-(1-p_{\text{ppl}})^{|S_1|}\right)\left(1-(1-p_{\text{puz}})^{|S_1|}\right) \geq (2p\log n)^2. \tag{3.1}$$

The expected number of vertices in R_1 that connect in this way is bounded below by

$$n(1-o(1))\frac{4c(\log n)^2}{n} \geq 2c(\log n)^2. \tag{3.2}$$

Each $v \in R_1$ merges with S_1 independently so we can use Chernoff bounds to find a lower bound for the probability that $|S_2|$ is at least half the expected value of $|S_2|$,

$$\mathbb{P}\left(|S_2| \geq c(\log n)^2\right) \geq 1-\exp\left(-\frac{1}{2}c(\log n)^2\right). \tag{3.3}$$

Assuming this lower bound for $|S_2|$ holds, the expectation of $|S_3|$ will be at least $2c|S_2|^2 = 2c^3(\log n)^4$. The Chernoff bounds show that

$$\mathbb{P}\left(|S_3| \geq c^3(\log n)^4\right) \geq 1-\exp\left(-\frac{1}{2}c^3(\log n)^4\right). \tag{3.4}$$

In general we have for $i > 2$ not too large ($(c\log n)^{2^{i-1}} \ll n^{1/3}$), the edges connecting a vertex in R_i to S_i exist independently in both the people and puzzle graph with probability p. The expectation of $|S_i|$ is greater than $\frac{2}{c}(c\log n)^{2^{i-1}}$ so again Chernoff bounds show that

$$\mathbb{P}\left(|S_i| \geq \frac{1}{c}\left(c\log n\right)^{2^{i-1}}\right) \geq 1-\exp\left(-\frac{1}{2c}\left(c\log n\right)^{2^{i-1}}\right). \tag{3.5}$$

By "not too large" we mean that for some positive $\alpha < 1/2$, $(c\log n)^{2^{i-1}} \ll n^{\alpha}$. In particular we have that $i < \operatorname{loglog} n$ which in turn makes $R_i \geq n - n^{1/2}\operatorname{loglog} n$. If $|S_i|$ is larger than $n^{1/2}$, $|S_{i+1}| \geq n(1-o(1))$ and $|S_{i+2}| = n$ with high probability.

3.2.2 $p_{\mathit{eff}} = 64\operatorname{loglog} n/n\log n$

For notational simplicity we let $x = \log n / \operatorname{loglog} n$, so $p_{\text{eff}} = 64x^{-1}n^{-1}$. We will also assume for some $0 < \epsilon < 1/2$, that $p_{\text{puz}} = 64x^{-1}n^{-1+\epsilon}$ and $p_{\text{ppl}} = n^{-\epsilon}$. The constant 64 has been chosen for expository purposes and has not been optimized.

Proof of Theorem 3.0.1 (b).

The sets S_i and R_i will have the same definition as before. We let E^i_{people} and E^i_{puzzle} denote the subsets of E_{people} and E_{puzzle} with one endpoint in S_i and one in R_i. The important thing to note is that E^i_{people} and E^i_{puzzle} are disjoint from $\cup_{j<i} E^j_{people}$ and $\cup_{j<i} E^j_{puzzle}$ respectively.

For large n, there exists an integer $i^* = \min\{i : |S_i| \geq n^\epsilon\}$ and a collection of values $\{\gamma_i = \frac{1}{2}x2^{2^{i-1}}\}_{1 \leq i \leq i^*}$ along with the events $T_i := \cap_{j=1}^{i}\{|S_j| \geq \gamma_j\}$. We will prove a series of lemmas that will combine to prove the theorem.

Lemma 3.2.1. $\mathbb{P}(T_1) = 1 - o(1)$.

Lemma 3.2.2. *For* $i < i^*$, $\mathbb{P}(T_{i+1} | T_i) \geq 1 - \exp\left(-\frac{1}{2}\gamma_i\right)$.

Lemma 3.2.3. $\mathbb{P}(T_{i^*}) \geq \prod_{i=1}^{i^*}\left(1 - \exp\left(-\frac{1}{2}\gamma_i\right)\right)$.

Lemma 3.2.4. $\mathbb{P}\left(\textbf{\textit{Solve}} \middle| T_{i^*}\right) = 1 - o(1)$

First we will show that T_1 occurs with high probability.

Proof of Lemma 3.2.1.

Let t_k denote the number of disjoint trees of size k from a graph with $p = p_{\text{eff}}$ and $|V_1| = n/2$. The expected value of t_k is given by

$$\lambda = \lambda_k = \mathbb{E}(t_k) = \binom{n/2}{k} k^{k-2} {p_{\text{eff}}}^{k-1} (1 - p_{\text{eff}})^{k(n/2-k)+\binom{k}{2}-k+1}. \tag{3.6}$$

From Theorem 5.1 in [29], the total variation distance between the distribution of t_k and a Poisson distribution with parameter λ has the following upper bound:

$$d_{TV}(\mathcal{L}(t_k), P_\lambda) \leq c_1(n/2, k, p_{\text{eff}}) + c_2(n/2, k, p_{\text{eff}})$$

where

$$c_1(n/2, k, p_{\text{eff}}) = \frac{2k^2}{n}\lambda$$

and

$$c_2(n/2, k, p_{\text{eff}}) \leq e^{-2k^2/n} k^2 p_{\text{eff}} (1 - p_{\text{eff}})^{-k^2} \lambda.$$

When $p_{\text{eff}} = 64x^{-1}n^{-1}$, the distribution of t_x will approach a Poisson distribution. Hence $\mathbb{P}(t_x \geq 1) \geq 1 - e^{-\lambda} - o(1)$. In this case $\lambda_x \to \infty$ so $\mathbb{P}(t_x \geq 1) = 1 - o(1)$. Therefore $\mathbb{P}(T_1) = 1 - o(1)$. □

Proof of Lemma 3.2.2.

We determined S_1 only using edges with both endpoints in V_1. For $1 \leq i < i^*$, and each $v \in R_i$ we have , the vertices of S_i will come from R_i and will be determined by edges from E^i_{puzzle} and E^i_{people}. Let $u, u' \in S_i$ and $v \in R_i$. We have

$$\mathbb{P}\left((u, v) \in E_{\text{puzzle}} \middle| T_i\right) = p_{\text{puz}}, \tag{3.7}$$

and similarly,

$$\mathbb{P}\left((u', v) \in E_{\text{people}} \middle| T_i\right) = p_{\text{ppl}}. \tag{3.8}$$

The probability that v connects to S_i can be expressed by

$$\mathbb{P}(v \in S_{i+1} \big| T_i) \geq \left(1 - (1 - p_{\text{puz}})^{|S_i|}\right)\left(1 - (1 - p_{\text{ppl}})^{|S_i|}\right). \tag{3.9}$$

This probability is independent for all vertices of R_i and is bounded below by

$$\frac{1}{4} p_{\text{eff}} |S_i|^2 \geq 8n^{-1}x^{-1}\gamma_2^2 = 2n^{-1}x2^{2^i} = 4n^{-1}\gamma_{i+1}. \tag{3.10}$$

The size of S_{i+1} conditioned on T_i stochastically dominates a binomial random variable, X_{i+1} with distribution $\text{Bin}(n - o(n), 4n^{-1}\gamma_{i+1})$. Therefore if T_i occurs, the expected value of $|S_{i+1}|$ is at least the expected value $\mathbb{E}X_{i+1} \geq 2\gamma_{i+1}$. Chernoff bounds then give

$$\mathbb{P}\left(T_{i+1} \big| T_i\right) \geq \mathbb{P}(X_{i+1} \geq \gamma_{i+1}) \geq 1 - \exp\left(-\frac{1}{2}\gamma_{i+1}\right) \tag{3.11}$$

This bound will hold for all i that allows for $(1-(1-p_{\text{ppl}})^{|S_i|})$ to be bounded by $|S_i|p_{\text{ppl}}/2$. Since $p_{\text{ppl}} = n^{-\epsilon}$, this bound holds when $|S_i| << n^{\epsilon}$.

□

Proof of Lemma 3.2.3.

This follows from the convergence of $\prod_j \left(1 - \exp\left(-\frac{1}{2}\gamma_i\right)\right)$. Using conditional probabilities we have

$$\mathbb{P}(T_{i^*}) \geq \mathbb{P}(T_{i^*} \big| T_{i^*-1})\mathbb{P}(T_{i^*-1}) = \left(1 - \exp\left(-\frac{1}{2}\gamma_{i^*}\right)\right)\mathbb{P}(T_{i^*-1}). \tag{3.12}$$

Iteratively apply the previous two lemmas to obtain

$$\mathbb{P}(T_{i^*}) \geq \prod_{j=1}^{i^*}\left(1 - \exp\left(-\frac{1}{2}\gamma_j\right)\right). \tag{3.13}$$

Taking the logarithm of both sides we have

$$\log \mathbb{P}(T_{i^*}) \geq \sum_{i=1}^{i^*} \log\left(1 - \exp\left(-\frac{1}{2}\gamma_i\right)\right) \geq \sum_{i=1}^{\infty} \log\left(1 - \exp\left(-\frac{1}{2}\gamma_i\right)\right). \tag{3.14}$$

We can use the approximation $\log(1-q) \geq -2q$ for q near 0. This gives

$$0 \geq \log \mathbb{P}(T_{i^*}) \geq \sum_i -\exp\left(-\frac{1}{2}x2^{2^{i-1}}\right) \geq -e^{-x}. \tag{3.15}$$

Since x is large when n is large this expression tends to zero. Hence

$$\mathbb{P}(T_{i^*}) \geq e^{-e^{-x}} \geq 1 - e^{-x} = 1 - o(1). \tag{3.16}$$

□

Proof of Lemma 3.2.4.

If T_{i^*} is true then not only do we have $|S_{i^*}| \geq n^\epsilon$, but also that $|R_{i^*}| \geq n/3$. Let $|S_{i^*}| = \alpha n^\epsilon$ where α is bounded below by a constant. For $v \in R_{i^*}$ we have similar to 3.9

$$\begin{aligned}\mathbb{P}(v \in S_{i*+1}) &\geq \left(1 - \left(1 - n^{-\epsilon}\right)^{|S_{i^*}|}\right)\left(1 - \left(1 - 64x^{-1}n^{-1+\epsilon}\right)^{|S_{i^*}|}\right). \\ &\geq \left(1 - e^{-\alpha}\right)\left(1 - \left(1 - 64x^{-1}n^{-1+\epsilon}\right)^{\alpha n^\epsilon}\right).\end{aligned}$$

There exists positive constant $\delta < 1$ such that

$$1 - \left(1 - 64x^{-1}n^{-1+\epsilon}\right)^{\alpha n^\epsilon} \geq n^{\epsilon+\delta-1}.$$

In this case the expected size of S_{i^*+1} is

$$\mathbb{E}|S_{i^*+1}| \geq cn^{\epsilon+\delta} \tag{3.17}$$

where c is some small positive constant. Then using the Chernoff bound again we have

$$\mathbb{P}(|S_{i^*+1} \geq cn^{\epsilon+\delta}/2) = 1 - \exp^{-cn^{\epsilon+\delta}/4}. \tag{3.18}$$

We modify the definition of R_{i^*+1} to include all remaining vertices and not just those in R_1,

$$R_{i^*+1} := V \backslash \bigcup_{j=1}^{i^*+1} S_j.$$

The edges of $E_{people}^{i^*+1}$ and $E_{puzzle}^{i^*+1}$ are still independent of all revealed edges so far. There exists $\delta' > 0$ such that for $v \in R_{i^*+1}$ the probability that there exists some $u \in S_{i^*+1}$ such that $(u, v) \in E_{people}^{i^*+1}$ is given by

$$\mathbb{P}(\ \exists\ u \in S_{i^*+1}\ s.t.\ (u,v) \in E_{people}^{i^*+1}) \geq 1 - \left(1 - n^{-\epsilon}\right)^{cn^{\epsilon+\delta}/2} \geq 1 - \exp(-n^{\delta'}). \tag{3.19}$$

Therefore every $v \in R_{i^*+1}$ is connected to S_{i^*+1} with a people edge with probability

$$\left(1 - \exp(-n^{\delta'})\right)^n \geq 1 - n\exp(-n^{\delta}) = 1 - o(1). \tag{3.20}$$

In the limit, everything connects to S_{i^*+1} with an edge in E_{people}. Since(V, E_{puzzle}) is connected, eventually everything will merge together.

□

The proof of Theorem 3.0.1 (b) follows because

$$\mathbb{P}(\mathbf{Solve}) \geq \mathbb{P}(\mathbf{Solve}\big|T_{i^*})\mathbb{P}(T_{i^*}) = (1 - o(1))(1 - o(1)) = 1 - o(1).$$

□

BIBLIOGRAPHY

[1] Robert Tarjan. Sorting using networks of queues and stacks. *J. ACM*, 19(2):341–346, April 1972.

[2] Sergi Elizalde. Fixed points and excedances in restricted permutations. *arXiv preprint math/0212221*, 2002.

[3] Sergi Elizalde. Multiple pattern avoidance with respect to fixed points and excedances. *Electron. J. Combin*, 11(1):40, 2004.

[4] Sergi Elizalde. *Statistics on pattern-avoiding permutations.* PhD thesis, Massachusetts Institute of Technology, 2004.

[5] Sergi Elizalde and Igor Pak. Bijections for refined restricted permutations. *Journal of Combinatorial Theory, Series A*, 105(2):207–219, 2004.

[6] W. Feller. *An Introduction to Probability Theory and its Applications, 2nd ed.* John Wiley, New York, 1957.

[7] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics.* Cambridge University Press, Cambridge, 2009.

[8] Ira M Gessel. Symmetric functions and p-recursiveness. *Journal of Combinatorial Theory, Series A*, 53(2):257–285, 1990.

[9] Christopher Hoffman, Douglas Rizzolo, and Erik Slivken. Pattern avoiding permutations and brownian excursion. *preprint*, 2014.

[10] W. D. Kaigh. An invariance principle for random walk conditioned by a late return to zero. *Ann. Probability*, 4(1):115–121, 1976.

[11] Sergey Kitaev. *Patterns in permutations and words.* Springer, 2011.

[12] Donald E. Knuth. *The art of computer programming. Vol. 1: Fundamental algorithms.* Second printing. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont, 1969.

[13] Percy A MacMahon. Combinatory analysis, 1915.

[14] Neal Madras and Lerna Pehlivan. Structure of random 312-avoiding permutations. *preprint*, 2013.

[15] Adam Marcus and Gábor Tardos. Excluded permutation matrices and the stanley–wilf conjecture. *Journal of Combinatorial Theory, Series A*, 107(1):153–160, 2004.

[16] Sam Miner and Igor Pak. The shape of random pattern avoiding permutations. *arXiv preprint arXiv:1303.7313*, 2013.

[17] Peter Mörters and Yuval Peres. *Brownian motion*, volume 30. Cambridge University Press, 2010.

[18] J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard.

[19] Vaughan R. Pratt. Computing permutations with double-ended queues, parallel stacks and parallel queues. In *Proceedings of the Fifth Annual ACM Symposium on Theory of Computing*, STOC '73, pages 268–277, New York, NY, USA, 1973. ACM.

[20] J Chalupa, P L Leath, and G R Reich. Bootstrap percolation on a Bethe lattice. *J. Phys. C*, 12(1):L31–L35, 1979.

[21] Roberto H Schonmann. On the behavior of some cellular automata related to bootstrap percolation. *Annals of Probability*, 20(1):174–193, January 1992.

[22] C. D. van Enter. Proof of Straley's argument for bootstrap percolation. *J. Statist. Phys.*, 48(3-4):943–945, 1987.

[23] Joan Adler and Uri Levi. Bootstrap percolation: visualizations and applications. *Brazilian Journal of Physics*, 33(3), 2003.

[24] M. Aizenman and J. L. Lebowitz. Metastability effects in bootstrap percolation. *J. Phys. A*, 21(19):3801–3813, 1988.

[25] Alexander E Holroyd. Sharp metastability threshold for two-dimensional bootstrap percolation. *Probability Theory Related Fields*, 125:195–224, 2003.

[26] Alexander E. Holroyd, Thomas M. Liggett, and Dan Romik. Integrals, partitions, and cellular automata. *Trans. Amer. Math. Soc.*, 356(8):3349–3368, 2004.

[27] József Balogh, Béla Bollobás, and Robert Morris. Bootstrap percolation in high dimensions. *Journal of Combinatorics, Probability and Computing*, 19(5-6):643–692, 2010.

[28] József Balogh and Boris Pittel. Bootstrap percolation on the random regular graph. *Random Structures and Algorithms*, 30(1-2):257–286, 2007.

[29] Béla Bollobás. *Random Graphs 2nd Ed.* Cambridge University Press, 2001.

[30] Nathan Ross. Fundamentals of Stein's method. *Probability Surveys*, 8:210–293, September 2001.

[31] Janko Gravner, Chris Hoffman, James Pfieffer, and David Sivakoff. Bootstrap percolation on the hamming torus. preprint, 2012.

[32] George Grimmett. *Percolation*. Springer, 1999.

[33] Béla Bollobás and Oliver Riordan. The critical probability for random Voronoi percolation in the plane is 1/2. *Probability Theory Related Fields*, 136:417–468, 2006.

[34] Charles D. Brummitt, Shirshendu Chatterjee, Partha S. Dey, and David Sivakoff. Jigsaw percolation: What social networks can collaboratively solve a puzzle. *arxiv*, 2012.

[35] J. Gravner and D. Sivakoff. Nucleation scaling in jigsaw percolation. *ArXiv e-prints*, October 2013.