Thresholds for Patterns in Random Compositions and Random Permutations

PhD Thesis

Dan Threlfall

Department of Mathematics and Statistics
University of Strathclyde, Glasgow

October 9, 2024

Chapters 2, 3 and 5 are based on [12] and are the result of the author's work in collaboration with David Bevan. Chapters 4 and 6 are based on [13] and are the result of the author's work in collaboration with David Bevan. The author contributed to all of the content within [12] with the exception of runs of equal terms, square patterns, upper and lower consecutive patterns, consecutive patterns specifying relative ordering (except for Carlitz compositions) and nonconsecutive patterns. The author contributed to all of the content within [13] with the exception of Proposition 4, the open questions and inversion sequence inequalities within the appendix.

Signed: Dan Threlfall

Date: 09/10/2024

Abstract

We explore how the asymptotic structure of a random permutation of $[n]$ with m inversions evolves, as m increases, establishing thresholds for the appearance and disappearance of any classical, consecutive or vincular pattern. Our investigation begins with exploring how the asymptotic structure of a random n-term weak integer composition of m evolves, as m increases from zero. The primary focus of our investigation into compositions is establishing thresholds for the appearance and disappearance of substructures, particularly the appearance and disappearance of consecutive composition patterns. We are then able to transfer the established composition threshold to establish the thresholds for classical, consecutive or vincular permutation patterns occurring within our random permutation model.

This thesis is based on the papers [12] and [13].

Contents

Acknowledgements

I am grateful to my supervisor, David Bevan, for his incredible support and guidance over the past four years of my PhD studies.

I would also like to thank the community at Strathclyde. Throughout my studies at this institution I have made some great friends and memories which I will forever cherish.

I am also thankful to my friends who I grew up with, their constant support and interest in my work has always been gratefully embraced.

Furthermore, I would like to thank the Permutation Patterns research community as well as the Scottish combinatorics community for their friendship, advice and interesting research over the past few years.

I would also like to thank the Engineering and Physical Sciences Research Committee (EPSRC) for the financial support of my PhD studies.

Finally, I would like to thank my family for their unwavering love and support they have given me throughout my life.

Notation

Throughout this thesis, numerous different symbols are used as such, we list the following notation for the benefit of the reader:

Notation	Definition
p	Probability
q	$1-p$ (The complement of p).
$\mathbb{P}[A]$	The probability of the event, A.
$\mathbb{E}[X]$	The expected value of the random variable, X.
$\mathrm{Var}[X]$	The variance of the random variable, X.
$\mathsf{Cov}[X,Y]$	The covariance between the random variables, X and Y.
$[n]$	The set $\{1,2,\ldots,n\}$.
C	An integer composition.
$C(i)$	The ith term of the integer composition, C.
$\|C\|$	The size of the random integer composition,C.
C^{+j}	The integer composition, C which has had 1 added to its jth term.
$\mathcal{C}_n$	The set of all n-compositions.
$\mathcal{C}_{n,m}$	The set of all n-compositions of m.
$\mathbf{C}_{n,m}$	The uniform random composition.
$\mathbf{C}_{n,p}$	The geometric random composition.
$\mathbf{C}_t$	The evolutionary random composition.
$\mathcal{Q}$	A property.
$\emptyset$	The empty set.
Δ	$\sum_{i\sim j}\mathbb{P}[A_i\wedge A_j]$ (The sum over the dependent pair of indices).
Λ	$\sum_{i\in I_n}\mathbb{P}[A_i]^2+\sum_{i\sim j}\mathbb{P}[A_i]\mathbb{P}[A_j]$.

Notation	Definition
$\mathsf{comp}_{\max}(C)$	The length of the longest component in C.
$\mathsf{gap}_{\max}(C)$	The length of the longest gap in C.
$\mathsf{comp}_{\min}(C)$	The length of the shortest component in C.
$\mathsf{gap}_{\min}(C)$	The length of the shortest gap in C.
$\max(C)$	The largest term of C.
$\log n$	The natural logarithm of n (unless stated otherwise.)
π	A pattern.
$\|\pi\|$	The size of the pattern, π.
${}^{=}\overline{\pi}$	An exact consecutive pattern.
σ	A permutation
$\sigma(i)$	The ith term of the permutation, σ.
$\overline{\sigma}$	The complement of the permutation, σ.
$\mathcal{S}_n$	The set of all n-permutations.
$\mathcal{S}_{n,m}$	The set of all n-permutations with m inversions.
$\boldsymbol{\sigma}_{n,m}$	The uniform random permutation.
e_σ	The inversion sequence of the permutation, σ.
$e_\sigma(i)$	The ith term of the inversion sequence of the permutation, σ.
$\mathcal{E}_{n,m}$	The set of all inversion sequences of n-permutations with m inversions.
e^{+j}	The inversion sequence, e which has had 1 added to its jth term.
$\mathbf{e}_{n,m}$	The uniform random inversion sequence.
$e[i,j]$	The terms of the inversion sequence, e from term i to term j.
$\sigma[i,j]$	The terms of the permutation, σ from term i to term j.
$\mathsf{inv}\,(\pi)$	The total number of inversions of the permutation pattern, π.
$\sigma \oplus \tau$	The direct sum of the permutations, σ and τ.

Furthermore, this thesis uses several different mathematical relations. If f and g are positive functions of n, then we use the following notation:

$$\begin{aligned}
f \lesssim g \quad & \text{if} \quad \limsup_{n\to\infty} f/g < \infty, \\
f \asymp g \quad & \text{if} \quad 0 < \liminf_{n\to\infty} f/g \quad \text{and} \quad \limsup_{n\to\infty} f/g < \infty, \\
f \sim g \quad & \text{if} \quad \lim_{n\to\infty} f/g = 1, \\
f \sim 0 \quad & \text{if} \quad \lim_{n\to\infty} f = 0, \\
f \ll g \ \text{ or } \ g \gg f \quad & \text{if} \quad \lim_{n\to\infty} f/g = 0.
\end{aligned}$$

In particular, $f \ll 1$ if $\lim_{n\to\infty} f = 0$, and $f \gg 1$ if $\lim_{n\to\infty} f = \infty$.
Note that we use the following equivalences within this thesis:

$$f \lesssim g \iff f = O(g), \qquad f \asymp g \iff f = \Theta(g), \qquad f \ll g \iff f = o(g).$$

Though $f \sim 0$ is nonstandard, it is used within this thesis to simplify the presentation of results.

Chapter 1

Introduction

We consider compositions and permutations from an evolutionary perspective, in an analogous manner to the Gilbert-Erdós-Renyi random graph [24, 25, 28]. Our two primary composition models are the *uniform random composition* $\mathbf{C}_{n,m}$, drawn uniformly from the family of n-term weak integer compositions of m, and the *geometric random composition* $\mathbf{C}_{n,p}$, an n-term weak integer composition in which each term is sampled independently from the geometric distribution where $0 < p < 1$ and with parameter $q = 1 - p$; that is, $\mathbb{P}\big[\mathbf{C}_{n,p}(i) = k\big] = qp^k$ for each $k \geq 0$ and $i \in [n]$. Our primary permutation model is the *uniform random permutation* $\boldsymbol{\sigma}_{n,m}$, drawn uniformly at random from the set of all n-permutations with exactly m inversions.

Our primary focus is to initially establish the threshold for a consecutive composition pattern to occur within $\mathbf{C}_{n,p}$ (generally the easiest of our models to work with) before transferring this threshold between our models to establish the threshold for a consecutive permutation pattern to occur with $\boldsymbol{\sigma}_{n,m}$. We conclude this thesis by establishing the thresholds for classical and vincular patterns to occur within $\boldsymbol{\sigma}_{n,m}$.

Within the second chapter we introduce our composition models, properties and thresholds as well as the methodology we utilise to establish many of our composition thresholds. The third chapter explores components (maximal runs of nonzero terms) and gaps (maximal runs of zero terms) of weak integer compositions. We also establish thresholds for different lengths of components and gaps occurring within our composition models. In our fourth chapter we briefly investigate the largest terms within our composition models. Within the fifth chapter we establish thresholds for exact consecutive composition patterns occurring within $\mathbf{C}_{n,p}$ and $\mathbf{C}_{n,m}$. Our sixth and final chapter introduces our primary permutation model, explores the relationship between compositions and inversion sequences as well as the relationship between inversion sequences and permutations. Finally we establish the thresholds for consecutive, classical and vincular patterns occurring within $\boldsymbol{\sigma}_{n,m}$.

Chapter 2

Random Compositions

An *n-term weak composition of m*, or just an *n-composition of m*, is a sequence of n nonnegative integers $(c_1, \ldots, c_n)$ such that $\sum_{i=1}^{n} c_i = m$. Compositions can be considered to be words over the nonnegative integers, and, if no term exceeds nine, we sometimes write specific compositions simply as a sequence of digits. See Figure 2.1 for an example. Alternatively, we can consider such a composition to consist of a sequence of n *boxes*, such that term c_i is the number of *balls* in box $i \in [n] := \{1, 2, \ldots, n\}$.

A "stars and bars" argument is a graphical aid for solving certain counting problems. In particular, we can calculate the number of ways of placing m balls in n boxes. By utilising this argument, the number of distinct n-compositions of m can be derived. See Figure 2.2 for an example of a representation of a 13-term composition of 27. In a "stars and bars" diagram, the "stars" represent balls and the "bars" represent the end of one box and the start of the next.

If we consider a "stars and bars" diagram representing placing m balls in n boxes, there are m "stars" and $n-1$ "bars". We can then consider having $m+n-1$ empty spaces and choosing where to place either the m "stars" or $n-1$ "bars". Using this argument it can be seen that the number of n-compositions of m is $\binom{m+n-1}{n-1} = \binom{m+n-1}{m}$.

Definition 2.0.1. If C is an integer composition, then we use $C(i)$ to denote its ith term, and $|C|$ to denote its *size*, the sum of its terms. Let $\mathcal{C}_n$ denote the set of all n-compositions, and let $\mathcal{C}_{n,m}$ be the set of all n-compositions of m.

We now present three models of random integer compositions.

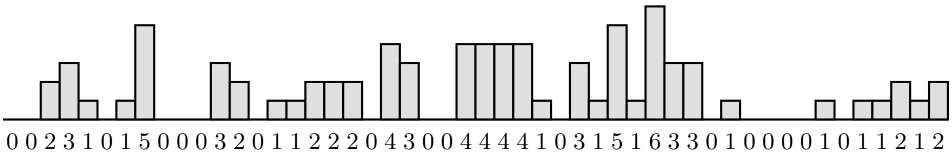

Figure 2.1: Bar-chart representation of a 50-term composition of 80

$$* * \mid * \mid\mid * * * * \mid * \mid * \mid * * * * * * * \mid * * * \mid * * \mid * * \mid\mid * \mid * * *$$

Figure 2.2: “Stars and bars” representation of the composition 2,1,0,4,1,1,7,3,2,2,0,1,3

2.1 The Uniform Random Composition $\mathbf{C}_{n,m}$

The *uniform random composition* $\mathbf{C}_{n,m}$ is drawn uniformly from $\mathcal{C}_{n,m}$. Thus, for every composition $C \in \mathcal{C}_{n,m}$,

$$\mathbb{P}\big[\mathbf{C}_{n,m} = C\big] \;=\; \binom{m+n-1}{m}^{-1},$$

each of the distinct n-compositions of m being equally likely. For example, the probability that $\mathcal{C}_{50,80}$ is the composition in Figure 2.1 is $\binom{129}{80}^{-1}$.

2.2 The Evolutionary Random Composition $\mathbf{C}_t$

An alternative, evolutionary, perspective comes from taking a dynamic view. We can consider a process on compositions, namely an infinite sequence of compositions, $0^n, C_1, C_2, C_3, \dots,$ where 0^n denotes the *empty* n-composition $(0, \dots, 0)$, and C_{t+1} is obtained from C_t by the addition of 1 to a single term. Note that there is no maximal n-composition (unlike the situation with random graphs).

Definition 2.2.1. The *evolutionary random composition* $(\mathbf{C}_t)_{t\geq 0}$ is the Markov chain satisfying $\mathbf{C}_0 = 0^n$ and, for each $t \geq 0$ and $j \in [n]$,

$$\mathbb{P}\big[\mathbf{C}_{t+1} = \mathbf{C}_t^{+j}\big] \;=\; \frac{\mathbf{C}_t(j)+1}{n+t},$$

where C^{+j} denotes the composition obtained from C by the addition of 1 to its jth term.

The evolutionary random composition $\mathbf{C}_t$ is uniformly distributed over n-compositions of t or alternatively, $\mathbf{C}_{t+1}$ is obtained from $\mathbf{C}_t$ by adding a star at a uniformly random place in the stars and bars diagram:

Proposition 2.2.2. *For each $t \geq 0$, the random composition $\mathbf{C}_t$ is uniformly distributed over $\mathcal{C}_{n,t}$.*

Proof. We use induction on t. Trivially, $\mathbf{C}_0$ is uniformly distributed over $\mathcal{C}_{n,0}$. Suppose $\mathbf{C}_t$ is uniformly distributed over $\mathcal{C}_{n,t}$, and that $C \in \mathcal{C}_{n,t+1}$. Let C^{-j} denote the composition obtained from C by the subtraction of 1 from its jth term (if this is possible). Then,

$$\begin{aligned}\mathbb{P}\big[\mathbf{C}_{t+1} = C\big] \;&=\; \sum_{j\in[n],\, C(j)\neq 0} \mathbb{P}\big[\mathbf{C}_t = C^{-j}\big]\frac{C(j)}{n+t} \\ &=\; \binom{n+t-1}{t}^{-1} \sum_{j\in[n]} \frac{C(j)}{n+t} \;=\; \frac{t!\,(n-1)!}{(n+t-1)!}\,\frac{t+1}{n+t} \;=\; \binom{n+t}{t+1}^{-1}. \;\square\end{aligned}$$

2.3 The Geometric Random Composition $\mathbf{C}_{n,p}$

If $p \in [0,1)$, then the *geometric random composition* $\mathbf{C}_{n,p}$ is distributed over $\mathcal{C}_n$ so that for each $C \in \mathcal{C}_n$, we have $\mathbb{P}\big[\mathbf{C}_{n,p} = C\big] = q^n p^{|C|}$, where $q = 1 - p$. Each term of $\mathbf{C}_{n,p}$ is sampled independently from the geometric distribution with parameter q; that is, $\mathbb{P}\big[C(i) = k\big] = qp^k$ for each $k \geq 0$ and $i \in [n]$. Note that $\mathbf{C}_{n,p}$ is not defined for $p = 1$. For example, the probability that $\mathbf{C}_{50,p}$ is the composition in Figure 2.1 equals $q^{50}p^{80}$.

Definition 2.3.1. To avoid unnecessary repetition, when considering $\mathbf{C}_{n,p}$ in this thesis, q always denotes $1 - p$. Moreover, we also assume that the definition of any annotated p also defines a similarly annotated q, so we have $q_1 = 1 - p_1$ and $q^+ = 1 - p^+$, without stating so explicitly.

We collect here a few basic facts about $\mathbf{C}_{n,p}$. Each term has mean p/q and variance p/q^2, and its size $|\mathbf{C}_{n,p}|$ satisfies a negative binomial distribution,

$$\mathbb{P}\big[|\mathbf{C}_{n,p}| = m\big] \;=\; \binom{m+n-1}{m} p^m q^n,$$

with mean $\mu_{n,p} = np/q$. Note that if $p \ll 1$ then $\mu_{n,p} = np/(1-p) \sim np/1 = np$, and if $q \ll 1$ then $\mu_{n,p} = n(1-q)/q \sim n(1)/q = n/q$ where $p = p(n)$ and $m = m(n)$ are functions of n.

We now show (Proposition 2.3.3) that $|\mathbf{C}_{n,p}|$ is *concentrated* in the sense that, for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\big[1 - \epsilon \leq \frac{|\mathbf{C}_{n,p}|}{\mu_{n,p}} \leq 1 + \epsilon\big] = 1.$$

Proposition 2.3.2 ([26, Proposition $III.3$])**.** *Consider a family of random variables* $\mathbf{X}_n$*. Assume that the means* $\mu_n = \mathbb{E}\big[\mathbf{X}_n\big]$ *and the standard deviations* $\sigma_n = \sigma(\mathbf{X}_n)$ *satisfy the condition*

$$\lim_{n\to\infty} \frac{\sigma_n}{\mu_n} = 0,$$

then the distribution of $\mathbf{X}_n$ *is concentrated.*

It can now be proved that $|\mathbf{C}_{n,p}|$ is concentrated.

Proposition 2.3.3. $|\mathbf{C}_{n,p}|$ *is concentrated when* $p \gg n^{-1}$.

Proof. The geometric random composition $\mathbf{C}_{n,p}$ has size $|\mathbf{C}_{n,p}|$, with mean $\mu_{n,p} = np/q$ and variance $\sigma^2_{n,p} = np/q^2$. If $p \gg n^{-1}$, then

$$\lim_{n\to\infty} \frac{\sigma_{n,p}}{\mu_{n,p}} = \lim_{n\to\infty} \frac{\sqrt{np}/q}{np/q} = \lim_{n\to\infty} \frac{1}{\sqrt{np}} = 0.$$

By Proposition 2.3.2, $|\mathbf{C}_{n,p}|$ is concentrated. □

The size $|\mathbf{C}_{n,p}|$ has variance np/q^2, and so exhibits a concentrated distribution as long as $p \gg n^{-1}$. We make a brief further exploration on the upper bound on the probability of deviation of $|\mathbf{C}_{n,p}|$ by utilising Chebyshev's inequality.

Lemma 2.3.4 (Chebyshev's inequality)**.** *If* $\mathbf{X}$ *is a random variable with finite mean and variance, then, for* $k > 0$,

$$\mathbb{P}\big[|\mathbf{X} - \mu| \geq k\sigma\big] \leq \frac{1}{k^2}.$$

By Chebyshev's inequality, we have the following:

Lemma 2.3.5. $\mathbb{P}\big[\big||\mathbf{C}_{n,p}| - np/q\big| \geq \alpha\sqrt{np}/q\big] \leq \alpha^{-2}$.

Geometric random compositions of size m are uniformly distributed over $\mathcal{C}_{n,m}$:

Proposition 2.3.6. *A random composition* $\mathbf{C}_{n,p}$ *whose terms sum to* m *is equally likely to be any one of the distinct* n*-compositions of* m*.*

Proof. Suppose $C \in \mathcal{C}_{n,m}$. Then,

$$\begin{aligned}\mathbb{P}\big[\mathbf{C}_{n,p} = C \;\big|\; |\mathbf{C}_{n,p}| = m\big] &= \frac{\mathbb{P}\big[\mathbf{C}_{n,p} = C \,\wedge\, |\mathbf{C}_{n,p}| = m\big]}{\mathbb{P}\big[|\mathbf{C}_{n,p}| = m\big]} \\ &= \frac{\mathbb{P}\big[\mathbf{C}_{n,p} = C\big]}{\mathbb{P}\big[|\mathbf{C}_{n,p}| = m\big]} \\ &= \frac{p^m q^n}{\binom{m+n-1}{m} p^m q^n} = \binom{m+n-1}{m}^{-1}. \qquad \square\end{aligned}$$

Thus, $\mathbf{C}_{n,p}$ conditioned on the event $|\mathbf{C}_{n,p}| = m$ is equal in distribution to $\mathbf{C}_{n,m}$. Note that this holds for any choice of p and m. As is the case with random graphs, the probabilistic model is more amenable to analysis, so we prefer to work with $\mathbf{C}_{n,p}$ and then transfer the results to $\mathbf{C}_{n,m}$ (see Propositions 2.4.5, 2.5.1 and 5.1.3 below).

2.4 Properties

Definition 2.4.1. We consider a *property* of n-compositions simply to be a subset of $\mathcal{C}_n$.

For example, the set of n-compositions with no zero terms is a property, as is the set of n-compositions with at least one term equal to three.

A property $\mathcal{Q}$ is *increasing* if $C \in \mathcal{Q}$ implies $C^{+j} \in \mathcal{Q}$ for every $j \in [n]$, or equivalently if $C \in \mathcal{Q}$ implies $C + C' \in \mathcal{Q}$, where $C + C'$ denotes the term-wise sum of two n-compositions. The complement of an increasing property is *decreasing*. A property that is either increasing or decreasing is *monotone*. For example, the n-compositions with no zero terms form an increasing property, adding to terms, once all terms are non-zero, will result in all subsequent n-compositions having no zero terms. On the other hand, the set of n-compositions with exactly one term equal to three is not monotone. For example, if we have the following 5-composition C given by $2, 0, 4, 1, 3$, exactly one term is equal to 3. By adding (enough) to terms $C(1), C(2)$ or $C(4)$ can result in the property of there being exactly one term equal to 3, no longer holding. Similarly, by taking 1 from either of the terms $C(3)$ or $C(5)$ will also result in the property no longer holding.

Both $\mathbf{C}_{n,m}$ and $\mathbf{C}_{n,p}$ behave monotonically with respect to monotone properties:

Proposition 2.4.2. *If $\mathcal{Q}$ is an increasing property and $m_1 < m_2$, then*

$$\mathbb{P}\big[\mathbf{C}_{n,m_1} \in \mathcal{Q}\big] \leq \mathbb{P}\big[\mathbf{C}_{n,m_2} \in \mathcal{Q}\big].$$

Proof. $\mathbb{P}\big[\mathbf{C}_{t+1} \in \mathcal{Q}\big] \geq \mathbb{P}\big[\mathbf{C}_{t+1} \in \mathcal{Q} \wedge \mathbf{C}_t \in \mathcal{Q}\big] = \mathbb{P}\big[\mathbf{C}_t \in \mathcal{Q}\big]$, since $\mathcal{Q}$ is increasing. The proposition now follows by Proposition 2.2.2. □

Proposition 2.4.3. *If $\mathcal{Q}$ is an increasing property and $p_1 < p_2$, then*

$$\mathbb{P}\big[\mathbf{C}_{n,p_1} \in \mathcal{Q}\big] \leq \mathbb{P}\big[\mathbf{C}_{n,p_2} \in \mathcal{Q}\big].$$

Proof. Let $\mathbf{C}_{n,p_1,p_2}$ denote a random n-composition, each of whose terms is sampled independently from the following distribution. For each $i \in [n]$,

$$\mathbb{P}\big[\mathbf{C}_{n,p_1,p_2}(i) = k\big] \;=\; \begin{cases} \frac{q_2}{q_1} & \text{if } k = 0, \\ \\ \frac{q_2}{q_1}\big(1 - \frac{p_1}{p_2}\big){p_2}^k & \text{if } k \geq 1. \end{cases}$$

We claim that $\mathbf{C}_{n,p_1} + \mathbf{C}_{n,p_1,p_2}$ has the same distribution as $\mathbf{C}_{n,p_2}$ if $\mathbf{C}_{n,p_1}$ and $\mathbf{C}_{n,p_1,p_2}$ are chosen independently, thus providing a way of building $\mathbf{C}_{n,p_2}$ from 0^n in two steps via $\mathbf{C}_{n,p_1}$.

To prove this equality of distribution, we use probability generating functions. Let

$$f_p(x) = \sum_{k\geq 0} \mathbb{P}\big[\mathbf{C}_{n,p}(i) = k\big] x^k = \frac{q}{1-px},$$
$$f_{p_1,p_2}(x) = \sum_{k\geq 0} \mathbb{P}\big[\mathbf{C}_{n,p_1,p_2}(i) = k\big] x^k = \frac{q_2(1-p_1x)}{q_1(1-p_2x)}.$$

Thus $f_{p_2}(x) = f_{p_1}(x) f_{p_1,p_2}(x)$, and equality of distribution then follows from the independence of each term in the random compositions.

Hence, by coupling $\mathbf{C}_{n,p_1}$ and $\mathbf{C}_{n,p_2}$,

$$\begin{aligned}\mathbb{P}\big[\mathbf{C}_{n,p_2} \in \mathcal{Q}\big] &= \mathbb{P}\big[\mathbf{C}_{n,p_1} + \mathbf{C}_{n,p_1,p_2} \in \mathcal{Q}\big] \\ &\geq \mathbb{P}\big[\mathbf{C}_{n,p_1} + \mathbf{C}_{n,p_1,p_2} \in \mathcal{Q} \,\wedge\, \mathbf{C}_{n,p_1} \in \mathcal{Q}\big] \quad \text{(where the } \mathbf{C}_{n,p_1} \text{ are the same)} \\ &= \mathbb{P}\big[\mathbf{C}_{n,p_1} \in \mathcal{Q}\big],\end{aligned}$$

since $\mathcal{Q}$ is increasing. □

Typically, we are interested in whether a property holds, or fails to hold, in the asymptotic limit.

Definition 2.4.4. We say that $\mathcal{Q}$ holds *asymptotically almost surely* (*a.a.s.*) or, synonymously, *with high probability* (*w.h.p.*) in $\mathbf{C}_{n,p}$ if $\mathbb{P}\big[\mathbf{C}_{n,p} \in \mathcal{Q}\big] \sim 1$, and analogously for $\mathbf{C}_{n,m}$. If a property holds a.a.s. then its complement asymptotically *almost never* holds.

The following trivial example can be considered. Let $p \in (0,1)$ be a constant and $\mathcal{Q}_1$ be the set of all compositions with at least one term greater than 0. Clearly, $\mathbb{P}\big[\mathbf{C}_{n,p} \in \mathcal{Q}_1\big] \sim 1$. On the other hand (for the same probability), let $\mathcal{Q}_2$ be the set of all n-compositions where all terms are equal to 0, then $\mathbb{P}\big[\mathbf{C}_{n,p} \in \mathcal{Q}_2\big] \sim 0$.

Since $|\mathbf{C}_{n,p}|$ is concentrated around its mean, it is reasonable to expect that, if n is large, then $\mathbf{C}_{n,p}$ and $\mathbf{C}_{n,m}$ should behave in a similar fashion when $m \sim np/q$ (the mean of $|\mathbf{C}_{n,p}|$), or equivalently, when $p \sim m/(m+n)$. This is indeed the case, and the following proposition enables us to transfer results from $\mathbf{C}_{n,p}$ to $\mathbf{C}_{n,m}$, the probability that an increasing property holds being the same in both models.

Proposition 2.4.5. *Let $\mathcal{Q}$ be an increasing property and $\alpha \in [0,1]$ be a constant. Suppose $p_0 = p_0(n)$ and $\delta = \delta(n) \gg \sqrt{p_0}/(q_0\sqrt{n})$ are such that $\mathbb{P}\big[\mathbf{C}_{n,p} \in \mathcal{Q}\big] \sim \alpha$ for all p for which p/q differs from p_0/q_0 by no more than δ. Then $\mathbb{P}\big[\mathbf{C}_{n,m_0} \in \mathcal{Q}\big] \sim \alpha$, where $m_0 = [np_0/q_0]$.*

Proof. Let p^- satisfy $p^-/q^- = p_0/q_0 - \delta$, and p^+ satisfy $p^+/q^+ = p_0/q_0 + \delta$.

Fix any $\epsilon > 0$ and suppose n is sufficiently large such that both $\mathbb{P}\big[|\mathbf{C}_{n,p^-}| > m_0\big] \le \epsilon$ and $\mathbb{P}\big[|\mathbf{C}_{n,p^+}| < m_0\big] \le \epsilon$. Note that $|m_0/n - p_0/q_0| < \delta$. This is possible by Observation 2.3.5 given that

$$\begin{aligned} m_0 - \frac{np^-}{q^-} &= n\delta \gg \frac{\sqrt{np_0}}{q_0} \sim \frac{\sqrt{np^-}}{q^-}, \quad \text{and} \\ \frac{np^+}{q^+} - m_0 &= n\delta \gg \frac{\sqrt{np_0}}{q_0} \sim \frac{\sqrt{np^+}}{q^+}. \end{aligned}$$

Then

$$\begin{aligned} \mathbb{P}\big[\mathbf{C}_{n,p^-} \in \mathcal{Q}\big] &= \sum_{k \le m_0} \mathbb{P}\big[\mathbf{C}_{n,k} \in \mathcal{Q}\big]\,\mathbb{P}\big[|\mathbf{C}_{n,p^-}| = k\big] \;+\; \sum_{k > m_0} \mathbb{P}\big[\mathbf{C}_{n,k} \in \mathcal{Q}\big]\,\mathbb{P}\big[|\mathbf{C}_{n,p^-}| = k\big] \\ &\le \mathbb{P}\big[\mathbf{C}_{n,m_0} \in \mathcal{Q}\big]\,\mathbb{P}\big[|\mathbf{C}_{n,p^-}| \le m_0\big] \;+\; \mathbb{P}\big[|\mathbf{C}_{n,p^-}| > m_0\big] \\ &\le \mathbb{P}\big[\mathbf{C}_{n,m_0} \in \mathcal{Q}\big] \;+\; \epsilon. \qquad (1) \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbb{P}\big[\mathbf{C}_{n,p^+} \in \mathcal{Q}\big] &\ge \sum_{k \geqslant m_0} \mathbb{P}\big[\mathbf{C}_{n,k} \in \mathcal{Q}\big]\,\mathbb{P}\big[|\mathbf{C}_{n,p^+}| = k\big] \\ &\ge \mathbb{P}\big[\mathbf{C}_{n,m_0} \in \mathcal{Q}\big]\,\mathbb{P}\big[|\mathbf{C}_{n,p^+}| \ge m_0\big] \\ &\ge (1-\epsilon)\,\mathbb{P}\big[\mathbf{C}_{n,m_0} \in \mathcal{Q}\big]. \qquad (2) \end{aligned}$$

So (1) and (2) imply the proposition. □

Note that, in general, we cannot remove the requirement that the property be increasing. For example, if $\mathcal{Q}$ is the set of n-compositions with no zero terms whose size is not a power of 2, then $\mathbb{P}\big[\mathbf{C}_{n,m} \in \mathcal{Q}\big] = 0$ whenever m is a power of 2, whereas $\mathcal{Q}$ holds a.a.s. in $\mathbf{C}_{n,p}$ once p is sufficiently large that both conditions hold w.h.p. However, in some situations we can transfer results concerning non-monotone properties from $\mathbf{C}_{n,p}$ to $\mathbf{C}_{n,m}$. For example, Proposition 5.1.3 enables us to do this for exact consecutive patterns.

2.5 Thresholds

Large random combinatorial objects often show a phenomenon of the sudden appearance and disappearance of properties. This nature of behaviour has been documented (but not limited to) graph theory [27], compositions and permutations. The latter two of which are explored within this thesis. We say that a function $m^\star = m^\star(n)$ is a *threshold* for an increasing property $\mathcal{Q}$ in $\mathbf{C}_{n,m}$ if

$$\mathbb{P}\big[\mathbf{C}_{n,m} \in \mathcal{Q}\big] \sim \begin{cases} 0 & \text{if } \; m \ll m^\star, \\ 1 & \text{if } \; m \gg m^\star, \end{cases}$$

and that $p^\star = p^\star(n)$ or $q^\star = q^\star(n)$ is a threshold for $\mathcal{Q}$ in $\mathbf{C}_{n,p}$ if

$$\mathbb{P}\big[\mathbf{C}_{n,p} \in \mathcal{Q}\big] \sim \begin{cases} 0 & \text{if } \; p/q \ll p^\star/q^\star, \\ 1 & \text{if } \; p/q \gg p^\star/q^\star. \end{cases}$$

That is, a property asymptotically almost never holds below its threshold, but holds asymptotically almost surely above it. Though it is not known if every monotone property of compositions has a threshold, we establish a number of them within this thesis.

In many situations, it can be determined that the threshold is more abrupt. A function $m^\star$ is a *sharp* threshold for a property $\mathcal{Q}$ in $\mathbf{C}_{n,m}$, and $p^\star$ is a sharp threshold for $\mathcal{Q}$ in $\mathbf{C}_{n,p}$, if, for every $\varepsilon > 0$,

$$\mathbb{P}\big[\mathbf{C}_{n,m} \in \mathcal{Q}\big] \sim \begin{cases} 0 & \text{if } \ m \leq (1-\varepsilon)m^\star, \\ 1 & \text{if } \ m \geq (1+\varepsilon)m^\star, \end{cases}$$

$$\mathbb{P}\big[\mathbf{C}_{n,p} \in \mathcal{Q}\big] \sim \begin{cases} 0 & \text{if } \ p/q \leq (1-\varepsilon)p^\star/q^\star, \\ 1 & \text{if } \ p/q \geq (1+\varepsilon)p^\star/q^\star. \end{cases}$$

Clearly, thresholds are not unique. Indeed, if a threshold for a property $\mathcal{Q}$ is not sharp, then a constant multiple is also a threshold for $\mathcal{Q}$. Sharp thresholds are not unique either, although a constant multiple of a sharp threshold for a property is not a threshold for that property. On the other hand, if $f_1(n) \gg 1$ is a sharp threshold for a property $\mathcal{Q}$ in $\mathbf{C}_{n,m}$ and $f_0(n) \ll f_1(n)$ then $f_1(n) + f_0(n)$ is also a threshold for $\mathcal{Q}$ in $\mathbf{C}_{n,m}$.

A consequence of Proposition 2.4.5 is that a threshold in $\mathbf{C}_{n,p}$ can be transferred to one in $\mathbf{C}_{n,m}$:

Proposition 2.5.1. *Let $\mathcal{Q}$ be an increasing property that has a threshold $p^\star \geq n^{-1}$ in $\mathbf{C}_{n,p}$. Then $np^\star/q^\star$ is a threshold for $\mathcal{Q}$ in $\mathbf{C}_{n,m}$.*

Proof. Let $m^\star = np^\star/q^\star$. Suppose $m \gg m^\star$ and $p^+ = m/(m+n)$, so $p^+/q^+ \gg p^\star/q^\star$. Now, since $p^\star \gg n^{-1}$, we also have $p^+/q^+ \gg \sqrt{p^+}/(q^+\sqrt{n})$, so we can find $\delta \gg \sqrt{p^+}/(q^+\sqrt{n})$ such that $p^+/q^+ - \delta \gg p^\star/q^\star$. Since $\mathcal{Q}$ holds a.a.s. in $\mathbf{C}_{n,p}$ when $p/q \gg p^\star/q^\star$, by Proposition 2.4.5, $\mathcal{Q}$ also holds a.a.s. in $\mathbf{C}_{n,m}$.

Similarly, suppose now that $m \ll m^\star$ and $p^- = m/(m+n)$, so $p^-/q^- \ll p^\star/q^\star$. Since $p^\star \gg n^{-1}$, we also have $p^\star/q^\star \gg \sqrt{p^-}/(q^-\sqrt{n})$, so we can find $\delta \gg \sqrt{p^-}/(q^-\sqrt{n})$ such that $p^-/q^- + \delta \ll p^\star/q^\star$. Since $\mathcal{Q}$ asymptotically almost never holds in $\mathbf{C}_{n,p}$ when $p/q \ll p^\star/q^\star$, then by Proposition 2.4.5, $\mathcal{Q}$ also asymptotically almost never holds in $\mathbf{C}_{n,m}$. □

To establish the presence of thresholds, we use the First Moment Method and the Second Moment Method. The First Moment Method is an immediate corollary of

Markov's Inequality and gives a sufficient condition for a property to asymptotically almost never hold.

Proposition 2.5.2 (First Moment Method [27, Lemma 20.2])**.** *If* $(X_n)_{n=1}^{\infty}$ *is a sequence of nonnegative integer-valued random variables and* $\mathbb{E}\big[X_n\big] \ll 1$*, then* $\mathbb{P}\big[X_n = 0\big] \sim 1$.

The Second Moment Method, which follows from Chebyshev's Inequality, gives a sufficient condition for a property to hold a.a.s.

The following presentation follows [2, Corollary 4.3.4].

Definition 2.5.3. Given an indexed set of events $\{A_i : i \in I\}$, we write $i \sim j$ if $i \neq j$ and the events A_i and A_j are not independent. We say that A_i and A_j are *correlated.* If $i \sim j$, we say that i, j is a *dependent* pair of indices.

For example, if, for each $i \in [n-1]$, the event A_i occurs if the ith and $(i+1)$th terms of $\mathbf{C}_{n,p}$ are identical, then $i \sim j$ precisely when $|i - j| = 1$.

Proposition 2.5.4 (Second Moment Method)**.** *Suppose, for each* $n \geq 1$*, that* $\{A_i : i \in I_n\}$ *is a set of events. Suppose* $X = X_n$ *is the random variable that records how many of these events occur, and let* $\Delta = \sum_{i \sim j} \mathbb{P}\big[A_i \wedge A_j\big]$*, where the sum is over dependent pairs of indices. If* $\mathbb{E}\big[X\big] \gg 1$ *and* $\Delta \ll \mathbb{E}\big[X\big]^2$*, then* $\mathbb{P}\big[X > 0\big] \sim 1$.

It is possible to determine the probability of a property holding *at* its threshold. To do this we use the Chen–Stein Method [17]. The basic idea is that if events are mostly independent (for some properly defined notion of "mostly"), then the number of these events that occur tends to a Poisson distribution. As noted in [3], under suitable conditions, Poisson convergence can be established by computing only the first and second moments. In particular, this holds in the case of *dissociated* events [6, 7], which is sufficient for our purposes.

Definition 2.5.5. Let $\mathcal{N}$ be a collection of subsets, drawn from the natural numbers, and suppose that, for each $J \in \mathcal{N}$, the random variable X_J can take only the values 0 or 1. The random variables $\{X_J : J \in \mathcal{J}\}$ are said to be dissociated if collections of them which have no index in common are independent. That is, $\{X_J : J \in \mathcal{J}\}$ is independent of $\{X_K : K \in \mathcal{K}\}$ whenever $\mathcal{J} \cap \mathcal{K} = \emptyset$.

We adapt the following proposition to be utilised in further results.

Proposition 2.5.6 ([35, Theorem 4])**.** *Consider a random variable W that can be written as a sum $\sum_{\alpha\in\Gamma} I_\alpha$ of 0-1 random variables, where Γ is a finite index set. Let $\lambda = \mathbb{E}\big[W\big] = \sum_\alpha p_\alpha$, where $p_\alpha = \mathbb{E}\big[I_\alpha\big] = \mathbb{P}\big[I_\alpha = 1\big] = \mathbb{P}\big[\alpha\big]$. Suppose that for each α there is a subset $\Gamma_\alpha \subset \Gamma$ such that I_α is independent of $\{I_\beta : \beta \notin \Gamma_\alpha\}$. Then the total variation distance between the distribution (or "law") $\mathcal{L}(W)$ of W and the Poisson distribution $Po(\lambda)$, with mean μ, is bounded above as follows:*

$$d_{tv}\left(\mathcal{L}(W), Po(\lambda)\right) \leq max\left(1, \lambda^{-1}\right)\left(\sum_{\alpha\in\Gamma} p_\alpha^2 + \sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma_\alpha\setminus\{\alpha\}}\left(p_\alpha p_\beta + \mathbb{E}\big[I_\alpha I_\beta\big]\right)\right).$$

The following is adapted from the above proposition.

Proposition 2.5.7 (Chen–Stein Method)**.** *Suppose, for each $n \geq 1$, that $\{A_i : i \in I_n\}$ is a set of events, and that $|I_n| \gg 1$. Suppose X_n is the random variable that records how many of these events occur, and let*

$$\Delta \;=\; \sum_{i\sim j}\mathbb{P}\big[A_i \wedge A_j\big] \qquad \textit{and} \qquad \Lambda \;=\; \sum_{i\in I_n}\mathbb{P}\big[A_i\big]^2 \;+\; \sum_{i\sim j}\mathbb{P}\big[A_i\big]\mathbb{P}\big[A_j\big].$$

If there exists a constant $\lambda > 0$ such that $\mathbb{E}\big[X_n\big] \sim \lambda$, and $\Delta + \Lambda \ll 1$, then X_n converges in distribution to a Poisson distribution with mean λ. In particular, the asymptotic probability that none of the events occur is $e^{-\lambda}$.

In particular, the following adaptations were made to [35, Theorem 4] to give Proposition 2.5.7:

Notation used in [35, Theorem 4]	Notation used in Proposition 2.5.7
Γ	$\{A_i : i \in I_n\}$
Γ_α	$\{A_j : i \sim j\} \cup \{A_i\}$, if $A_i = \alpha$
W	X_n
$\sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma_\alpha\setminus\{\alpha\}}$	$\sum_{i\sim j}$
$\sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma_\alpha\setminus\{\alpha\}}\mathbb{E}\big[I_\alpha I_\beta\big]$	Δ

In Proposition 2.5.7, the equation has been separated into Δ and Λ, since Δ is used in the Second Moment Method.

Chapter 3

Components and Gaps

In the next two sections, we shall investigate the behaviour of how the random composition evolves as its size increases. We primarily establish our results for $\mathbf{C}_{n,p}$, before transferring to $\mathbf{C}_{n,m}$.

Perhaps intuitively, whenever $p \ll n^{-1}$ in $\mathbf{C}_{n,p}$, all terms are equal to 0 a.a.s. As the expectation of the number of nonzero terms is equal to $np \ll n\left(n^{-1}\right) = 1$, then the expectation is asymptotically equal to 0. So, by the First Moment Method, every term is equal to 0 a.a.s.

Our focus in this current section is on components and gaps.

Definition 3.0.1. A *component* of a weak integer composition is a maximal run of nonzero terms. A *gap* is a maximal run of zero terms.

For example, the composition in Figure 2.1 on page 5 has 10 components, the longest having length 7. It also has 10 gaps, the longest having length 4.

Components in $\mathbf{C}_{n,p}$ are equivalent to maximal runs of heads in sequences of coin tosses. Here, a head is obtained with probability p and a tail is obtained with probability $q = 1 - p$. This topic of study has previously been explored in great detail [23, 30, 32] (see also [26, pages 308–312]). Furthermore, components are equivalent to maximal runs of a singular repeating letter in a *word* over a binary *alphabet* $\{0, 1\}$. A word is a sequence of the elements of the alphabet.

Components and gaps are dual in $\mathbf{C}_{n,p}$. Any statement about components can be converted into one about gaps by reversing the roles of p and q and vice versa. For example, once $q \ll n^{-1}$ the expected number of terms in $\mathbf{C}_{n,p}$ equal to zero, is equal to $nq \ll 1$. Thus, a.a.s., every term is nonzero and so $\mathbf{C}_{n,p}$ consists of a single component with no gaps. Similarly, once $p \ll n^{-1}$, we see that $\mathbf{C}_{n,p}$ consists of a single gap with no components as every term is equal to zero a.a.s. Here, there would be no gaps and there would be a lot of structure to investigate.

Below we determine thresholds for the appearance and disappearance of components of a given length. Initially, however, we have a brief look at the number of components in $\mathbf{C}_{n,p}$.

Proposition 3.0.2. *In $\mathbf{C}_{n,p}$, the expected number of components equals $nqp + p^2$, and the expected number of gaps equals $nqp + q^2$. Therefore, for any positive constant α, asymptotically,*

$$\mathbb{E}\big[\textit{number of components in } \mathbf{C}_{n,p}\big] \sim \begin{cases} 0 & \textit{if } \; p \ll n^{-1}, \\ \alpha & \textit{if } \; p \sim \alpha n^{-1}, \\ np & \textit{if } \; n^{-1} \ll p \ll 1, \\ npq & \textit{if } \; p \textit{ is constant}, \\ nq & \textit{if } \; 1 \gg q \gg n^{-1}, \\ \alpha + 1 & \textit{if } \; q \sim \alpha n^{-1}, \\ 1 & \textit{if } \; n^{-1} \gg q, \end{cases}$$

and

$$\mathbb{E}\big[\textit{number of gaps in } \mathbf{C}_{n,p}\big] \sim \begin{cases} 1 & \textit{if } \; p \ll n^{-1}, \\ \alpha + 1 & \textit{if } \; p \sim \alpha n^{-1}, \\ np & \textit{if } \; n^{-1} \ll p \ll 1, \\ npq & \textit{if } \; p \textit{ is constant}, \\ nq & \textit{if } \; 1 \gg q \gg n^{-1}, \\ \alpha & \textit{if } \; q \sim \alpha n^{-1}, \\ 0 & \textit{if } \; n^{-1} \gg q. \end{cases}$$

Proof. We count the left ends of components. The probability that the jth term of $\mathbf{C}_{n,p}$ is the start of a component equals p if $j = 1$ and qp if $2 \le j \le n$. Thus the expected number of components equals $p + (n-1)qp = nqp + p^2$.

- If $p \ll n^{-1}$, then $nqp + p^2 = np - np^2 + p^2 \ll 1 + n^{-2} \sim 1$. Thus $nqp + p^2 \sim 0$.
- If $p \sim \alpha n^{-1}$, then $nqp + p^2 = np - np^2 + p^2 \sim \alpha - \alpha^2 n^{-1} + \alpha^2 n^{-2} \sim \alpha$. Thus $nqp + p^2 \sim \alpha$.
- If $n^{-1} \ll p \ll 1$, then $nqp + p^2 = np - np^2 + p^2 \sim np$. Thus $nqp + \ p^2 \sim \ np$.
- If p is constant, then $nqp + p^2 \sim nqp$.
- If $q \sim \alpha n^{-1}$, then $nqp + p^2 = nq(1-q) + 1 - 2q + q^2 \sim \alpha(1 - \alpha n^{-1}) + 1 - 2\alpha n^{-1} + (\alpha n^{-1})^2 \sim \ \alpha \ + \ 1$. Thus, $nqp + p^2 \sim \ \alpha \ + \ 1$.
- If $n^{-1} \gg q$, then $nqp + p^2 = nq(1-q) + 1 - 2q + q^2 \sim 0(1-0) + 1 - 2(0) + 0^2 = 1$. Thus, $nqp + p^2 \sim 1$.

Due to the duality of gaps and components, the result of the expected number of gaps can be established by reversing the roles of p and q. □

Thus we have established that whenever $p \lesssim n^{-1}$, there are a finite number of components and gaps. By the duality of components and gaps, we have also established that whenever $q \lesssim n^{-1}$, there are also a finite number of components and gaps. We now establish that for any fixed $k \ge 2$, that $p \asymp n^{-1}$ is the lower bound and $q \asymp n^{-1}$ is the upper bound for there being at least k components and at least k gaps.

Proposition 3.0.3. *Suppose $k \ge 2$ is constant. Then,*

$$\mathbb{P}\big[\mathbf{C}_{n,p} \textit{ has at least } k \textit{ components}\big] \sim \begin{cases} 0 & \textit{if } \ p \ll n^{-1}, \\ 1 & \textit{if } \ n^{-1} \ll p \ \ \textit{and } \ q \gg n^{-1}, \\ 0 & \textit{if } \ n^{-1} \gg q, \end{cases}$$

and

$$\mathbb{P}\big[\mathbf{C}_{n,p} \textit{ has at least } k \textit{ gaps}\big] \sim \begin{cases} 0 & \textit{if } \ p \ll n^{-1}, \\ 1 & \textit{if } \ n^{-1} \ll p \quad \textit{and} \quad q \gg n^{-1}, \\ 0 & \textit{if } \ n^{-1} \gg q. \end{cases}$$

Proof. For each $i \in [n]$, let B_i be the event that the ith term of $\mathbf{C}_{n,p}$ is the beginning of a component. Thus, $\mathbb{P}\big[B_1\big] = p$, and $\mathbb{P}\big[B_i\big] = qp$ if $i > 1$. Suppose $\mathbf{i} := (i_1, i_2, \ldots, i_k) \in [n]^k$ is a vector such that $i_{j+1} \geq i_j + 2$ for each $j \in [k-1]$, and let $A_{\mathbf{i}} = B_{i_1} \wedge B_{i_2} \wedge \ldots \wedge B_{i_k}$. If $i_1 = 1$, then $\mathbb{P}\big[A_{\mathbf{i}}\big] = q^{k-1}p^k$; otherwise $\mathbb{P}\big[A_{\mathbf{i}}\big] = q^k p^k$.

If X is the total number of these k-tuples of components in $\mathbf{C}_{n,p}$, then by linearity of expectation, their expected number equals

$$\begin{aligned} E_k := \mathbb{E}\big[X\big] &= \binom{n-k}{k} q^k p^k + \binom{n-k}{k-1} q^{k-1} p^k \\ &= \left(\frac{(n-k)!}{k!(n-2k)!}\right) q^k p^k + \left(\frac{(n-k)!}{(k-1)!(n-2k+1)!}\right) q^{k-1} p^k \\ &\sim \left(\frac{n^k}{k!}\right) q^k p^k + \left(\frac{n^{k-1}}{(k-1)!}\right) q^{k-1} p^k \\ &= \left(\frac{n^k}{k!}\right) q^k p^k + \left(\frac{n^k k}{nq(k)!}\right) q^k p^k \\ &= \frac{n^k q^k p^k}{k!} \left(1 + \frac{k}{nq}\right). \end{aligned} \tag{3.1}$$

Suppose $p \ll n^{-1}$. So $p = n^{-1}/\omega$ with $\omega \gg 1$.

$$\begin{aligned} \mathbb{E}\big[X\big] &\sim \left(\frac{n^k \left(1 - \frac{n^{-1}}{\omega}\right)\left(\frac{n^{-1}}{\omega}\right)^k}{k!}\right)\left(1 + \frac{k}{n\left(1 - \frac{n^{-1}}{\omega}\right)}\right) \\ &\sim \left(\frac{n^k \left(\frac{n^k}{\omega^k}\right)}{k!}\right)\left(1 + \frac{k}{n\left(1 - \frac{1}{\omega}\right)}\right) \\ &\sim \frac{1}{\omega^k k!} \ll 1. \end{aligned}$$

Thus, by the First Moment Method (Proposition 2.5.2), $X = 0$ a.a.s., or equivalently, w.h.p. $\mathbf{C}_{n,p}$ has fewer than k components.

Now suppose that $q \ll n^{-1}$, so $q = n^{-1}/\omega$ with $\omega \gg 1$.

$$\begin{aligned}
\mathbb{E}\big[X\big] &\sim \left(\frac{n^k \left(\frac{n^{-1}}{\omega}\right)^k \left(1-\frac{n^{-1}}{\omega}\right)^k}{k!}\right)\left(1+\frac{k}{n\left(\frac{n^{-1}}{\omega}\right)}\right) \\
&\sim \left(\frac{1}{\omega^k k!}\right)(1+\omega k) \\
&\sim \frac{1}{\omega^{k-1}k!} \ll 1,
\end{aligned}$$

since $k \geq 2$. Therefore, by the First Moment Method, $X = 0$ a.a.s., or equivalently, $\mathbf{C}_{n,p}$ has less than k components.

Finally suppose that $p \gg n^{-1}$ and $q \gg n^{-1}$. We will use the Second Moment Method to prove the final part of the proposition. Distinct events $A_{\mathbf{i}}$ and $A_{\mathbf{j}}$ are correlated ($\mathbf{i} \sim \mathbf{j}$) if there exists a pair of indices i_r in $\mathbf{i}$ and j_s in $\mathbf{j}$ such that $|i_r - j_s| \leq 1$. If, for any such pair, their difference equals 1, then $\mathbb{P}\big[A_{\mathbf{i}} \wedge A_{\mathbf{j}}\big] = 0$. Otherwise, the event $A_{\mathbf{i}} \wedge A_{\mathbf{j}}$ represents, for some $t \in [k-1]$, the presence of $k+t$ component left ends, with the indices of $k-t$ of these occurring in both $\mathbf{i}$ and $\mathbf{j}$. So, $\mathbb{P}\big[A_{\mathbf{i}} \wedge A_{\mathbf{j}}\big] = E_{k+t}$. Thus, for some constant C_k,

$$\begin{aligned}
\Delta := \sum_{\mathbf{i}\sim\mathbf{j}} \mathbb{P}\big[A_{\mathbf{i}} \wedge A_{\mathbf{j}}\big] &= \sum_{t=1}^{k-1} \binom{k+t}{k}\binom{k}{t} E_{k+t} \\
&< C_k \sum_{t=1}^{k-1} E_{k+t} \\
&\sim C_k \sum_{t=1}^{k-1} \left(\frac{n^{k+t}q^{k+t}p^{k+t}}{(k+t)!}\right)\left(1+\frac{k+t}{nq}\right) \quad \text{(from equation 3.1)} \\
&< \frac{C_k}{k!}\left(1+\frac{2k}{nq}\right)\sum_{t=1}^{k-1}(nqp)^{k+t} \\
&= \frac{C_k}{k!}\left(1+\frac{2k}{nq}\right)\left(\frac{(npq)^{2k}-(npq)^{k+1}}{npq-1}.\right)
\end{aligned}$$

Suppose that $p \ll 1$. So $p = n^{-1}\omega$ where $1 \ll \omega \ll n$.

$$\begin{aligned}
\mathbb{E}[X] &\sim \frac{n^k p^k q^k}{k!}\left(1+\frac{k}{nq}\right) \\
&= \left(\frac{n^k\left(n^{-1}\omega\right)^k\left(1-n^{-1}\omega\right)^k}{k!}\right)\left(1+\frac{k}{n\left(1-n^{-1}\omega\right)}\right) \\
&\sim \frac{\omega^k}{k!} \gg 1,
\end{aligned}$$

and

$$\begin{aligned}
\Delta &\sim \frac{C_k}{k!}\left(1+\frac{2k}{nq}\right)\left(\frac{(npq)^{2k}-(npq)^{k+1}}{npq-1}\right) \\
&= \frac{C_k}{k!}\left(1+\frac{2k}{n\left(1-n^{-1}\omega\right)}\right)\left(\frac{\left(n\left(n^{-1}\omega\right)\left(1-n^{-1}\omega\right)\right)^{2k}-\left(n\left(n^{-1}\omega\right)\left(1-n^{-1}\omega\right)\right)^{k+1}}{n\left(n^{-1}\omega\right)\left(1-n^{-1}\omega\right)-1}\right) \\
&= \frac{C_k}{k!}\left(1+\frac{2k}{n-\omega}\right)\left(\frac{\left(\omega-\frac{\omega^2}{n}\right)^{2k}-\left(\omega-\frac{\omega^2}{n}\right)^{k+1}}{\omega-\frac{\omega^2}{n}-1}\right) \\
&\sim \frac{C_k}{k!}\left(\frac{\omega^{2k}}{\omega}\right) \\
&= \frac{C_k}{k!}\omega^{2k-1}.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\mathbb{E}[X]^2 &\sim \left(\left(\frac{n^k p^k q^k}{k!}\right)\left(1+\frac{k}{nq}\right)\right)^2 \\
&= \left(\frac{\left(n\left(n^{-1}\omega\right)\left(1-n^{-1}\omega\right)\right)^{2k}}{(k!)^2}\right)\left(1+\frac{k}{n\left(1-n^{-1}\omega\right)}\right)^2 \\
&= \left(\frac{\left(\omega-\frac{\omega^2}{n}\right)^{2k}}{(k!)^2}\right)\left(1+\frac{k}{n-\omega}\right)^2 \\
&\sim \left(\frac{\omega^k}{k!}\right)^2.
\end{aligned}$$

Therefore,

$$\frac{\Delta}{\mathbb{E}[X]^2} \sim \left(\frac{C_k}{k!}\omega^{2k-1}\right) \Big/ \left(\frac{\omega^k}{k!}\right)^2 \sim \frac{k!C_k}{\omega} \ll 1.$$

Similarly, if $\omega \ll n$ and $q = n^{-1}\omega$, then

$$\Delta \sim \frac{C_k}{k!}\left(1 + \frac{2k}{nq}\right)\left(\frac{(npq)^{2k} - (npq)^{k+1}}{npq - 1}\right) \sim \frac{C_k}{k!}\omega^{2k-1}$$

and

$$\frac{\Delta}{\mathbb{E}[X]^2} \sim \frac{k!C_k}{\omega} \ll 1.$$

Finally, if p is asymptotically a constant, then

$$\mathbb{E}[X] \sim \frac{n^k p^k q^k}{k!}\left(1 + \frac{k}{nq}\right) \sim \frac{n^k p^k q^k}{k!} = \beta_k n^k$$

for some constant β_k. Furthermore,

$$\Delta \sim \frac{C_k}{k!}\left(1 + \frac{2k}{nq}\right)\left(\frac{(npq)^{2k} - (npq)^{k+1}}{npq - 1}\right) \sim \frac{C_k}{k!}(npq)^{2k-1}.$$

Therefore,

$$\frac{\Delta}{\mathbb{E}[X]^2} \sim \frac{\frac{C_k}{k!}(npq)^{2k-1}}{n^{2k}\beta_k} = \left(\frac{C_k}{\beta_k^2 k!}\right)(pq)^{2k-1}\left(\frac{1}{n}\right) \ll 1$$

since $\left(\frac{C_k}{\beta_k^2 k!}\right)(pq)^{2k-1}$ is constant.

So by the Second Moment Method (Proposition 2.5.4), if both $n^{-1} \ll p$ and $q \gg n^{-1}$ then $X > 0$ a.a.s., or equivalently, w.h.p. $\mathbf{C}_{n,p}$ has at least k components. By duality between components and gaps, the threshold is identical for gaps as it is for components. □

3.1 The Longest Component and Longest Gap

We now establish thresholds for $\mathbf{C}_{n,p}$ to have a component or gap exceeding a specified length. This topic of research has been studied in some detail before with similar methods utilised and, in some cases, applications to other related areas of mathematics (see [4], [30]). These properties are monotone, that is, we cannot 'destroy' the property that a component is of at least a certain length by increasing a term of the composition by 1. Similarly, if a gap is of at least a certain length, we cannot 'destroy' this property by removing 1 from a term of the composition. If C is a composition, let $\mathsf{comp}_{\max}(C)$ be the length of the longest component of C, and $\mathsf{gap}_{\max}(C)$ be the length of the longest gap in C.

Given some value of k, for each $i \in [n+1-k]$, let A_i be the event "$\mathbf{C}_{n,p}(i), \ldots, \mathbf{C}_{n,p}(i+k-1)$ are all nonzero". Then $\mathbb{P}[A_i] = p^k$. So, if X is the total number of runs of k nonzero terms in $\mathbf{C}_{n,p}$, then by linearity of expectation, $\mathbb{E}[X] = (n+1-k)p^k \sim np^k$, as long as $k \ll n$. Note that if we have a run of exactly $k+4$ nonzero terms for example, we would take that to mean there are 5 runs of k nonzero terms within these $k+4$ nonzero terms.

Distinct events A_i and A_j are correlated ($i \sim j$) if $|i-j| < k$. If $i \sim j$ and $i < j$, then $j = i + t$ for some $t \in [k-1]$, and $\mathbb{P}[A_i \wedge A_j] = p^{k+t}$. So,

$$\Delta := \sum_{i \sim j} \mathbb{P}[A_i \wedge A_j] < np^k \sum_{t=1}^{k-1} p^t < np^k \sum_{t=1}^{\infty} p^t \;=\; np^{k+1}/q,$$

and $\Delta/\mathbb{E}[X]^2 \lesssim p/np^k q$. Moreover,

$$\Lambda := \sum_{i} \mathbb{P}[A_i]^2 \;+\; \sum_{i \sim j} \mathbb{P}[A_i]\mathbb{P}[A_j] \;<\; (n+1-k)p^{2k} + nkp^{2k} \;\sim\; nkp^{2k}.$$

To apply the Chen–Stein Method (Proposition 2.5.7), it is sufficient to show that $\Delta \ll 1$ and $\Lambda \ll 1$.

We now establish the threshold for the appearance of a component of at least fixed length k in $\mathbf{C}_{n,p}$ to be $p \asymp n^{-1/k}$. Similarly we establish the threshold for the appearance of a gap of at least fixed length k in $\mathbf{C}_{n,p}$ to be $q \asymp n^{-1/k}$.

Proposition 3.1.1. *Suppose $k \geq 1$ is constant. Then, for any positive constant α,*

$$\mathbb{P}\big[\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) \geq k\big] \sim \begin{cases} 0 & \text{if } p \ll n^{-1/k}, \\ 1 - e^{-\alpha^k} & \text{if } p \sim \alpha n^{-1/k}, \\ 1 & \text{if } n^{-1/k} \ll p, \end{cases}$$

$$\mathbb{P}\big[\mathsf{gap}_{\max}(\mathbf{C}_{n,p}) \geq k\big] \sim \begin{cases} 1 & \text{if } q \gg n^{-1/k}, \\ 1 - e^{-\alpha^k} & \text{if } q \sim \alpha n^{-1/k}, \\ 0 & \text{if } n^{-1/k} \gg q. \end{cases}$$

Proof. If $p \ll n^{-1/k}$ and X is the total number of runs of k nonzero terms in $\mathbf{C}_{n,p}$ then $\mathbb{E}\big[X\big] \sim np^k \ll 1$, so by the First Moment Method, $X = 0$ a.a.s., or equivalently, $\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) < k$ a.a.s.

If $n^{-1/k} \ll p$, then $\mathbb{E}\big[X\big] \gg 1$. If $p \ll 1$, then

$$\Delta/\mathbb{E}\big[X\big]^2 \lesssim \frac{p^{1-k}}{nq} \sim \frac{p^{1-k}}{n} \ll n^{-1/k} \ll 1.$$

So by the Second Moment Method, $X > 0$ a.a.s., or equivalently, $\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) \geq k$ a.a.s. Since the property of having a component of length at least k is increasing, then by Proposition 2.4.3 this also holds for larger p.

Finally, suppose that $p \sim \alpha n^{-1/k}$. Then $\mathbb{E}\big[X\big] \sim \alpha^k$ and $\Delta < \alpha^k p/q \ll 1$. Moreover, we have $\Lambda < nkp^{2k} \sim \alpha^{2k} k n^{-1} \ll 1$. So, by the Chen–Stein Method (Proposition 2.5.7), the number of components in $\mathbf{C}_{n,p}$ of length k converges in distribution to a Poisson distribution with mean α^k. In particular, the probability that no components have length k or greater is asymptotically $e^{-\alpha^k}$ as $n \to \infty$. □

Thus (using Proposition 2.5.1 to transfer the thresholds from $\mathbf{C}_{n,p}$ to $\mathbf{C}_{n,m}$), as m increases, for some constant α, we first see components of length 2 in $\mathbf{C}_{n,m}$ when $m \sim \frac{n(\alpha n^{-1/2})}{1-n^{-1/2}} \sim \alpha n^{1/2}$ a.a.s. Furthermore, for some constant α, we first see components of length 3 when $m \sim \frac{n(\alpha n^{-1/3})}{1-\alpha n^{-1/3}} \sim \alpha n^{2/3}$ a.a.s and so forth.

Ergo, if $m \sim \alpha n^c$, for positive constants α and $c = 1 - 1/k$ such that $k \in \mathbb{N}$, then

$\mathsf{comp}_{\max}(\mathbf{C}_{n,m})$ is equal to k or $k-1$ a.a.s. This is due to $\mathbb{P}\big[\mathsf{comp}_{\max}(\mathbf{C}_{n,m}) \geq k\big] \sim 1 - e^{-\alpha^k} \notin \{0,1\}$. If $c \in (0,1)$ and $c \neq 1 - 1/k$, then $\mathsf{comp}_{\max}(\mathbf{C}_{n,m})$ only takes one value a.a.s.

However, in $\mathbf{C}_{n,p}$, once $p \gg n^{-1/k}$ or, in $\mathbf{C}_{n,m}$, once $m \gg \frac{n(n^{-1/k})}{1-n^{-1/k}} \sim n^{1-1/k}$ for every k (for example, $m = n/\log n$), a.a.s. the length of the longest component exceeds any fixed value. In this thesis log denotes $\log_e$ unless stated otherwise.

Similarly, a.a.s. gaps of length 3 'vanish' in $\mathbf{C}_{n,p}$ once $q \ll n^{-1/3}$, or in $\mathbf{C}_{n,m}$, once $m \gg \frac{n(1-n^{-1/3})}{n^{-1/3}} \sim n^{4/3}$, every gap has length 1, in $\mathbf{C}_{n,p}$, when $q \ll n^{-1/2}$, or in $\mathbf{C}_{n,m}$ once $m \gg \frac{n(1-n^{-1/2})}{n^{-1/2}} \sim n^{3/2}$, and there are no gaps at all, in $\mathbf{C}_{n,p}$ once $q \ll n^{-1}$, or in $\mathbf{C}_{n,m}$ once $m \gg \frac{n(1-n^{-1})}{n^{-1}} \sim n^2$.

We now turn our attention to the appearance and disappearance of gaps and components of length k as k increases with n.

Proposition 3.1.2. *Suppose $1 \ll k \ll \log n$. Then, for any $\omega \gg 1$ and constant α,*

$$\mathbb{P}\big[\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) \geq k\big] \sim \begin{cases} 0 & \textit{if } \; p = e^{-(\log n+\omega)/k}, \\ 1 - e^{-e^{\alpha}} & \textit{if } \; p = e^{-(\log n-\alpha)/k}, \\ 1 & \textit{if } \; p = e^{-(\log n-\omega)/k}, \end{cases}$$

$$\mathbb{P}\big[\mathsf{gap}_{\max}(\mathbf{C}_{n,p}) \geq k\big] \sim \begin{cases} 1 & \textit{if } \; q = e^{-(\log n-\omega)/k}, \\ 1 - e^{-e^{\alpha}} & \textit{if } \; q = e^{-(\log n-\alpha)/k}, \\ 0 & \textit{if } \; q = e^{-(\log n+\omega)/k}. \end{cases}$$

Proof. First suppose $p = e^{-(\log n+\omega)/k}$ and X is the total number of runs of k nonzero terms in $\mathbf{C}_{n,p}$, then

$$\mathbb{E}\big[X\big] \sim np^k = n\left(e^{-\frac{\log n+\omega}{k}}\right)^k = ne^{-\log n-\omega} = (n)\left(\frac{1}{n}\right)e^{-\omega} = e^{-\omega} \ll 1.$$

So by the First Moment Method, $X = 0$ a.a.s., or equivalently $\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) < k$ a.a.s.

Now suppose $p = e^{-(\log n-\omega)/k}$, then $\mathbb{E}\big[X\big] \sim e^{\omega} \gg 1$. If $\omega \ll \log n$, then $p \ll 1$ and

$$\Delta/\mathbb{E}\big[X\big]^2 < \frac{\big(np^{k+1}/q\big)}{\big(np^k\big)^2} = \frac{e^{\omega}p}{e^{2\omega}q} = \frac{p}{q}e^{-\omega} \ll 1.$$

So by the Second Moment Method, $X > 0$ or equivalently, $\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) \geq k$ a.a.s. Since the property of having a component of length at least k is increasing, then by Proposition 2.4.3 this also holds for larger p (faster growing ω).

Finally, if $p \sim e^{-(\log n - \alpha)/k}$, then $\mathbb{E}\big[X\big] \sim e^{\alpha}$ and $\Delta < pe^{\alpha}/q \ll 1$. Moreover, we have $\Lambda < nkp^{2k} \sim \frac{nke^{2\alpha}}{n^2} = e^{2\alpha}kn^{-1} \ll 1$. So, by the utilisation of the Chen-Stein method (Proposition 2.5.7), the number of components in $\mathbf{C}_{n,p}$ of length k converges in distribution to a Poisson distribution with mean e^{α}. In particular, the probability that no components have length k or greater is asymptotically $e^{-e^{\alpha}}$ as $n \to \infty$.

□

We continue our investigation into gaps and components of length k as k increases with n. Once more, we establish a sharp threshold.

Proposition 3.1.3. *Suppose $k = c\log n$ for some constant c. Then, for any $\omega \gg 1$,*

$$\mathbb{P}\big[\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) \geq k\big] \sim \begin{cases} 0 & \textit{if} \;\; p = e^{-1/c - \omega/\log n}, \\ 1 & \textit{if} \;\; p = e^{-1/c + \omega/\log n}, \end{cases}$$

$$\mathbb{P}\big[\mathsf{gap}_{\max}(\mathbf{C}_{n,p}) \geq k\big] \sim \begin{cases} 1 & \textit{if} \;\; q = e^{-1/c + \omega/\log n}, \\ 0 & \textit{if} \;\; q = e^{-1/c - \omega/\log n}. \end{cases}$$

Proof. If $p = e^{-1/c - \omega/\log n}$ and X is the total number of runs of k nonzero terms in $\mathbf{C}_{n,p}$. Then suppose

$$\mathbb{E}\big[X\big] \sim np^k = n\left(e^{-1/c - \omega/\log n}\right)^{c\log n} = ne^{-\log n - \omega c} = n\left(\frac{1}{n}\right)e^{-\omega c} = e^{-c\omega} \ll 1.$$

So by the First Moment Method, $X = 0$ a.a.s. or equivalently, $\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) < k$ a.a.s.

Now suppose $p = e^{-1/c + \omega/\log n}$, then $\mathbb{E}\big[X\big] \sim e^{c\omega} \gg 1$. If $\omega \ll \log n$, then p tends to a finite value and

$$\frac{\Delta}{\mathbb{E}\left[X\right]^2} < \frac{n\left(p^{k+1}/q\right)}{\left(np^k\right)^2} = \frac{p}{np^k q} = \frac{pe^{-c\omega}}{q} \ll 1.$$

So $\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) \geq k$ a.a.s. by the Second Moment Method. Since the property of having a component of length at least k is increasing, then by Proposition 2.4.3 this also holds for larger p (faster growing ω). □

We now establish one final pair of thresholds in this section, one for the longest gap and one for the longest component. Here we investigate the length $k = n^c$ for some $c \in (0, 1)$ and establish the threshold is $q = k^{-1} \log n$ for the appearance of these components and $p = k^{-1} \log n$ for the appearance of these gaps.

Proposition 3.1.4. *Suppose $k = n^c$ for some $c \in (0, 1)$. Then, for any $\omega \gg 1$,*

$$\mathbb{P}\left[\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) \geq k\right] \sim \begin{cases} 0 & \text{if} \;\; q = k^{-1}(\log n + \omega), \\ 1 & \text{if} \;\; q = k^{-1}\big((1-c)\log n - \omega\big), \end{cases}$$

$$\mathbb{P}\left[\mathsf{gap}_{\max}(\mathbf{C}_{n,p}) \geq k\right] \sim \begin{cases} 1 & \text{if} \;\; p = k^{-1}\big((1-c)\log n - \omega\big), \\ 0 & \text{if} \;\; p = k^{-1}(\log n + \omega). \end{cases}$$

Proof. First suppose $q = (\log n + \omega)/n^c$, then $p = 1 - (\log n + \omega)/n^c$ and X is the total number of runs of k nonzero terms in $\mathbf{C}_{n,p}$.

$$\mathbb{E}\left[X\right] \sim np^k = n\big(1 - (\log n + \omega)/k\big)^k \sim ne^{-\log n - \omega} = e^{-\omega} \ll 1.$$

So by the First Moment Method, $X = 0$ a.a.s or equivalently, $\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) < k$ a.a.s.

We need to use an alternative bound on Δ.

$$\Delta := \sum_{i \sim j} \mathbb{P}\left[A_i \wedge A_j\right] < np^k \sum_{t=1}^{k-1} p^t < nkp^{k+1}.$$

Thus $\Delta/\mathbb{E}\left[X\right]^2 \lesssim pk/n(1-q)^k$.

Now suppose $q = (\log n - \omega)/n^c$, then $p = 1 - (\log n - \omega)/n^c$ and $\mathbb{E}\left[X\right] \sim e^{\omega} \gg 1$.

Furthermore,

$$\Delta/\mathbb{E}\big[X\big]^2 < \frac{nkp^{k+1}}{\left(np^k\right)^2} = \frac{kp}{np^k} \sim kpe^{-\omega} \ll 1.$$

So $\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) \geq k$ a.a.s. by the Second Moment Method. □

3.2 The Shortest Component and Shortest Gap

We now establish thresholds for $\mathbf{C}_{n,p}$ to have a component shorter than a certain length and also for $\mathbf{C}_{n,p}$ to have a gap shorter than a certain length. These are not monotone properties. For example, adding one to the last term of the composition 413300 yields 413301 which reduces the length of the shortest component from 4 to 1. On the other hand, removing one from the last term of the composition 413301 yields 413300 which increases the length of the shortest component from 1 to 4. Similarly, removing one from the last term of the composition 00021 gives 00020 which reduces the length of the shortest gap from 3 to 1. On the other hand adding one to the last term of the composition 00020 gives 00021 which increases the length of the shortest gap from 1 to 3. If C is a composition, let $\mathsf{comp}_{\min}(C)$ be the length of the shortest component in C, and $\mathsf{gap}_{\min}(C)$ be the length of the shortest gap in C.

For each $\ell \in [n-1]$ and each $i \in [n+1-\ell]$, let $A_{i,\ell}$ be the event that the ith term of $\mathbf{C}_{n,p}$ is the start of a component of length ℓ. Then

$$\mathbb{P}\big[A_{i,\ell}\big] \;=\; \begin{cases} qp^\ell & \text{if } i = 1 \text{ or } i = n+1-\ell, \\ q^2p^\ell & \text{otherwise.} \end{cases}$$

So, assuming $\ell \ll n$, if X_ℓ is the number of components of length ℓ in $\mathbf{C}_{n,p}$, then

$$\mathbb{E}\big[X_\ell\big] \;=\; (n-1-\ell)q^2p^\ell + 2qp^\ell \sim nq^2p^\ell.$$

Given some $k \ll n$ and assuming $kq \ll 1$ (so $1-p^k = 1-(1-q)^k = kq + O(q^2) \sim kq$), let X be the total number of components of length at most k in $\mathbf{C}_{n,p}$. Then

$$\mathbb{E}\big[X\big] \;=\; \sum_{\ell=1}^{k} \mathbb{E}\big[X_\ell\big] \sim nq^2(p + p^2 + \ldots + p^k) \;=\; npq(1-p^k) \sim knq^2.$$

Distinct events $A_{i,r}$ and $A_{j,s}$ $(i \leq j)$ are correlated $(i, r \sim j, s)$ in two situations. If $j \leq i + r$, then the corresponding components overlap in a contradictory manner, so $\mathbb{P}\big[A_{i,r} \wedge A_{j,s}\big] = 0$. If $j = i + r + 1$, then the corresponding components are separated by a single zero term and $\mathbb{P}\big[A_{i,r} \wedge A_{j,s}\big] = q^3 p^{r+s}$, except when the pair of components occur at the start or end of the composition, in which case $\mathbb{P}\big[A_{i,r} \wedge A_{j,s}\big] = q^2 p^{r+s}$. Thus,

$$\begin{aligned}\Delta := \sum_{i,r \sim j,s} \mathbb{P}\big[A_{i,r} \wedge A_{j,s}\big] \;&=\; \sum_{r=1}^{k}\sum_{s=1}^{k}(n-2-r-s)q^3p^{r+s} + 2q^2p^{r+s} \\ &\sim n\sum_{r=1}^{k}\sum_{s=1}^{k} q^3p^{r+s} \;=\; np^2q\big(1-p^k\big)^2 \sim k^2nq^3.\end{aligned}$$

Thus $\Delta/\mathbb{E}\big[X\big]^2 \sim 1/nq$, which tends to zero as long as $q \gg n^{-1}$. Moreover,

$$\begin{aligned}\Lambda &:= \sum_{i,\ell} \mathbb{P}\big[A_{i,\ell}\big]^2 + \sum_{i,r \sim j,s} \mathbb{P}\big[A_{i,r}\big]\mathbb{P}\big[A_{j,s}\big] \\ &\sim n\sum_{r=1}^{k}(r+2)\sum_{s=1}^{k} q^4p^{r+s} \sim \frac{1}{2}k^2(k+5)nq^4 \;\lesssim\; k^3nq^4.\end{aligned}$$

The first of our results in this section establishes the threshold for fixed k, that the appearance of gaps of length less than k occurs at $p \asymp n^{-1/2}$, whereas the disappearance of components of length less than k occurs at $q \asymp n^{-1/2}$.

Proposition 3.2.1. *Suppose $k \geq 1$ is constant. Then, for any positive constant α,*

$$\mathbb{P}\big[\mathsf{comp}_{\min}(\mathbf{C}_{n,p}) > k\big] \sim \begin{cases} 0 & \text{if } q \gg n^{-1/2}, \\ e^{-\alpha^2 k} & \text{if } q \sim \alpha n^{-1/2}, \\ 1 & \text{if } n^{-1/2} \gg q, \end{cases}$$

$$\mathbb{P}\big[\mathsf{gap}_{\min}(\mathbf{C}_{n,p}) > k\big] \sim \begin{cases} 1 & \text{if } p \ll n^{-1/2}, \\ e^{-\alpha^2 k} & \text{if } p \sim \alpha n^{-1/2}, \\ 0 & \text{if } n^{-1/2} \ll p. \end{cases}$$

Proof. Suppose $\omega \gg 1$ and $q = n^{-1/2}/\omega$ and that X is the total number of components of length at most k in $\mathbf{C}_{n,p}$. Then

$$\mathbb{E}\big[X\big] \sim knq^2 = nk\left(n^{-1/2}\omega^{-1}\right)^2 = k/\omega^2 \ll 1,$$

so by the First Moment Method, $X = 0$ a.a.s. or equivalently there are no components of length k or less, and $\mathsf{comp}_{\min}(\mathbf{C}_{n,p}) > k$ a.a.s.

Now suppose $q = n^{-1/2}\omega$, then $\mathbb{E}\big[X\big] \sim knq^2 = nk\left(n^{-1/2}\omega\right)^2 = k\omega^2 \gg 1$. Furthermore,

$$\Delta/\mathbb{E}\big[X\big]^2 \sim \frac{k^2nq^3}{(knq^2)^2} = \frac{k}{nq} = \frac{k}{n^{1/2}\omega} \ll 1.$$

So by the Second Moment Method, $X > 0$ a.a.s or equivalently, $\mathsf{comp}_{\min}(\mathbf{C}_{n,p}) \leq k$ a.a.s.

Finally, suppose $q \sim \alpha n^{-1/2}$, then $\mathbb{E}\big[X\big] \sim \alpha^2 k$. Also,

$$\Delta \sim k^2nq^3 \sim k^2n\left(\alpha n^{-1/2}\right)^3 \sim \alpha^3k^2n^{-1/2} \ll 1.$$

Moreover, we have

$$\Lambda \sim \frac{1}{2}k^2(k+5)nq^4 \sim \frac{1}{2}k^2(k+5)n\left(\alpha n^{-1/2}\right)^4 = \frac{\frac{1}{2}\alpha^4k(k+5)}{n} \ll 1.$$

So, the number of components in $\mathbf{C}_{n,p}$ of length at most k converges in distribution to a Poisson distribution with mean $\alpha^2 k$. In particular, the probability that no components have length k or less is asymptotically $e^{-\alpha^2 k}$ as $n \to \infty$. $\square$

Thus, when $p \gg n^{-1/2}$, we see gaps of length 1 in $\mathbf{C}_{n,p}$ a.a.s. While $p \ll n^{-1/2}$, or while $m \sim \frac{n(n^{-1/2})}{1-n^{-1/2}} \sim n^{-1/2}$ in $\mathbf{C}_{n,m}$, there is no gap of any fixed length a.a.s. On the other hand, components of length 1 do not disappear until $q \asymp n^{-1/2}$ in $\mathbf{C}_{n,p}$, or until $m \asymp n^{3/2}$ in $\mathbf{C}_{n,m}$. Here however, by Proposition 3.1.4, the longest components have length of the order of $\sqrt{n}\log n$ a.a.s. Once $n^{-1/2} \gg q$ however, no component of any fixed length remains.

We conclude this chapter by establishing the thresholds for the shortest gap in $\mathbf{C}_{n,p}$ and shortest component in $\mathbf{C}_{n,p}$ as k grows with n.

Proposition 3.2.2. *Suppose* $1 \ll k \ll n$. *Then, for any positive constant* α*,*

$$\mathbb{P}\big[\mathsf{comp}_{\min}(\mathbf{C}_{n,p}) > k\big] \sim \begin{cases} 0 & \text{if } q \gg 1/\sqrt{kn}, \\ e^{-\alpha^2} & \text{if } q \sim \alpha/\sqrt{kn}, \\ 1 & \text{if } 1/\sqrt{kn} \gg q, \end{cases}$$

$$\mathbb{P}\big[\mathsf{gap}_{\min}(\mathbf{C}_{n,p}) > k\big] \sim \begin{cases} 1 & \text{if } p \ll 1/\sqrt{kn}, \\ e^{-\alpha^2} & \text{if } p \sim \alpha/\sqrt{kn}, \\ 0 & \text{if } 1/\sqrt{kn} \ll p. \end{cases}$$

Note that $kq \sim \sqrt{k/n} \ll 1$, *as required for our asymptotics to be valid.*

Proof. Suppose $\omega \gg 1$ and $q = \omega^{-1}/\sqrt{kn}$ and that let X be the total number of components of length at most k in $\mathbf{C}_{n,p}$. Then

$$\mathbb{E}\big[X\big] \sim knq^2 = nk\left(\omega^{-1}/\sqrt{kn}\right)^2 = nk\left(\omega^{-2}/kn\right) = \omega^{-2} \ll 1.$$

So by the First Moment Method, $X = 0$ a.a.s. or equivalently, there are no components of length k or less and $\mathsf{comp}_{\min}(\mathbf{C}_{n,p}) > k$ a.a.s.

Now suppose $q = \omega/\sqrt{kn}$, then $\mathbb{E}[X] \sim knq^2 = kn\left(\omega^2/kn\right) = \omega^2 \gg 1$. Furthermore,

$$\Delta/\mathbb{E}[X]^2 \sim \frac{k^2nq^3}{(knq^2)^2} = \frac{k}{nq} = \frac{k}{n\left(\omega(kn)^{-1/2}\right)} = \frac{k^{3/2}}{\omega n^{1/2}} \ll 1.$$

So by the Second Moment Method, $X > 0$ a.a.s. or equivalently, $\mathsf{comp}_{\min}(\mathbf{C}_{n,p}) \leq k$ a.a.s.

Finally, suppose $q \sim \alpha/\sqrt{kn}$, then $\mathbb{E}[X] \sim knq^2 \sim \alpha^2$ and

$$\Delta \sim k^2nq^3 \sim k^2n\left(\alpha^3(kn)^{-3/2}\right) = \frac{k^{1/2}\alpha^3}{n^{1/2}} \ll 1.$$

Moreover, we have

$$\Lambda \sim \frac{1}{2}k^2(k+5)nq^4 \sim \frac{1}{2}k^2(k+5)n\left(\frac{\alpha}{\sqrt{kn}}\right)^4 = \frac{\frac{1}{2}(k+5)\alpha^4}{n} \ll 1.$$

So, the number of components in $\mathbf{C}_{n,p}$ of length at most k converges in distribution to a Poisson distribution with mean α^2. In particular, the probability that no components have length k or less is asymptotically $e^{-\alpha^2}$ as $n \to \infty$. $\square$

Chapter 4

Largest Composition Terms

In this short chapter, we establish thresholds for the largest terms in a composition that are of at least a specified size.

We first introduce some new notation. Let C be a composition, then $\max(C)$ is the largest term of C.

Proposition 4.0.1. *Suppose $\omega \gg 1$ and $p = \omega^{-1}$. Then, for any constant α,*

$$\mathbb{P}\big[\mathsf{max}(\mathbf{C}_{n,p}) \geq r\big] \;\sim\; \begin{cases} 0 & \text{if } \; r - \frac{\log n}{\log \omega} \gg \frac{1}{\log \omega}, \\ 1 - e^{-e^{-\alpha}} & \text{if } \; r = \frac{\log n + \alpha}{\log \omega}, \\ 1 & \text{if } \; \frac{\log n}{\log \omega} - r \gg \frac{1}{\log \omega}. \end{cases}$$

Proof. The probability that term $i \in [n]$ is less than r is given by

$$\mathbb{P}\big[C(i) < r\big] = q + qp + qp^2 + \cdots + qp^{r-1} = q \sum_{k=0}^{r-1} p^k = 1 - p^r = 1 - \omega^{-r}.$$

Suppose $r = \frac{\log n + \delta}{\log \omega}$, then

$$\begin{aligned}1-\omega^{-r} &= 1-\omega^{-\frac{\log n+\delta}{\log\omega}}\\ &= 1-\exp\left(-\left(\frac{\log n+\delta}{\log\omega}\right)\log\omega\right)\\ &= 1-\exp(-\log n-\delta)\\ &= 1-\frac{e^{-\delta}}{n}.\end{aligned}$$

Therefore,

$$\mathbb{P}\big[\mathsf{max}(\mathbf{C}_{n,p})<r\big] \;=\; \left(1-\frac{e^{-\delta}}{n}\right)^n \;\sim\; \begin{cases}1 & \text{if } \delta\gg 1,\\ e^{-e^{-\alpha}} & \text{if } \delta\sim\alpha,\\ 0 & \text{if } -\delta\gg 1.\end{cases} \qquad \square$$

Thus, $\mathsf{max}(\mathbf{C}_{n,p}) \sim \log n/\log(1/p)$ if $p \ll 1$ a.a.s. We now consider the case where p is constant.

Proposition 4.0.2. *If p is constant, then for any constant c,*

$$\mathbb{P}\big[\mathsf{max}(\mathbf{C}_{n,p})\geq r\big] \;\sim\; \begin{cases}0 & \textit{if } r-\log_{1/p} n\gg 1,\\ 1-e^{-p^c} & \textit{if } r=\log_{1/p} n+c,\\ 1 & \textit{if } \log_{1/p} n-r\gg 1.\end{cases}$$

Proof. The probability that term $i\in[n]$ is less than r is given by

$$\mathbb{P}\big[C(i)<r\big]=q+qp+qp^2+\cdots+qp^{r-1}=q\sum_{k=0}^{r-1}p^k=1-p^r.$$

If $r=\log_{1/p} n+\delta$, then

$$\begin{aligned}
1-p^r &= 1-p^{\log_{1/p} n+\delta} \\
&= 1-\exp\left(\left(\log_{1/p} n+\delta\right)\log p\right) \\
&= 1-\exp\left(\left(\log_{1/p} n\right)(\log p)+\log p^{\delta}\right) \\
&= 1-\exp\left(-\left(\frac{\log n}{\log p}\right)(\log p)+\log p^{\delta}\right) \\
&= 1-\exp\left(\log n^{-1}+\log p^{\delta}\right) \\
&= 1-\frac{p^{\delta}}{n}.
\end{aligned}$$

Therefore,

$$\mathbb{P}\big[\mathsf{max}(\mathbf{C}_{n,p})<r\big] \;=\; \left(1-\frac{p^{\delta}}{n}\right)^n \;\sim\; \begin{cases} 1 & \text{if } \;\delta \gg 1, \\ e^{-p^c} & \text{if } \;\delta \sim c, \\ 0 & \text{if } \;-\delta \gg 1. \end{cases}$$

□

Thus, when p is constant, the distribution of the largest term is concentrated around $\log_{1/p} n$. We now investigate the case for when p tends to 1.

Proposition 4.0.3. *If $q \ll 1$, then for any $\varepsilon > 0$,*

$$\mathbb{P}\big[\mathsf{max}(\mathbf{C}_{n,p})\geq r\big] \;\sim\; \begin{cases} 0 & \textit{if } \; r \geq (1+\varepsilon)q^{-1}\log n, \\ 1 & \textit{if } \; r \leq (1-\varepsilon)q^{-1}\log n. \end{cases}$$

Proof. The probability that term $i \in [n]$ is less than r is given by

$$\mathbb{P}\big[C(i)<r\big] = q+qp+qp^2+\cdots+qp^{r-1} = q\sum_{k=0}^{r-1} p^k = 1-p^r.$$

Thus,

$$\begin{aligned}\mathbb{P}\big[\mathsf{max}(\mathbf{C}_{n,p}) < r\big] &= (1-p^r)^n \\ &= (1-(1-q)^r)^n \\ &= (1-\exp(r\log(1-q)))^n\,.\end{aligned}$$

Let $L = \log\mathbb{P}\big[\mathsf{max}(\mathbf{C}_{n,p}) < r\big] = n\log\big(1-\exp\big(r\log(1-q)\big)\big)$.

Now, for small enough x, we have $-2x < \log(1-x) < -x$.

So, if $r = (1+\varepsilon)q^{-1}\log n$, then for sufficiently large n,

$$L > n\log\big(1-\exp(-rq)\big) \;=\; n\log\big(1-n^{-(1+\varepsilon)}\big) > -2n^{-\varepsilon}.$$

Thus $L \sim 0$, and $\mathbb{P}\big[\mathsf{max}(\mathbf{C}_{n,p}) < r\big] \sim 1$.

Similarly, using the tighter bound $-x-x^2 < \log(1-x)$, if now $r = (1-\varepsilon)q^{-1}\log n$, then for sufficiently large n,

$$\begin{aligned}L &= n\log(1-\exp(r\log(1-q))) \\ &< n\log\left(1-\exp\left(r\left(-q-q^2\right)\right)\right) \\ &< n\log\big(1-\exp(-rq(1+q))\big) \\ &= \; n\log\big(1-n^{-(1-\varepsilon)(1+q)}\big) \\ &< -n^{\varepsilon-(1-\varepsilon)q}.\end{aligned}$$

Thus $\lim_{n\to\infty} L = -\infty$ (since $q \ll 1$), and $\mathbb{P}\big[\mathsf{max}(\mathbf{C}_{n,p}) < r\big] \sim 0$. □

Hence, when $m \gg n$, a.a.s. $\mathsf{max}(\mathbf{C}_{n,m}) \sim \frac{m}{n}\log n$, a factor of $\log n$ more than the value of the average term.

We conclude this chapter by utilising the previous three results in the following proposition.

Proposition 4.0.4.

$$\lim_{n\to\infty}\mathbb{P}\big[\,\max(\mathbf{C}_{n,m}) \gg \frac{m}{n}\log n\big] \;=\; 0.$$

Proof. First suppose $p \ll 1$. By Proposition 4.0.1, the maximum term in $\mathbf{C}_{n,p}$ is asymptotic to $\log n / \log(1/p)$ and

$$\lim_{n\to\infty} \mathbb{P}\big[\max(\mathbf{C}_{n,p}) \gg \frac{\log n}{\log(1/p)}\big].$$

By Proposition 2.5.1,

$$\lim_{n\to\infty} \mathbb{P}\big[\max(\mathbf{C}_{n,m}) \gg \frac{\log n}{\log\left(\frac{n}{m}\right)} \sim \frac{\log n}{\log n - \log m} \gg \frac{m}{n}\log n\big] = 0,$$

since $m \ll n$.

Now suppose p is asymptotically a constant. By Proposition 4.0.2, the maximum term in $\mathbf{C}_{n,p}$ is asymptotic to $\log_{\frac{1}{p}} n$ and

$$\lim_{n\to\infty} \mathbb{P}\big[\max(\mathbf{C}_{n,p}) \gg \log_{\frac{1}{p}} n\big].$$

By Proposition 2.5.1 and for some constants α, β, γ,

$$\lim_{n\to\infty} \mathbb{P}\big[\max(\mathbf{C}_{n,m}) \gg \log_{\frac{n}{\gamma m}} n \sim \log_{\frac{1}{\beta}} n = \delta \log n\big] = 0.$$

Since $m \sim \alpha n$ for some constant α, then $\frac{m}{n}\log n = \alpha \log n$ and therefore,

$$\lim_{n\to\infty} \mathbb{P}\big[\max(\mathbf{C}_{n,m}) \gg \frac{m}{n}\log n\big] = 0.$$

Now suppose $q \ll 1$. By Proposition 4.0.3, the maximum term in $\mathbf{C}_{n,p}$ is asymptotic to $(1-\epsilon)q^{-1}\log n$ and

$$\lim_{n\to\infty} \mathbb{P}\big[\max(\mathbf{C}_{n,p}) \gg q^{-1}\log n\big] = 0.$$

By Proposition 2.5.1,

$$\lim_{n\to\infty} \mathbb{P}\big[\max(\mathbf{C}_{n,m}) \gg \frac{m}{n}\log n\big] = 0$$

since $m \gg n$. □

Chapter 5

Patterns

The focus of this chapter investigates the appearance and disappearance of *composition patterns*. A composition pattern is simply a sub-composition, under some notion of *containment*. In particular, this section focuses on *exact composition patterns*, in which terms must take specified values. We then finish by determining the threshold for $\mathbf{C}_{n,p}$ to be a *Carlitz* composition (having no adjacent pair of equal terms).

There is quite an extensive literature on patterns in compositions and words. This includes comprehensive expositions by Heubach and Mansour [33] and Kitaev [37]. However, their approach is different and they do not consider exact patterns.

5.1 Exact Consecutive Patterns

The *exact consecutive pattern* ${}^{=}\overline{r_1 \ldots r_k}$ occurs at position i in a composition C if, for each $j \in [k]$, we have $C(i-1+j) = r_j$. In the language of combinatorics on words, such a pattern occurs in a composition if it is a *factor* of the composition. See Figure 5.1 for an illustration. A pattern is *nonzero* if at least one of its terms is positive.

The presence of an exact pattern is not a monotone property, for example ${}^{=}\overline{55}$ occurs in the compositions 554 and 655, but does not occur in 654. However, thresholds for a nonzero exact pattern ${}^{=}\overline{\pi}$ can be established, one for its appearance (lower threshold) and one for its disappearance (upper threshold). Rather interestingly, if ${}^{=}\overline{\pi} = \overline{r_1 \ldots r_k}$, then the lower threshold, for the appearance of ${}^{=}\overline{\pi}$, depends on its size $|{}^{=}\overline{\pi}| = \sum_{i=1}^{k} r_i$,

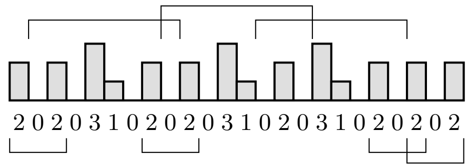

Figure 5.1: A composition containing four occurrences of the exact consecutive pattern $^{=}\overline{202}$ and three occurrences of $^{=}\overline{02031020}$

whereas the upper threshold, for the disappearance of the consecutive pattern, depends on its length, k.

Proposition 5.1.1. *If $^{=}\overline{\pi}$ is a nonzero exact consecutive pattern of length k, then for any positive constant α,*

$$\mathbb{P}\big[\mathbf{C}_{n,p} \text{ contains } {}^{=}\overline{\pi}\big] \sim \begin{cases} 0 & \text{if } p \ll n^{-1/|\pi|}, \\ 1-e^{-\alpha^{|\pi|}} & \text{if } p \sim \alpha n^{-1/|\pi|}, \\ 1 & \text{if } n^{-1/|\pi|} \ll p \text{ and } q \gg n^{-1/k}, \\ 1-e^{-\alpha^{k}} & \text{if } q \sim \alpha n^{-1/k}, \\ 0 & \text{if } n^{-1/k} \gg q. \end{cases}$$

The expected number of occurrences of $^{=}\overline{\pi}$ in $\mathbf{C}_{n,p}$ is maximal when $p/q = |\pi|/k$.

Proof. Suppose $\pi = r_1 \dots r_k$ and $|\pi| = s$. For each $i \in [n+1-k]$, let A_i be the event that $^{=}\overline{\pi}$, which is of length k and size s, occurs at position i in $\mathbf{C}_{n,p}$, and let X be the number of occurrences of $^{=}\overline{\pi}$ in $\mathbf{C}_{n,p}$. Then, $\mathbb{P}\big[A_i\big] = q^k p^s$, and $\mathbb{E}\big[X\big] \sim nq^k p^s$, which, by elementary calculus, is seen to be maximal when $p = s/(k+s)$.

We begin by handling the first and last ranges of values in the statement of the proposition.

If $p \ll n^{-1/s}$, then $\mathbb{E}\big[X\big] \sim np^s \ll 1$. Similarly, if $q \ll n^{-1/k}$, then $\mathbb{E}\big[X\big] \sim nq^k \ll 1$. Thus, by the First Moment Method, in either case, w.h.p. $^{=}\overline{\pi}$ doesn't occur in $\mathbf{C}_{n,p}$.

Distinct events A_i and A_j are correlated if $t = |j-i| < k$. If $r_\ell \neq r_{\ell+t}$ for some $\ell \in [k-t]$, then $\mathbb{P}\big[A_i \wedge A_j\big] = 0$. Otherwise, $\mathbb{P}\big[A_i \wedge A_j\big] \leq q^{k+1}p^{s+1}$, since π is nonzero.

Thus,

$$\Delta := \sum_{i \sim j} \mathbb{P}\big[A_i \wedge A_j\big] \le n \sum_{t=1}^{k-1} q^{k+1} p^{s+1} \le nkq^{k+1}p^{s+1}$$

and

$$R := \Delta / \mathbb{E}\big[X\big]^2 \lesssim \frac{nkq^{k+1}p^{s+1}}{(nq^k p^s)^2} = \frac{k}{nq^{k-1}p^{s-1}}.$$

Moreover,

$$\Lambda := \sum_{i} \mathbb{P}\big[A_i\big]^2 + \sum_{i \sim j} \mathbb{P}\big[A_i\big]\mathbb{P}\big[A_j\big] \sim nkp^{2s}q^{2k}.$$

We now consider the third case in the statement of the proposition.

Suppose $p = \omega n^{-1/s} \ll 1$ and $q \ll \omega n^{-1/k}$ for some $\omega \gg 1$. Then

$$\mathbb{E}\big[X\big] \sim nq^k p^s \sim np^s = n\left(\omega n^{-1/s}\right)^s = \omega^s \gg 1$$

and

$$R \lesssim \frac{k}{nq^{k-1}p^{s-1}} = \frac{k}{np^{s-1}} = \frac{k}{n\left(\omega n^{-1/s}\right)^{s-1}} = \frac{k}{\omega^{s-1}n^{1/s}} \ll 1.$$

Similarly, since $q = \omega n^{-1/k} \ll 1$ for some $\omega \gg 1$, then

$$\mathbb{E}\big[X\big] \sim \omega^k \gg 1$$

and

$$R \lesssim \frac{k}{\omega^{k-1}n^{1/k}} \ll 1.$$

Finally, if p is asymptotically bounded away from both 0 and 1, then $\mathbb{E}\big[X\big] \asymp n \gg 1$ and $R \asymp n^{-1} \ll 1$. Hence, by the Second Moment Method, if $n^{-1/s} \ll p$ and $q \gg n^{-1/k}$, w.h.p. ${}^{=}\pi$ occurs in $\mathbf{C}_{n,p}$.

Finally, we analyse the second and fourth cases.

Suppose $p \sim \alpha n^{-1/s}$. Then $\mathbb{E}\big[X\big] \sim \alpha^s$ and $\Delta \le \alpha^s kp \ll 1$, and $\Lambda \sim \alpha^{2s}k/n \ll 1$. So, by the Chen–Stein Method, the number of occurrences of ${}^{=}\pi$ is asymptotically Poisson with mean α^s. Similarly, if $q = \alpha n^{-1/k}$ then $\mathbb{E}\big[X\big] \sim \alpha^k$ and $\Delta \le \alpha^k kq \ll 1$, and $\Lambda \sim \alpha^{2k}k/n \ll 1$, so the number of occurrences of ${}^{=}\pi$ is asymptotically Poisson with mean α^k. □

We therefore see how the length and size of consecutive composition patterns affect their appearance and disappearance in $\mathbf{C}_{n,p}$. Smaller patterns appear before larger ones and longer patterns disappear before shorter ones.

We now introduce a result to be utilised in the subsequent proposition.

Proposition 5.1.2 (Stirling's Approximation [14])**.** *For positive n,*

$$n! \leq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{-12/n}$$

or, alternatively

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{-c/n}$$

for some positive constant c.

The following proposition enables us to transfer the thresholds for exact consecutive patterns from $\mathbf{C}_{n,p}$ to $\mathbf{C}_{n,m}$.

Proposition 5.1.3. *If $^{=}\pi$ is an exact consecutive pattern and $m \sim np/q \gg 1$, then*

$$\lim_{n\to\infty} \mathbb{P}\big[\mathbf{C}_{n,m} \text{ contains } {}^{=}\pi\big] = \lim_{n\to\infty} \mathbb{P}\big[\mathbf{C}_{n,p} \text{ contains } {}^{=}\pi\big].$$

Proof. Suppose π has length k and size s. For each $i \in [n+1-k]$, let P_i be the probability that $^{=}\pi$ occurs at position i in $\mathbf{C}_{n,m}$. Then,

$$P_i \;=\; \binom{m-s+n-k-1}{m-s}\binom{m+n-1}{m}^{-1}.$$

For brevity, let $n_1 = n-1$, $n_k = n_1 - k$ and $m_s = m - s$. Note that $n_1 \sim n_k \sim n$ and $m_s \sim m$ (as along as $k \ll n$ and $s \ll m$), and also that $p \sim m/(m+n)$ and $q \sim n/(m+n)$.

Then, by Stirling's approximation, Proposition 5.1.2,

$$
\begin{aligned}
P_i &= \binom{m_s+n_k}{m_s}\binom{m+n_1}{m}^{-1} \\
&= \left(\frac{(m_s+n_k)!}{m_s!n_k!}\right)\left(\frac{m!n_1!}{(m+n_1)!}\right) \\
&= \left(\sqrt{\frac{(m_s+n_k)\,m\,n_1}{m_s\,n_k\,(m+n_1)}}\right)\left(\frac{(m_s+n_k)^{m_s+n_k}}{m_s^{m_s}\,n_k^{n_k}}\right)\left(\frac{m^m\,n_1^{n_1}}{(m+n_1)^{m+n_1}}\right)\left(e^{(1+c)(m_s+n_k+m+n_1-m_s-n_k-m-n_1)}\right) \\
&\sim \left(\frac{(m_s+n_k)^{m_s+n_k}}{{m_s}^{m_s}{n_k}^{n_k}}\right)\left(\frac{m^m{n_1}^{n_1}}{(m+n_1)^{m+n_1}}\right) \\
&= (m_s+n_k)^{m+n_1}(m_s+n_k)^{-s-k}{m_s}^{-m_s}{n_k}^{-n_k}m^{m_s}m^s{n_1}^{n_k}{n_1}^k(m+n_1)^{-(m+n_1)} \\
&\sim \left(\frac{m_s+n_k}{m+n_1}\right)^{m+n_1}(m_s+n_k)^{-s-k}\left(\frac{m}{m_s}\right)^{m_s}m^s\left(\frac{n_1}{n_k}\right)^{n_k}n_1^k \\
&= \left(\frac{m-s+n_1-k}{m+n_1}\right)^{m+n_1}\left(\frac{m_s+s}{m_s}\right)^{m_s}\left(\frac{n_k+k}{n_k}\right)\left(\frac{m^s{n_1}^k}{(m_s+n_k)^{s+k}}\right) \\
&= \left(1-\frac{s+k}{m+n_1}\right)^{m+n_1}\left(1+\frac{s}{m_s}\right)^{m_s}\left(1+\frac{k}{n_k}\right)^{n_k}\frac{m^s\,n_1^k}{(m_s+n_k)^{s+k}} \\
&\sim e^{-s-k}\,e^s\,e^k\,p^s\,q^k \;=\; p^sq^k \sim \mathbb{P}\big[{}^{=}\overline{\pi} \text{ occurs at position } i \text{ in } \mathbf{C}_{n,p}\big].
\end{aligned}
$$

The result then follows from the fact that the probability of $\mathbf{C}_{n,p}$ or $\mathbf{C}_{n,m}$ containing an exact consecutive pattern depends only on the asymptotic probabilities of exact consecutive patterns occurring at a given position.

Specifically, the First and Second Moment Methods and the Chen–Stein Method make use only of asymptotic probabilities, which we have shown to be identical for $\mathbf{C}_{n,p}$ and $\mathbf{C}_{n,m}$, when $m \sim np/q \gg 1$. Note that a correlated pair of occurrences of an exact consecutive pattern is simply an occurrence of a larger exact consecutive pattern, so the corresponding asymptotic probabilities also match. Thus, the argument in the proof of Proposition 5.1.1 could be copied with only trivial changes to yield exactly the same asymptotic probabilities for $\mathbb{P}\big[\mathbf{C}_{n,m} \text{ contains } {}^{=}\overline{\pi}\big]$ as Proposition 5.1.1 gives for $\mathbb{P}\big[\mathbf{C}_{n,p} \text{ contains } {}^{=}\overline{\pi}\big]$. □

Thus, as the threshold for an exact composition pattern of size s appearing in $\mathbf{C}_{n,p}$ is $p = n^{-1/s}$, then the threshold of the appearance of the pattern occurring in $\mathbf{C}_{n,m}$

is $m \sim \frac{n\left(n^{-1/s}\right)}{1-n^{-1/s}} \sim n^{1-1/s}$. In general, if $\gamma \in (0,1)$, and $m \sim n^{\gamma}$, then any exact consecutive pattern of size less than $\frac{1}{1-\gamma}$ exists in $\mathbf{C}_{n,m}$ a.a.s. However, any pattern with size greater than $\frac{1}{1-\gamma}$ does not exist in $\mathbf{C}_{n,m}$ a.a.s.

Similarly, as the threshold for an exact composition pattern of length k disappearing in $\mathbf{C}_{n,p}$ is $q = n^{-1/k}$, the threshold of the disappearance of the pattern in $\mathbf{C}_{n,m}$ is $m \sim \frac{n\left(1-n^{-1/k}\right)}{n^{-1/k}} \sim n^{1+1/k}$. If $\gamma \in (1,2)$ and $m \sim n^{\gamma}$, then every exact consecutive pattern with length less than $\frac{1}{\gamma-1}$ exists in $\mathbf{C}_{n,m}$ a.a.s. However any pattern with length greater than $\frac{1}{1-\gamma}$ does not exist in $\mathbf{C}_{n,m}$ a.a.s.

These results establish when any given exact consecutive pattern is present. For example, w.h.p. the pattern ${}^{=}\overline{2718281}$ appears when $m \asymp n^{28/29}$ and has disappeared once $m \gg n^{8/7}$. If π_1 is both shorter and smaller than π_2, then ${}^{=}\overline{\pi_1}$ arrives before ${}^{=}\overline{\pi_2}$ and leaves after ${}^{=}\overline{\pi_2}$. For example, w.h.p. ${}^{=}\overline{110}$, ${}^{=}\overline{21}$, ${}^{=}\overline{4}$ and ${}^{=}\overline{2021}$ arrive in that order, but depart in the order ${}^{=}\overline{2021}$, ${}^{=}\overline{110}$, ${}^{=}\overline{21}$, ${}^{=}\overline{4}$.

We are now able to formally establish the threshold for $\mathbf{C}_{n,m}$ to contain an exact composition pattern.

Proposition 5.1.4. *If c is a non-zero exact composition pattern of length k with $\|c\| = s$, then for any positive constant a,*

$$\lim_{n\to\infty} \mathbb{P}\big[\mathbf{C}_{n,m} \text{ contains } c\big] = \begin{cases} 0 & \text{if } m \ll n^{1-1/s}, \\ 1 - e^{-a^s} & \text{if } m \sim an^{1-1/s} \text{ and } s > 1, \\ 1 & \text{if } m \sim a \text{ and } s = 1, \\ 1 & \text{if } n^{1-1/s} \ll m \ll n^{1+1/k}. \end{cases}$$

Proof. This follows from Proposition 5.1.1 and Proposition 5.1.3. If $m < s$, then $\mathbf{C}_{n,m}$ doesn't contain c. If m is bounded and $m \geq s > 1$, then

$$\mathbb{P}\big[\mathbf{C}_{n,m} \text{ contains } c\big] < n \frac{\binom{m-s+n-k-1}{m-s}}{\binom{m+n-1}{m}} \sim \frac{m!}{(m-s)!} n^{1-s} \ll 1.$$

If $m \sim a$ and $s = 1$, then a.a.s. $\mathbf{C}_{n,m}$ contains exactly a occurrences of c, this being the same as having the first few and last few terms equal to zero, and avoiding a finite number of patterns, each of size greater than one. □

We now prove position independence of exact composition patterns occurring in $\mathbf{C}_{n,m}$.

Proposition 5.1.5. *Let c be any exact composition pattern of length k. Then, for any $i, j \in [n+1-k]$,*

$$\mathbb{P}\big[c \textit{ occurs at } i \textit{ in } \mathbf{C}_{n,m}\big] \;=\; \mathbb{P}\big[c \textit{ occurs at } j \textit{ in } \mathbf{C}_{n,m}\big].$$

Proof. The probability of c appearing at position i in $\mathbf{C}_{n,m}$ is equal to

$$\binom{(m-|c|)+(n-k)-1}{m-|c|} \times \binom{m+n-1}{m}^{-1},$$

which does not depend on i. □

Therefore, for any $i, j \in [n+1-k]$, the probability of a consecutive composition pattern occurring at position i in $\mathbf{C}_{n,m}$ is equally likely to occur at position j in $\mathbf{C}_{n,m}$.

5.2 Carlitz Compositions

We finish this part of the thesis by establishing the threshold for a composition to be Carlitz, that is no adjacent pair of terms being equal. These have previously been well-studied [29, 36, 38, 39].

Proposition 5.2.1. *If $\mathcal{Q}$ is the set of Carlitz compositions, then*

$$\lim_{n\to\infty} \mathbb{P}\big[\mathbf{C}_{n,p} \in \mathcal{Q}\big] \;=\; \begin{cases} 0 & \textit{if } \; q \gg 1/n, \\ 1 & \textit{if } \; q \ll 1/n. \end{cases}$$

Proof. For each $i \in [n-1]$, let A_i be the event "$\mathbf{C}_{n,p}(i) = \mathbf{C}_{n,p}(i+1)$" and let X be the number of pairs of adjacent terms that are equal.

By summing the probabilities that a pair of adjacent terms both equal k, for each i:

$$\mathbb{P}\big[A_i\big] \;=\; \sum_{k=0}^{\infty} p^{2k}q^2 \;=\; \frac{q^2}{1-p^2} \;=\; \frac{q}{1+p}$$

Then, by linearity of expectation,

$$\mathbb{E}\big[X\big] \;=\; (n-1)\mathbb{E}\big[A_i\big] \sim nq/(1+p).$$

So, if $q \ll 1/n$, then $p \to 1$ and $\mathbb{E}\big[X\big] \to 0$, and, by the First Moment Method, with high probability no two adjacent terms of $\mathbf{C}_{n,p}$ are equal.

Distinct events A_i and A_j are correlated $(i \sim j)$ if $|i-j| = 1$. For each $i \in [n-1]$,

$$\mathbb{P}\big[A_i \wedge A_{i+1}\big] \;=\; \sum_{k=0}^{n} p^{3k}q^3 \sim \frac{q^3}{1-p^3}.$$

So, given that there are $2n-2$ ordered pairs of adjacent terms,

$$\Delta := \sum_{i \sim j} \mathbb{P}\big[A_i \wedge A_j\big] \sim \frac{2nq^3}{1-p^3}.$$

Then,

$$\frac{\Delta}{\mathbb{E}\big[X\big]^2} \sim \frac{2nq^3}{1-p^3} \bigg/ \left(\frac{nq}{1+p}\right)^2 \;=\; \frac{2}{n}\left(\frac{1+2p+p^2}{1+p+p^2}\right).$$

Thus, if $q \gg 1/n$ then $\mathbb{E}\big[X\big] \to \infty$ and $\Delta \ll \mathbb{E}\big[X\big]^2$. Therefore, by the Second Moment Method, with high probability $\mathbf{C}_{n,p}$ contains an adjacent pair of equal terms. $\square$

An interesting point to note is that we begin to stop seeing adjacent terms at the same time we begin to no longer see any gaps within the composition a.a.s. ($q \sim \frac{1}{n}$).

Chapter 6

Permutations

Our attention now turns to permutations, though particular results from the exploration into compositions are required to complete our investigation into permutations. A *permutation* or *n-permutation* is considered to be simply an arrangement of the numbers $[n] := \{1, 2, \ldots, n\}$. Let $\mathcal{S}_n$ denote the set of all n-permutations. We often display an n-permutation σ using its plot, the set of points $(i, \sigma(i))$ in the Euclidean plane, for $i = 1, \ldots, n$.

Definition 6.0.1. If σ is an n-permutation, we define its complement, denoted $\overline{\sigma}$, to be the permutation such that $\overline{\sigma}(i) = n + 1 - \sigma(i)$ for every $i \in [n]$. Thus the plot of $\overline{\sigma}$ is the reflection of σ about a horizontal axis.

See Figure 6.1 for the plots of a 9-permutation and its complement.

We consider three different forms of permutation pattern containment. For a very brief introduction to permutation patterns, see [10]; for more extended expositions, see either Bóna [15] or Kitaev [37].

The first form of patterns that we consider in this part of the thesis is *consecutive permutation patterns*. A k-permutation π occurs as a consecutive pattern at position j in a permutation σ if the consecutive subsequence $\sigma(j) \ldots \sigma(j + k - 1)$ has the same relative ordering as π. For example, the consecutive pattern $\underline{132}$ occurs twice in the permutation at the left of Figure 6.1, at positions 2, and 6. See [20–22] for investigations of consecutive permutation patterns.

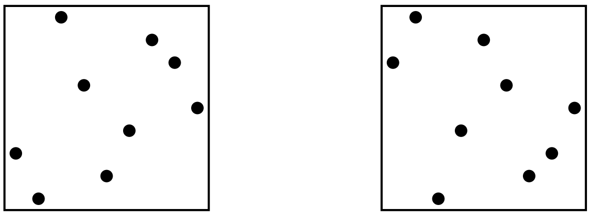

Figure 6.1: The permutation $\sigma = 319624875$ and its complement $\overline{\sigma} = 791486235$

We take a dynamic, or evolutionary, view by considering a process on n-permutations, namely a sequence of permutations $\sigma_0, \sigma_1, \sigma_2, \ldots, \sigma_{\binom{n}{2}}$, where σ_{t+1} is obtained from σ_t by the addition of one inversion (see below). Similar to that of compositions, as a permutation evolves, we see an abrupt appearance and disappearance of substructures.

6.1 Permutations and Inversion Sequences

In this section, we explore the relationship between permutations and *inversion sequences,* in particular the representation of permutations as inversion sequences. Given an n-permutation σ, its inversion sequence e_σ is the sequence of integers $(e_\sigma(j))_{j=1}^n$, where

$$e_\sigma(j) = |\{i : i < j \text{ and } \sigma(i) > \sigma(j)\}|$$

is the number of inversions involving $\sigma(j)$ and the terms of σ preceding $\sigma(j)$, or equivalently the number of points to the upper left of $(j, \sigma(j))$ in the plot of σ.

See Figure 6.2 for an example. The Figure displays the plot of the permutation 314862759 while the numbers at the bottom represent the inversion sequence of the permutation. The first point has no points to its upper left and so the first term of the inversion sequence is 0 (this is always the case for the first term of an inversion sequence). The second point has one point to its upper left and so the second term of the inversion sequence is 1 and so forth.

Clearly, for each $j \in [n]$, it is the case that $0 \leq e_\sigma(j)$ and as only $j-1$ terms have come before the j^{th} term in the permutation, then $e_\sigma(j) < j$. Every integer sequence satisfying the previous condition whose terms sum to m is the inversion sequence of an n-permutation with m inversions.

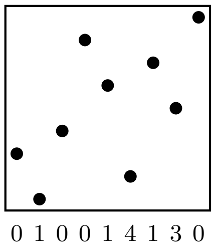

Figure 6.2: A permutation and its inversion sequence

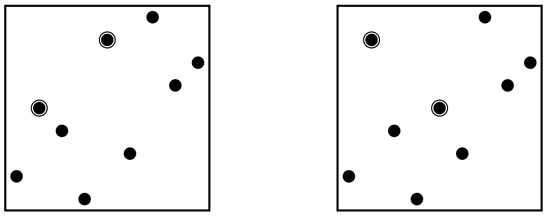

Figure 6.3: The permutations $\sigma = 254183967$ and $\sigma' = 284153967$

Definition 6.1.1. We use $\mathcal{E}_{n,m}$ to denote the set of such inversion sequences (or integer sequences) of length n whose terms sum to m and for each $j \in [n]$ it is the case that $0 \leq e_\sigma(j) \leq j-1$.

Given an inversion sequence e, if $e(i) < j-1$, then e^{+j} denotes the inversion sequence obtained from e by the addition of 1 to its j^{th} term. By increasing a term in the inversion sequence e_σ of σ by 1, values of two terms in the permutation σ, switch positions. See Figure 6.3 for an example, in which $e_\sigma = 001303022$ and $e_{\sigma'} = e_\sigma^{+5} = 001313022$.

Observation 6.1.2. *Let σ be a permutation. Suppose $e_\sigma(j) < j-1$, and that σ' is the permutation with inversion sequence e_σ^{+j}. Let $i < j$ be the index such that*

$$\sigma(i) \;=\; \max\big\{\sigma(k) \;:\; k < j \quad \textit{and} \quad \sigma(k) < \sigma(j)\big\}.$$

Then, $\sigma'(i) = \sigma(j)$ and $\sigma'(j) = \sigma(i)$, and $\sigma'(k) = \sigma(k)$ for each $k \neq i, j$.

6.2 The Uniform Random Permutation

Let $\mathcal{S}_{n,m}$ denote the set of all n-permutations with exactly m inversions. The *uniform random permutation*, denoted $\boldsymbol{\sigma}_{n,m}$, is a permutation drawn uniformly at random from

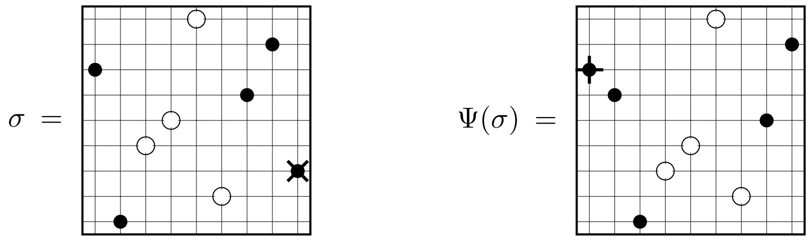

Figure 6.4: The bijection used in the proof of Proposition 6.2.1: the point marked **×** is replaced by that marked **+**; the consecutive pattern 2341 occurs at position 3 in σ and at position 4 in $\Psi(\sigma)$

$\mathcal{S}_{n,m}$. The only prior work specifically on $\boldsymbol{\sigma}_{n,m}$, of which we are aware is that of Acan and Pittel [1]. Their primary result is a determination of the (sharp) threshold at which $\boldsymbol{\sigma}_{n,m}$, becomes indecomposable at $m \sim (6/\pi^2)n \log n$.

We prove that the distribution of a consecutive permutation pattern in $\boldsymbol{\sigma}_{n,m}$ is independent of its position for any given n and m. This result allows us to only need to consider the occurrence of patterns at position 1 in $\boldsymbol{\sigma}_{n,m}$, making our proofs of following results significantly simpler. This proposition first appeared in the unpublished preprint [9].

Proposition 6.2.1. *For any consecutive permutation pattern π of length k and any $i, j \in [n+1-k]$,*

$$\mathbb{P}\big[\pi \textit{ occurs at position } i \textit{ in } \boldsymbol{\sigma}_{n,m}\big] \;=\; \mathbb{P}\big[\pi \textit{ occurs at position } j \textit{ in } \boldsymbol{\sigma}_{n,m}\big].$$

This result follows from the existence of an operation that removes the last point from a permutation and adds a new first point in such a way as to preserve the number of inversions. This operation shifts patterns rightwards.

Proof. As illustrated in Figure 6.4, let $\Psi : \mathcal{S}_{n,m} \to \mathcal{S}_{n,m}$ be defined by

$$\Psi(\sigma) \;=\; \Psi(\sigma_1\sigma_2 \ldots \sigma_n) \;=\; \sigma^\bullet \;=\; \sigma^\bullet_0\sigma^\bullet_1 \ldots \sigma^\bullet_{n-1},$$

where $\sigma^\bullet_0 = n+1-\sigma_n$, and for $1 \le i < n$,

$$\sigma^\bullet_i = \begin{cases} \sigma_i + 1 & \text{if } \sigma^\bullet_0 \le \sigma_i < \sigma_n, \\ \sigma_i - 1 & \text{if } \sigma_n < \sigma_i \le \sigma^\bullet_0, \\ \sigma_i & \text{otherwise.} \end{cases}$$

Note that σ_n contributes $n - \sigma_n$ inversions to σ, and $\sigma^\bullet_0$ contributes the same number of inversions to $\sigma^\bullet$. For $0 < i < n$, the point $\sigma^\bullet_i$ contributes the same number of inversions to $\sigma^\bullet$ as σ_i does to σ. So $\mathsf{inv}(\sigma^\bullet) = \mathsf{inv}(\sigma)$, where $\mathsf{inv}(\sigma^\bullet)$ and $\mathsf{inv}(\sigma)$ are the total number of inversions in $\sigma^\bullet$ and σ respectively. Since Ψ preserves length and has a well-defined inverse, it is a bijection on $\mathcal{S}_{n,m}$.

If π occurs at position $j \le n-k$ in σ, then π occurs at position $j+1$ in $\Psi(\sigma)$. Hence, if $1 \le i, j \le n+1-k$, then π occurs at position i in σ if and only if π occurs at position j in $\Psi^{j-i}(\sigma)$ since applying $\Psi^{j-i}(\sigma)$ shifts the pattern one space to the right a total of $j-i$ times. □

Similar to that of the previous part of this thesis, we introduce thresholds, this time for $\boldsymbol{\sigma}_{n,m}$. A function $m^\star = m^\star(n)$ is a threshold in $\boldsymbol{\sigma}_{n,m}$ for a property $\mathcal{Q}$ of permutations if

$$\lim_{n\to\infty} \mathbb{P}\big[\boldsymbol{\sigma}_{n,m} \text{ satisfies } \mathcal{Q}\big] = \begin{cases} 0 & \text{if } m \ll m^\star, \\ 1 & \text{if } m^\star \ll m \ll m^+, \end{cases}$$

for some function $m^+ \gg m^\star$. We also say that $\binom{n}{2} - m \sim m^\star$ is a threshold in $\boldsymbol{\sigma}_{n,m}$ for the *disappearance* of a property $\mathcal{Q}$ if,

$$\lim_{n\to\infty} \mathbb{P}\big[\boldsymbol{\sigma}_{n,m} \text{ satisfies } \mathcal{Q}\big] = \begin{cases} 1 & \text{if } m^\star \ll \binom{n}{2} - m \ll m^+, \\ 0 & \text{if } \binom{n}{2} - m \ll m^\star. \end{cases}$$

In this part of the thesis, we determine the thresholds for the appearance and disappearance of patterns in $\boldsymbol{\sigma}_{n,m}$, such as within our investigation into compositions. Before we are able to begin our investigation into these thresholds, we first introduce one final model.

6.3 Uniform Random Inversion Sequences

If $m \in [0, \binom{n}{2}]$, then we use $\mathbf{e}_{n,m}$ to denote an inversion sequence chosen uniformly from $\mathcal{E}_{n,m}$. We call $\mathbf{e}_{n,m}$ the *uniform random inversion sequence*. Since $\mathcal{E}_{n,m}$ and $\mathcal{S}_{n,m}$ are in bijection, then we deduce that $\mathbf{e}_{n,m}$ and $e_{\boldsymbol{\sigma}_{n,m}}$ have the same distribution.

If a consecutive permutation pattern π occurs at position 1 in a permutation σ, then e_π occurs at position 1 in e_σ. This is due to the fact that no terms occur to the left of π in σ and therefore no terms can occur to the upper-left of π in the plot of σ which therefore means that the only inversions are the pairs of terms that are inversions within π itself. On the other hand, if π occurs at position $j \neq 1$ in σ, then e_π does not necessarily occur at position j in e_σ. For example, $\underline{21}$ occurs at positions 1 and 3 in $\sigma = 4231$. Here, $e_\pi = e_{\underline{21}} = 01$ and $e_\sigma = e_{4231} = 0113$. We observe that $e_{\underline{21}}$ occurs at position 1 in e_{4231} but not at position 3.

We now give a formal proof showing that if π occurs at the start of a permutation σ, then e_π occurs at the start of e_σ.

Proposition 6.3.1. *Let π be any consecutive permutation pattern. If π occurs at position 1 in a permutation σ, then e_π occurs at position 1 in e_σ.*

Proof. If π has length k, then for each $j \in [k]$,

$$e_\sigma(j) \;=\; \big|\{i : i < j \text{ and } \sigma(i) > \sigma(j)\}\big| \;=\; \big|\{i : i < j \text{ and } \pi(i) > \pi(j)\}\big| \;=\; e_\pi(j). \quad \square$$

We now prove the following implication in the opposite direction to Proposition 6.3.1.

Proposition 6.3.2. *Let π be any consecutive permutation pattern. If σ is a permutation and e_π occurs at position j in e_σ, then π occurs at position j in σ. Moreover, if π has length k, then for all $i < j$ and $\ell = j, \ldots, j + k - 1$, we have $\sigma(i) < \sigma(\ell)$.*

Proof. We proceed by induction on the length of the pattern. If π has length 1, then $\pi = 1$ and $e_\pi = 0$. Hence, $e_\sigma(j) = 0$, so there is no point in the plot of σ to the upper left of $\sigma(j)$.

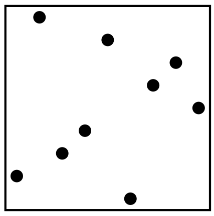

Figure 6.5: The permutation $\sigma = 293481675$

Suppose now that the proposition holds for patterns of length less than k, and that π has length k. Let π' be the permutation of length $k-1$ that results from the removal of the last point of π. If e_π occurs at position j in e_σ then $e_{\pi'}$ also occurs at position j in e_σ. So, by the induction hypothesis, π' occurs at position j in σ, with no point of σ to the upper left of any of the $k-1$ points $\sigma(j), \ldots, \sigma(j+k-2)$ that form its occurrence.

Since $e_\pi(k) < k$, at most $k-1$ points of σ are to the upper left of $\sigma(j+k-1)$, all of which must therefore be part of the occurrence of π', forming an occurrence of π at position j in σ. □

For example, let $\sigma = 293481675$ (see Figure 6.5 for its plot), so its inversion sequence is $e_\sigma = 001115224$. The consecutive permutation pattern $\pi = \underline{312}$ occurs at positions 2 and 5 in σ, whereas $e_\pi = e_{\underline{312}} = 011$ occurs at position 2 in e_σ but not at position 5. We observe that e_π occurs at position 2 as $\sigma(1) < \sigma(\ell)$ for all $\ell \in \{2, 3, 4\}$.

Propositions 6.2.1 and 6.3.1 immediately imply the following result.

Proposition 6.3.3. *For any consecutive permutation pattern π of length k and any $j \in [n+1-k]$,*

$$\begin{aligned} \mathbb{P}\big[\pi \textit{ occurs at position } j \textit{ in } \boldsymbol{\sigma}_{n,m}\big] &= \mathbb{P}\big[\pi \textit{ occurs at position } 1 \textit{ in } \boldsymbol{\sigma}_{n,m}\big] \\ &= \mathbb{P}\big[e_\pi \textit{ occurs at position } 1 \textit{ in } \mathbf{e}_{n,m}\big]. \end{aligned}$$

6.4 Compositions and Inversion Sequences

In this section, we investigate the relationship between compositions and inversion sequences. An inversion sequence e of length n with m inversions is a string of n

integers such that the terms sum to m and for each $j \in [n]$ we have $0 \leq e(j) < j$. Therefore the set of all such inversion sequences can be considered a special subset of the set of all n-compositions of m.

We now prove a few results that allow us to transfer thresholds from compositions to inversion sequences. We begin by establishing the threshold for when $\mathbf{C}_{n,m}$ is an inversion sequence. Recall that $\mathbf{C}_{n,m}$ is the uniform random composition.

Proposition 6.4.1. *The threshold for* $\mathbf{C}_{n,m}$ *to be an inversion sequence is given by*

$$\lim_{n\to\infty} \mathbb{P}\big[\mathbf{C}_{n,m} \in \mathcal{E}_{n,m}\big] \;=\; \begin{cases} 1 & \textit{if } \; m \ll n, \\ 0 & \textit{if } \; m \gg n. \end{cases}$$

Proof. We first establish the threshold for the geometric random composition $\mathbf{C}_{n,p}$ to be an inversion sequence. Recall that $C \in \mathcal{E}_{n,m}$ if $C(i) < i$ for each $i \in [n]$.

Now, $\mathbb{P}\big[\mathbf{C}_{n,p}(i) < i\big] = \sum_{k=0}^{i-1} p^k q = 1 - p^i$. So

$$\mathbb{P}\big[\mathbf{C}_{n,p} \text{ is an inversion sequence}\big] \;=\; \prod_{i=1}^{n} \big(1-p^i\big) \leq \prod_{i=1}^{\infty} \big(1-p^i\big)\,.$$

By Euler's Pentagonal Number Theorem (see [31]),

$$\prod_{i=1}^{\infty} \big(1-p^i\big) \;=\; 1 + \sum_{k=1}^{\infty} (-1)^k \Big(p^{k(3k+1)/2} + p^{k(3k-1)/2}\Big) \;=\; 1 - p - p^2 + p^5 + p^7 - \ldots.$$

If $p \ll 1$, then this converges to 1 as n tends to infinity, and so a.a.s. $\mathbf{C}_{n,p}$ is an inversion sequence.

On the other hand,

$$\mathbb{P}\big[\mathbf{C}_{n,p} \text{ is an inversion sequence}\big] \;=\; q \prod_{i=2}^{n} \big(1-p^i\big) \;\leq\; q.$$

If $q \ll 1$, then this converges to 0 as n tends to infinity, and so a.a.s. $\mathbf{C}_{n,p}$ is not an inversion sequence.

Not being an inversion sequence is an increasing property. This enables us to transfer the threshold from $\mathbf{C}_{n,p}$ to $\mathbf{C}_{n,m}$ using Proposition 2.5.1: If $\mathcal{Q}$ is an increasing property that has a threshold $p^\star \geq n^{-1}$ in $\mathbf{C}_{n,p}$, then $np^\star/q^\star$ is a threshold for $\mathcal{Q}$ in $\mathbf{C}_{n,m}$, where $q^\star = 1 - p^\star$. We can take $p^\star = q^\star = \frac{1}{2}$ to be a threshold for $\mathbf{C}_{n,p}$ to be an inversion sequence. So, $m \sim \frac{n(1/2)}{1/2} = n$ is a threshold for $\mathbf{C}_{n,m}$ to be an inversion sequence. □

In the following result we establish, under certain conditions, that if an exact composition pattern a.a.s. occurs in $\mathbf{C}_{n,m}$, then it also occurs in a suffix of $\mathbf{e}_{n,m}$ a.a.s.

Proposition 6.4.2. *Suppose c is an exact composition pattern, and that $m^- \gg 1$ and $m^+ \ll n^2/\log^2 n$ are such that a.a.s. $\mathbf{C}_{n,m}$ contains c whenever $m^- \ll m \ll m^+$. Then, a.a.s. $\mathbf{e}_{n,m}$ also contains c under the same conditions on m.*

Proof. Suppose $m \ll n^2/\log^2 n$. Then,

$$\frac{m}{n}\log n \ll \sqrt{m}\frac{n}{\log n}\frac{\log n}{n} = \sqrt{m} \ll \frac{n}{\log n} \ll n.$$

Let k satisfy $\frac{m}{n}\log n \ll k \ll \sqrt{m}$. Then, by Proposition 4.0.4, a.a.s. no term of $\mathbf{C}_{n,m}$ is greater than k.

Suppose $s \ll m$. Then $m^-(n) \ll m \ll m^+(n)$ implies $m^-(n-k) \ll m-s \ll m^+(n-k)$. So, if a.a.s. $\mathbf{C}_{n,m}$ contains c whenever $m^- \ll m \ll m^+$, then it is also the case that a.a.s. $\mathbf{C}_{n-k,m-s}$ contains c whenever $m^- \ll m \ll m^+$.

Now consider the suffix $\mathbf{e}' = \mathbf{e}_{n,m}[k+1, n]$ of $\mathbf{e}_{n,m}$. Clearly, $\mathbf{e}'(i) < k + i$ for each $i \in [n-k]$, and $m - \binom{k}{2} \leq \|\mathbf{e}'\| \leq m$, with $\binom{k}{2} \ll m$ by the definition of k.

Hence,

$$\begin{aligned}
&\text{a.a.s. } \mathbf{C}_{n,m} \text{ contains } c \text{ whenever } m^- \ll m \ll m^+ \\
\Longrightarrow\ &\text{a.a.s. } \mathbf{C}_{n-k,\|\mathbf{e}'\|} \text{ contains } c \text{ whenever } m^- \ll m \ll m^+ \\
\Longrightarrow\ &\text{a.a.s. } \mathbf{e}' \text{ contains } c \text{ whenever } m^- \ll m \ll m^+ \\
\Longrightarrow\ &\text{a.a.s. } \mathbf{e}_{n,m} \text{ contains } c \text{ whenever } m^- \ll m \ll m^+,
\end{aligned}$$

as required. □

We have now constructed the necessary framework to establish the thresholds for the appearance and disappearance of consecutive permutation patterns in $\boldsymbol{\sigma}_{n,m}$.

Theorem 6.4.3. Let π be any consecutive permutation pattern of length k. If $s = \mathsf{inv}(\pi)$ and $s' = \mathsf{inv}(\overline{\pi})$, then for any positive constant a,

$$\lim_{n\to\infty} \mathbb{P}\big[\boldsymbol{\sigma}_{n,m} \text{ contains } \pi\big] = \begin{cases} 0 & \text{if } m \ll n^{1-1/s}, \\ 1 - e^{-a^s} & \text{if } m \sim an^{1-1/s}, \\ 1 & \text{if } m \sim a \text{ and } s = 1, \\ 1 & \text{if } n^{1-1/s} \ll m \ll n^{1+1/k}, \end{cases}$$

$$\lim_{n\to\infty} \mathbb{P}\big[\boldsymbol{\sigma}_{n,m} \text{ contains } \pi\big] = \begin{cases} 1 & \text{if } n^{1+1/k} \gg \binom{n}{2} - m \gg n^{1-1/s'}, \\ 1 & \text{if } \binom{n}{2} - m \sim a \text{ and } s = 1, \\ 1 - e^{-a^{s'}} & \text{if } \binom{n}{2} - m \sim an^{1-1/s'}, \\ 0 & \text{if } \binom{n}{2} - m \ll n^{1-1/s'}, \end{cases}$$

as long as $s > 0$ and $s' > 0$, respectively.

Proof. If $m \ll n$, then by Proposition 6.4.1, a.a.s. $\mathbf{C}_{n,m}$ is an inversion sequence. So, by Propositions 6.3.3, 6.4.1 and 5.1.5, for any $i, j \in [n + 1 - |\pi|]$,

$$\begin{aligned} &\mathbb{P}\big[\pi \text{ occurs at position } j \text{ in } \boldsymbol{\sigma}_{n,m}\big] \\ &= \mathbb{P}\big[e_\pi \text{ occurs at position 1 in } \mathbf{e}_{n,m}\big] && \text{(by Proposition 6.3.3)} \\ &\sim \mathbb{P}\big[e_\pi \text{ occurs at position 1 in } \mathbf{C}_{n,m}\big] && \text{(by Proposition 6.4.1)} \\ &= \mathbb{P}\big[e_\pi \text{ occurs at position } i \text{ in } \mathbf{C}_{n,m}\big] && \text{(by Proposition 5.1.5).} \end{aligned}$$

Therefore $\mathbb{P}\big[\boldsymbol{\sigma}_{n,m} \text{ contains } \pi\big] \sim \mathbb{P}\big[\mathbf{C}_{n,m} \text{ contains } e_\pi\big]$.

From Proposition 5.1.4, if $m \ll n^{1-1/s}$ then a.a.s. $\mathbf{C}_{n,m}$ avoids e_π, and so a.a.s. $\boldsymbol{\sigma}_{n,m}$ avoids π. The same proposition also gives us the probability at the threshold.

Whenever $n^{1-1/s} \ll m \ll n^{1+1/k}$, then, by Proposition 5.1.4, a.a.s. $\mathbf{C}_{n,m}$ contains e_π. So, by Proposition 6.4.2, a.a.s. $\mathbf{e}_{n,m}$ contains e_π, and so a.a.s. $\boldsymbol{\sigma}_{n,m}$ contains π. Trivially, if $s = 1$ and $m = a \geq 1$, then $\boldsymbol{\sigma}_{n,m}$ contains π.

The threshold for the disappearance of π then follows from $\boldsymbol{\sigma}_{n,\binom{n}{2}-m} = \overline{\boldsymbol{\sigma}_{n,m}}$ since $\mathsf{inv}(\pi) + \mathsf{inv}(\overline{\pi}) = \binom{n}{2}$. □

$\asymp m$	Consecutive permutation patterns	Corresponding inversion sequences
1	$\underline{21}$, $\underline{132}$, $\underline{213}$, $\underline{1243}$, $\underline{1324}$, $\underline{2134}$	01, 001, 010, 0001, 0010, 0100
$\sqrt{n}$	$\underline{231}$, $\underline{312}$, $\underline{1342}$, $\underline{1423}$, $\underline{2143}$, $\underline{2314}$, $\underline{3124}$	002, 011, 0002, 0011, 0101, 0020, 0110
$n^{2/3}$	$\underline{321}$, $\underline{1432}$, $\underline{2341}$, $\underline{2413}$, $\underline{3142}$, $\underline{3214}$, $\underline{4123}$	012, 0012, 0003, 0021, 0102, 0120, 0111
$n^{3/4}$	$\underline{2431}$, $\underline{3241}$, $\underline{3412}$, $\underline{4132}$, $\underline{4213}$	0013, 0103, 0022, 0112, 0121
$n^{4/5}$	$\underline{3421}$, $\underline{4231}$, $\underline{4312}$	0023, 0113, 0122
$n^{5/6}$	$\underline{4321}$	0123

Table 6.1: Thresholds for the appearance in $\boldsymbol{\sigma}_{n,m}$ of short consecutive patterns

Therefore, $m \sim n^{1-1/\,\mathsf{inv}(\pi)}$ is the threshold for the appearance of a consecutive pattern π occurring in $\boldsymbol{\sigma}_{n,m}$, and $\binom{n}{2} - m \sim n^{1-1/\,\mathsf{inv}(\overline{\pi})}$ is the threshold for its disappearance from $\boldsymbol{\sigma}_{n,m}$. Thus, if $\gamma \in (0,1)$ and $m \sim n^{\gamma}$, then $\boldsymbol{\sigma}_{n,m}$ contains any consecutive permutation pattern with fewer than $\frac{1}{1-\gamma}$ inversions a.a.s. but avoids consecutive permutation patterns with more than $\frac{1}{1-\gamma}$ inversions a.a.s.

Table 6.1 displays all patterns of lengths $2, 3$ and 4 that have at least one inversion, as well as the thresholds at which they appear.

6.5 Classical and Vincular Patterns

In this concluding section of this part of the thesis, we establish two more pairs of thresholds, once more for the appearance and disappearance of permutation patterns. Here however, we are looking at two different types of permutation patterns, the first of which are *classical permutation patterns.*

We say that a classical pattern π *occurs at* $[i, j]$ *in* σ if $\sigma(i)$ is the first term and $\sigma(j)$ the last term in an occurrence of π. Such an occurrence has *width* $w = j + 1 - i$. We use $\sigma[i, j]$ to denote the permutation of $[w]$ that has the same relative order as $\sigma(i), \ldots, \sigma(j)$.

Given an n-permutation σ, we say that it is *decomposable* if there exists some $k < n$ such that

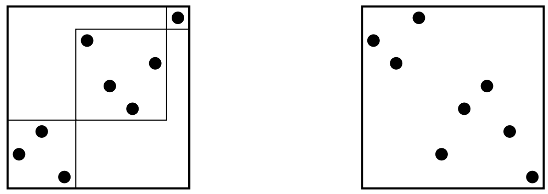

Figure 6.6: The sum decomposition of a decomposable permutation, and an indecomposable permutation

$$\{\sigma(1), \sigma(2), \ldots, \sigma(k)\} = \{1, 2, \ldots, k\}.$$

If a permutation is not decomposable, we say it is *indecomposable*. For example, Figure 6.6 displays the plot of a decomposable permutation pattern σ_1 at the left and the plot of an indecomposable permutation pattern σ_2 at the right. For the permutation $\sigma_1 = 23175468$, we can see that $\{\sigma(1), \sigma(2), \sigma(3)\} = \{1, 2, 3\}$ and $\{\sigma(4), \sigma(5), \sigma(6), \sigma(7)\} = \{4, 5, 6, 7\}$ and $\{\sigma(8)\} = \{8\}$. On the other hand, for the permutation $\sigma_2 = 76824531$, there does not exist any $k < 8$ such that

$$\{\sigma(1), \sigma(2), \ldots, \sigma(k)\} \;=\; \{1, 2, \ldots, k\}.$$

Any pattern that is decomposable can be expressed as the combination of two or more shorter permutations. Given two permutations σ and τ with lengths k and ℓ respectively, their *direct sum* $\sigma \oplus \tau$ is the permutation of length $k + \ell$ consisting of σ followed by a shifted copy of τ :

$$(\sigma \oplus \tau)(i) = \begin{cases} \sigma(i) & \text{if } \; i \leq k, \\ k + \tau(i - k) & \text{if } \; k + 1 \leq i \leq k + \ell. \end{cases}$$

For example, the permutation at the left of Figure 6.6 is $231 \oplus 4213 \oplus 1$. Every permutation has a unique representation as the direct sum of a sequence of one or more indecomposable permutations, which we call its *components*. This representation is known as its *sum decomposition*. The complement of a decomposable permutation is indecomposable. The indecomposable permutation at the right of Figure 6.6 is the complement of the decomposable permutation at the left of the figure.

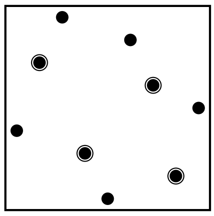

Figure 6.7: A permutation containing an indecomposable classical permutation pattern

The following two propositions consider occurrences of indecomposable classical patterns occurring within a permutation.

Proposition 6.5.1. *Suppose α is an indecomposable classical pattern of length $k \geq 2$. If α occurs at $[i, j]$ in a permutation σ, with width $w = j + 1 - i$, then*

$$\mathsf{inv}(\sigma[i, j]) \;\geq\; \mathsf{inv}(\alpha) \;+\; w \;-\; k.$$

Proof. If $i < \ell < j$ and $\sigma(\ell)$ does not lie in the occurrence of α then $\sigma(\ell)$ forms an inversion with some term in the occurrence of α. Otherwise we would have $\alpha = \beta \oplus \gamma$, with β lying to the left and below $\sigma(\ell)$ and γ lying to the right and above $\sigma(\ell)$. But α is indecomposable. Thus each of the $w - k$ terms of $\sigma[i, j]$ not in the occurrence of α contributes at least 1 to the number of inversions in $\sigma[i, j]$ and there are $\mathsf{inv}(\alpha)$ inversions with both end points in α. □

Here we give an example for the above proposition. The plot of the permutation $\sigma = 479318625$ in Figure 6.7 contains the indecomposable classical pattern $\alpha = 4231$ at $[2, 8]$ in σ. We observe that, for $\ell \in [3, 7]$, each $\sigma(\ell)$ forms at least one inversion with some term in the occurrence of α. Thus, each of the three terms of $\sigma[2, 8]$ not in the occurrence of α, contributes at least 1 to the number of inversions in $\sigma[2, 8]$. Here, α has width $w = 7$, length $k = 4$ and $\mathsf{inv}(\alpha) = 10$ so,

$$\mathsf{inv}\,(\sigma[2, 8]) = 17 \geq 13 = \mathsf{inv}(\alpha) + w - k.$$

We now prove that the containment of an indecomposable pattern implies containment of a consecutive pattern with the same number of inversions whose length is

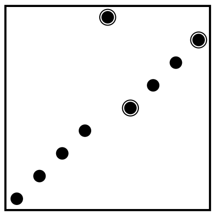

Figure 6.8: A permutation containing an indecomposable classical permutation pattern

bounded.

Proposition 6.5.2. *Suppose α is an indecomposable classical pattern of length $k \geq 2$ with s inversions. If α occurs in a permutation σ, then σ contains a consecutive pattern with at least s inversions of length at most ks.*

Proof. Suppose α occurs at $[i, j]$ in σ with width $w = j+1-i \geq k$. Let $t = \mathsf{inv}(\sigma[i, j])$. By Proposition 6.5.1, we have $t \geq s + w - k$. Note that $t \geq s \geq 1$.

Let $d = \lfloor t/s \rfloor$ and partition $e_{\sigma[i,j]}$ into d consecutive blocks of almost equal length, each block having length either $\lfloor w/d \rfloor$ or $\lceil w/d \rceil$. Since $t/d > s - 1$, by the pigeonhole principle, there is a block b, where its terms, $e_{b_1}, e_{b_2}, e_{b_3} \ldots$, are such that $\sum_{i=1} e_{b_i} = |b|$ and $|b| \geq s$.

Now, $w \leq k + t - s$, and

$$d \;=\; \left\lfloor \frac{t}{s} \right\rfloor \;\geq\; \frac{t+1}{s} - 1 \;=\; \frac{1+t-s}{s}.$$

So the length of each block is bounded above by

$$\begin{aligned} \left\lceil \frac{w}{d} \right\rceil \;<\; \frac{w}{d} + 1 \;\leq\; 1 + \frac{(k+t-s)s}{1+t-s} \\ \;=\; 1 + s + \frac{(k-1)s}{1+t-s} \;\leq\; 1 + s + (k-1)s \;=\; 1 + ks. \end{aligned}$$

Thus, since this is a strict inequality, there is a consecutive subsequence of $e_{\sigma[i,j]}$ of length no more than ks with at least s inversions, and so σ contains a consecutive pattern with at least s inversions of length at most ks. □

Here we give an example for the above. The plot of the permutation $\sigma = 123495678$

in Figure 6.8 contains the indecomposable classical pattern $\alpha = 312$ at $[5,9]$ in σ of width $w = 5$, length $k = |\alpha| = 3$ and $s = \mathsf{inv}(\alpha) = 2$. Let $t = \mathsf{inv}(\sigma[5,9]) = 4$ and $d = \left\lfloor \frac{t}{s} \right\rfloor = \left\lfloor \frac{4}{2} \right\rfloor = 2$. Therefore, we partition $e_{\sigma[5,9]} = 01111$ into $d = 2$ consecutive blocks. Each of these blocks are to have length $\left\lfloor \frac{w}{d} \right\rfloor = \left\lfloor \frac{5}{2} \right\rfloor = 2$ or length $\left\lceil \frac{w}{d} \right\rceil = \left\lceil \frac{5}{2} \right\rceil = 3$. So, we partition $e_{\sigma[5,9]}$ into two blocks, one of length two and one of length three. These blocks are therefore partitioned as either $e_{\sigma[5,6]} = 01$ and $e_{\sigma[7,9]} = 111$ or as $e_{\sigma[5,7]} = 011$ and $e_{\sigma[8,9]} = 11$. In either case, there is at least one of the partitioned blocks that sums to at least $s = 2$. We also observe that in this example there is a consecutive subsequence of $e_{\sigma[5,9]}$ which has length of no more than $ks = 6$ with at least $s = 2$ inversions and, by extension, σ contains a consecutive pattern with at least $s = 2$ inversions of length at most $ks = 6$.

We are now able to establish the thresholds for the appearance and disappearance of classical patterns in $\boldsymbol{\sigma}_{n,m}$.

Theorem 6.5.3. Let π be any classical permutation pattern. If s is the greatest number of inversions in a component of π, and s' is the greatest number of inversions in a component of $\overline{\pi}$, then for any positive constant a,

$$\lim_{n\to\infty} \mathbb{P}\big[\boldsymbol{\sigma}_{n,m} \text{ contains } \pi\big] = \begin{cases} 0 & \text{if } m \ll n^{1-1/s}, \\ 1 & \text{if } n^{1-1/s} \ll m \ll n, \end{cases}$$

$$\lim_{n\to\infty} \mathbb{P}\big[\boldsymbol{\sigma}_{n,m} \text{ contains } \pi\big] = \begin{cases} 1 & \text{if } n \gg \binom{n}{2} - m \gg n^{1-1/s'}, \\ 0 & \text{if } \binom{n}{2} - m \ll n^{1-1/s'}, \end{cases}$$

as long as $s > 0$ and $s' > 0$, respectively.

Proof. We first prove that below the threshold a.a.s. $\boldsymbol{\sigma}_{n,m}$ avoids π. Indeed, a.a.s. it contains no indecomposable pattern with s inversions.

By Proposition 6.5.2, if $\boldsymbol{\sigma}_{n,m}$ were to contain an indecomposable pattern α of length k then it would also contain some consecutive pattern of length at most ks with at least s inversions. There are only finitely many such consecutive patterns. Now suppose that $m \ll n^{1-1/s}$. From Theorem 6.4.3, we know that a.a.s. $\mathbf{C}_{n,m}$ contains no fixed finite set of consecutive patterns with s or more inversions. Thus $\mathbf{C}_{n,m}$ avoids α, and hence also avoids π.

We now prove that above the threshold a.a.s. $\boldsymbol{\sigma}_{n,m}$ contains π. Suppose π has sum decomposition $\pi = \alpha_1 \oplus \cdots \oplus \alpha_r$.

Let $\mathbf{C} = \mathbf{C}_{n,m}$. For $0 \leq j \leq r$, let $i_j = \lfloor jn/r \rfloor$, and, for each $j \in [r]$, let $\mathbf{C}_j = \mathbf{C}[i_{j-1} + 1, i_j]$. Thus, $\mathbf{C}_1, \ldots, \mathbf{C}_r$ is a partition of the terms of $\mathbf{C}$, each $\mathbf{C}_j$ having length $n_j \in \{\lfloor n/r \rfloor, \lceil n/r \rceil\}$. Let $m_j = |\mathbf{C}_j|$.

Since $|\mathbf{C}|$ is constant, the covariance between any two distinct terms of $\mathbf{C}$ is negative. Indeed, straightforward calculations show that

$$\mathrm{Var}\big[\mathbf{C}(i)\big] \;=\; \frac{(n-1)m(m+n)}{n^2(n+1)},$$

and

$$\mathsf{Cov}\big[\mathbf{C}(i_1), \mathbf{C}(i_2)\big] \;=\; -\frac{m(m+n)}{n^2(n+1)} \quad \text{if } i_1 \neq i_2.$$

Hence,

$$\mathrm{Var}\big[m_j/n_j\big] \;=\; \mathrm{Var}\big[m_j\big]/n_j^2 < n_j \,\mathrm{Var}\big[\mathbf{C}(i)\big]/n_j^2 \sim \frac{rm(m+n)}{n^3},$$

which tends to zero as long as $m \ll n^{3/2}$. Thus (by Chebyshev's inequality), for this range of values for m the sum of terms in each C_j satisfies a law of large numbers.

Thus, for each j and any $\varepsilon > 0$, a.a.s. we have $m_j > (1-\varepsilon)m/r$. Therefore, if $m \gg n^{1-1/s}$, then $m_j \gg n_j^{1-1/s}$ for each $j \in [r]$.

Thus, if $n^{1-1/s} \ll m \ll n$, for each $j \in [r]$, we have the following sequence of implications:

- By Proposition 5.1.4, a.a.s. $\mathbf{C}_{n_j,m_j}$ contains a consecutive occurrence of e_{α_j}.
- Thus a.a.s. $\mathbf{C} = \mathbf{C}_{n,m}$ contains consecutive occurrences of $e_{\alpha_1}, \ldots e_{\alpha_r}$ in that order.
- Since, by Proposition 6.4.1, $\mathbf{C}_{n,m}$ is a.a.s. an inversion sequence, a.a.s. $\mathbf{e}_{n,m}$ contains consecutive occurrences of $e_{\alpha_1}, \ldots e_{\alpha_r}$ in that order.

- By Proposition 6.3.2, these correspond to occurrences of $\alpha_1, \ldots, \alpha_r$ as consecutive patterns in $\boldsymbol{\sigma}_{n,m}$, such that no point of $\boldsymbol{\sigma}_{n,m}$ is to the upper left of any point in any of these occurrences.

- Thus a.a.s. $\pi = \alpha_1 \oplus \ldots \oplus \alpha_r$ occurs in $\boldsymbol{\sigma}_{n,m}$.

The threshold for the disappearance of π then follows as $\boldsymbol{\sigma}_{n,\binom{n}{2}-m} = \overline{\boldsymbol{\sigma}_{n,m}}$. □

Interestingly, classical patterns that are equal in length and that have the same total number of inversions do not necessarily share thresholds for their appearance and disappearance in $\boldsymbol{\sigma}_{n,m}$. Examples of this can be seen in Table 6.2.

	Pattern	
$n^{2/3}$	321654	$\binom{n}{2} - n^{8/9}$
$n^{4/5}$	423165	$\binom{n}{2} - n^{8/9}$
$n^{8/9}$	561324	$\binom{n}{2} - n^{4/5}$
$n^{8/9}$	456123	$\binom{n}{2} - n^{2/3}$

Table 6.2: Thresholds in $\boldsymbol{\sigma}_{n,m}$ for the appearance and disappearance of four classical patterns of length six with six inversions

We conclude this part of the thesis with establishing the appearance and disappearance of vincular permutation patterns in $\boldsymbol{\sigma}_{n,m}$.

In a *vincular pattern* only some terms are required to be adjacent. Consecutive terms in a vincular pattern that must be adjacent are underlined. For example, the vincular patterns $\underline{31}2$ and $3\underline{12}$ each occur once in the permutation at the left of Figure 6.1. (see [5, 8, 11, 16, 18, 19, 34])

We introduce one more definition before establishing the final pair of thresholds.

A *vincular* pattern with sum decomposition $\alpha_1 \oplus \ldots \oplus \alpha_k$, has a unique (possibly coarser) representation as a direct sum $\beta_1 \oplus \ldots \oplus \beta_\ell$ for some $\ell \leq k$, such that

- each $\beta_j = \alpha_{i_j} \oplus \alpha_{i_j+1} \oplus \ldots \oplus \alpha_{i_j+r_j}$ for some i_j and r_j, and

- α_i and α_{i+1} are components of the same β_j only if the last term of α_i is required to be adjacent to the first term of α_{i+1}.

We say that $\beta_1, \ldots, \beta_\ell$ are the pattern's *supercomponents*. For example, $231\underline{75}4\underline{68}$ has supercomponent decomposition $231 \oplus 42135$, whereas $2\underline{31754}68$ decomposes as

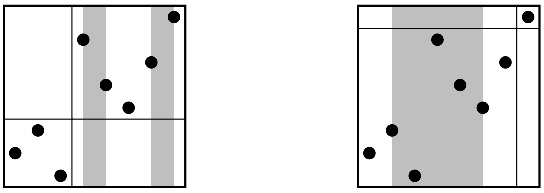

Figure 6.9: The supercomponents of vincular patterns $231\underline{75}4\underline{68}$ and $2\underline{31754}68$

$2317546 \oplus 1$. See Figure 6.9 for an illustration, in which the adjacency criteria are shown by shading.

Before we prove our final theorem, we introduce two results. These following propositions are similar to that of Propositions 6.5.1 and 6.5.2 where "indecomposable classical pattern" is replaced with "supercomponent".

Proposition 6.5.4. *Suppose α is a supercomponent of length $k \geq 2$. If α occurs at $[i, j]$ in a permutation σ, with width $w = j + 1 - i$, then*

$$\mathsf{inv}(\sigma[i,j]) \;\geq\; \mathsf{inv}(\alpha) \;+\; w \;-\; k.$$

Proof. If $i < \ell < j$ and $\sigma(\ell)$ does not lie in the occurrence of α then $\sigma(\ell)$ forms an inversion with some term in the occurrence of α. Otherwise we would have $\alpha = \beta \oplus \gamma$, with β lying to the left and below $\sigma(\ell)$ and γ lying to the right and above $\sigma(\ell)$. But α is indecomposable. Thus each of the $w - k$ terms of $\sigma[i, j]$ not in the occurrence of α contributes at least 1 to the number of inversions in $\sigma[i, j]$ and there are $\mathsf{inv}(\alpha)$ inversions with both end points in α. □

Proposition 6.5.5. *Suppose α is a supercomponent of length $k \geq 2$ with s inversions. If α occurs in a permutation σ, then σ contains a consecutive pattern with at least s inversions of length at most ks.*

Proof. Suppose α occurs at $[i, j]$ in σ with width $w = j + 1 - i \geq k$. Let $t = \mathsf{inv}(\sigma[i, j])$. By Proposition 6.5.4, we have $t \geq s + w - k$. Note that $t \geq s \geq 1$.

Let $d = \lfloor t/s \rfloor$ and partition $e_{\sigma[i,j]}$ into d consecutive blocks of almost equal length, each block having length either $\lfloor w/d \rfloor$ or $\lceil w/d \rceil$. Since $t/d > s - 1$, by the pigeonhole principle, there is a block b, where its terms, $e_{b_1}, e_{b_2}, e_{b_3} \ldots$, are such that $\sum_{i=1} e_{b_i} = |b|$

and $|b| \geq s$.

Now, $w \leq k + t - s$, and

$$d = \left\lfloor \frac{t}{s} \right\rfloor \geq \frac{t+1}{s} - 1 = \frac{1+t-s}{s}.$$

So the length of each block is bounded above by

$$\begin{aligned} \left\lceil \frac{w}{d} \right\rceil < \frac{w}{d} + 1 \leq 1 + \frac{(k+t-s)s}{1+t-s} \\ = 1 + s + \frac{(k-1)s}{1+t-s} \leq 1 + s + (k-1)s = 1 + ks. \end{aligned}$$

Thus, since this is a strict inequality, there is a consecutive subsequence of $e_{\sigma[i,j]}$ of length no more than ks with at least s inversions, and so σ contains a consecutive pattern with at least s inversions of length at most ks. □

The threshold for the appearance of a vincular pattern depends on the greatest number of inversions in one of its supercomponents. The proof for the following theorem is very similar to the proof for Theorem 6.5.3, where the two previous results are utilised and "indecomposable classical pattern" is replaced with "supercomponent".

Theorem 6.5.6. Let π be any vincular permutation pattern. If s is the greatest number of inversions in a supercomponent of π, and s' is the greatest number of inversions in a supercomponent of $\overline{\pi}$, then for any positive constant a,

$$\lim_{n\to\infty} \mathbb{P}\big[\boldsymbol{\sigma}_{n,m} \text{ contains } \pi\big] = \begin{cases} 0 & \text{if } m \ll n^{1-1/s}, \\ 1 & \text{if } n^{1-1/s} \ll m \ll n, \end{cases}$$

$$\lim_{n\to\infty} \mathbb{P}\big[\boldsymbol{\sigma}_{n,m} \text{ contains } \pi\big] = \begin{cases} 1 & \text{if } n \gg \binom{n}{2} - m \gg n^{1-1/s'}, \\ 0 & \text{if } \binom{n}{2} - m \ll n^{1-1/s'}, \end{cases}$$

as long as $s > 0$ and $s' > 0$, respectively.

Proof. We first prove that below the threshold a.a.s. $\boldsymbol{\sigma}_{n,m}$ avoids π. Indeed, a.a.s. it contains no supercomponent with s inversions.

By Proposition 6.5.5, if $\boldsymbol{\sigma}_{n,m}$ were to contain a supercomponent α of length k then it would also contain some consecutive pattern of length at most ks with at least s inversions. There are only finitely many such consecutive patterns. Now suppose that $m \ll n^{1-1/s}$. From Theorem 6.4.3, we know that a.a.s. $\mathbf{C}_{n,m}$ contains no fixed finite set of consecutive patterns with s or more inversions. Thus $\mathbf{C}_{n,m}$ avoids α, and hence also avoids π.

We now prove that above the threshold a.a.s. $\boldsymbol{\sigma}_{n,m}$ contains π. Suppose π has sum decomposition $\pi = \alpha_1 \oplus \cdots \oplus \alpha_r$.

Let $\mathbf{C} = \mathbf{C}_{n,m}$. For $0 \le j \le r$, let $i_j = \lfloor jn/r \rfloor$, and, for each $j \in [r]$, let $\mathbf{C}_j = \mathbf{C}[i_{j-1} + 1, i_j]$. Thus, $\mathbf{C}_1, \ldots, \mathbf{C}_r$ is a partition of the terms of $\mathbf{C}$, each $\mathbf{C}_j$ having length $n_j \in \{\lfloor n/r \rfloor, \lceil n/r \rceil\}$. Let $m_j = |\mathbf{C}_j|$.

Since $|\mathbf{C}|$ is constant, the covariance between any two distinct terms of $\mathbf{C}$ is negative. Indeed, straightforward calculations show that

$$\operatorname{Var}\big[\mathbf{C}(i)\big] \;=\; \frac{(n-1)m(m+n)}{n^2(n+1)},$$

and

$$\mathsf{Cov}\big[\mathbf{C}(i_1), \mathbf{C}(i_2)\big] \;=\; -\frac{m(m+n)}{n^2(n+1)} \;\text{ if } i_1 \neq i_2.$$

Hence,

$$\operatorname{Var}\big[m_j/n_j\big] \;=\; \operatorname{Var}\big[m_j\big]/n_j^2 < n_j \operatorname{Var}\big[\mathbf{C}(i)\big]/n_j^2 \sim \frac{rm(m+n)}{n^3},$$

which tends to zero as long as $m \ll n^{3/2}$. Thus (by Chebyshev's inequality), for this range of values for m the sum of terms in each C_j satisfies a law of large numbers.

Thus, for each j and any $\varepsilon > 0$, a.a.s. we have $m_j > (1-\varepsilon)m/r$. Therefore, if $m \gg n^{1-1/s}$, then $m_j \gg n_j^{1-1/s}$ for each $j \in [r]$.

Thus, if $n^{1-1/s} \ll m \ll n$, for each $j \in [r]$, we have the following sequence of implications:

- By Proposition 5.1.4, a.a.s. $\mathbf{C}_{n_j,m_j}$ contains a consecutive occurrence of e_{α_j}.
- Thus a.a.s. $\mathbf{C} = \mathbf{C}_{n,m}$ contains consecutive occurrences of $e_{\alpha_1}, \ldots e_{\alpha_r}$ in that

order.

- Since, by Proposition 6.4.1, $\mathbf{C}_{n,m}$ is a.a.s. an inversion sequence, a.a.s. $\mathbf{e}_{n,m}$ contains consecutive occurrences of $e_{\alpha_1}, \ldots e_{\alpha_r}$ in that order.

- By Proposition 6.3.2, these correspond to occurrences of $\alpha_1, \ldots, \alpha_r$ as consecutive patterns in $\boldsymbol{\sigma}_{n,m}$, such that no point of $\boldsymbol{\sigma}_{n,m}$ is to the upper left of any point in any of these occurrences.

- Thus a.a.s. $\pi = \alpha_1 \oplus \ldots \oplus \alpha_r$ occurs in $\boldsymbol{\sigma}_{n,m}$.

The threshold for the disappearance of π then follows as $\boldsymbol{\sigma}_{n,\binom{n}{2}-m} = \overline{\boldsymbol{\sigma}_{n,m}}$. $\square$

Bibliography

[1] Hüseyin Acan and Boris Pittel. On the connected components of a random permutation graph with a given number of edges. *J. Combin. Theory Ser. A*, 120(8):1947–1975, 2013.

[2] Noga Alon and Joel H. Spencer. *The Probabilistic Method.* John Wiley & Sons, fourth edition, 2016.

[3] R. Arratia, L. Goldstein, and L. Gordon. Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.*, 17(1):9–25, 1989.

[4] Richard Arratia, Larry Goldstein, and Louis Gordon. Poisson approximation and the Chen-Stein method. *Statist. Sci.*, 5(4):403–434, 1990.

[5] Eric Babson and Einar Steingrímsson. Generalized permutation patterns and a classification of the Mahonian statistics. *Sém. Lothar. Combin.*, 44: Art. B44b, 18 pp., 2000.

[6] A. D. Barbour and G. K. Eagleson. Poisson convergence for dissociated statistics. *J. Roy. Statist. Soc. Ser. B*, 46(3):397–402, 1984.

[7] A. D. Barbour and G. K. Eagleson. An improved Poisson limit theorem for sums of dissociated random variables. *J. Appl. Probab.*, 24(3):586–599, 1987.

[8] Andrew M. Baxter and Lara K. Pudwell. Enumeration schemes for vincular patterns. *Discrete Math.*, 312(10):1699–1712, 2012.

[9] David Bevan. Permutations with few inversions are locally uniform. *Preprint.* arXiv:1908.07277.

[10] David Bevan. Permutation patterns: basic definitions and notation. arXiv:1506.06673, 2015.

[11] David Bevan, Gi-Sang Cheon, and Sergey Kitaev. On naturally labelled posets and permutations avoiding 12–34. arXiv:2311.08023, 2023.

[12] David Bevan and Dan Threlfall. On the evolution of random integer compositions. *Preprint.* arXiv:2309.06287.

[13] David Bevan and Dan Threlfall. Thresholds for patterns in random permutations with a given number of inversions. *Electron. J. Combin.*, 31(4): P4.6, 2024.

[14] Gaurav Bhatnagar and Krishnan Rajkumar. Telescoping continued fractions for the error term in stirling's formula. 293, 2023.

[15] Miklós Bóna. *Combinatorics of Permutations.* CRC Press, second edition, 2012.

[16] Mireille Bousquet-Mélou, Anders Claesson, Mark Dukes, and Sergey Kitaev. $(2+2)$-free posets, ascent sequences and pattern avoiding permutations. *J. Combin. Theory Ser. A*, 117(7):884–909, 2010.

[17] Louis H. Y. Chen. Poisson approximation for dependent trials. *Ann. Probability*, 3(3):534–545, 1975.

[18] Anders Claesson. Generalized pattern avoidance. *European J. Combin.*, 22(7):961–971, 2001.

[19] Sergi Elizalde. Asymptotic enumeration of permutations avoiding generalized patterns. *Adv. in Appl. Math.*, 36(2):138–155, 2006.

[20] Sergi Elizalde. A survey of consecutive patterns in permutations. In *Recent Trends in Combinatorics*, pages 601–618. Springer, 2016.

[21] Sergi Elizalde and Marc Noy. Consecutive patterns in permutations. *Adv. in Appl. Math.*, 30(1–2):110–125, 2003.

[22] Sergi Elizalde and Marc Noy. Clusters, generating functions and asymptotics for consecutive patterns in permutations. *Adv. in Appl. Math.*, 49(3–5):351–374, 2012.

[23] P. Erdős and P. Révész. On the length of the longest head-run. In *Topics in information theory (Second Colloq., Keszthely, 1975)*, Colloq. Math. Soc. János Bolyai, Vol. 16, pages 219–228. North-Holland, 1977.

[24] P. Erdős and A. Rényi. On random graphs I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[25] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Kőzl.*, 5:17–61, 1960.

[26] Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics.* Cambridge University Press, 2009.

[27] Alan Frieze and Michał Karoński. *Introduction to Random Graphs.* Cambridge University Press, 2015.

[28] E. N. Gilbert. Random graphs. *Ann. Math. Statist.*, 30(4):1141–1144, 1959.

[29] William M. Y. Goh and Pawel Hitczenko. Average number of distinct part sizes in a random Carlitz composition. *European J. Combin.*, 23(6):647–657, 2002.

[30] Louis Gordon, Mark F. Schilling, and Michael S. Waterman. An extreme value theory for long head runs. *Probab. Theory Relat. Fields*, 72(2):279–287, 1986.

[31] Ian P. Goulden and David M. Jackson. *Combinatorial Enumeration.* Dover Publications Inc., 2004.

[32] L. J. Guibas and A. M. Odlyzko. Long repetitive patterns in random sequences. *Z. Wahrsch. Verw. Gebiete*, 53(3):241–262, 1980.

[33] Silvia Heubach and Toufik Mansour. *Combinatorics of compositions and words.* CRC Press, 2010.

[34] Lisa Hofer. A central limit theorem for vincular permutation patterns. *Discrete Math. Theor. Comput. Sci.*, 19(2):Paper No. 9, 26, 2018.

[35] Svante Janson. Coupling and Poisson approximation. *Acta Appl. Math.*, 34(1-2):7–15, 1994.

[36] Boris L. Kheyfets. The number of part sizes of a given multiplicity in a random Carlitz composition. *Adv. in Appl. Math.*, 35(3):335–354, 2005.

[37] Sergey Kitaev. *Patterns in Permutations and Words.* Springer, 2011.

[38] Arnold Knopfmacher and Helmut Prodinger. On Carlitz compositions. *European J. Combin.*, 19(5):579–589, 1998.

[39] Guy Louchard and Helmut Prodinger. Probabilistic analysis of Carlitz compositions. *Discrete Math. Theor. Comput. Sci.*, 5(1):71–95, 2002.