On the evolution of random integer compositions

David Bevan   Dan Threlfall

Abstract

We explore how the asymptotic structure of a random n-term weak integer composition of m evolves, as m increases from zero. The primary focus is on establishing thresholds for the appearance and disappearance of substructures. These include the longest and shortest runs of zero terms or of nonzero terms, longest increasing runs, longest runs of equal terms, largest squares (runs of k terms each equal to k), as well as a wide variety of other patterns. Of particular note is the dichotomy between the appearance and disappearance of exact consecutive patterns, with *smaller* patterns appearing before *larger* ones, whereas *longer* patterns disappear before *shorter* ones.

Mathematics Subject Classifications: 60C05, 05A05, 05C80

1 Introduction

We initiate the study of integer compositions from an evolutionary perspective, in an analogous manner to the Gilbert–Erdős–Rényi random graph [13, 14, 20]. Our two primary 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 with parameter $q = 1 - p$; that is, $\mathbb{P}\big[\mathbf{C}_{n,p}(i) = k\big] = qp^k$ for each $k \geqslant 0$ and $i \in [n]$. We are interested in how, for large n, the structure of $\mathbf{C}_{n,m}$ or $\mathbf{C}_{n,p}$ evolves as m or p increases from zero. The primary emphasis is on the establishment of thresholds for the appearance and disappearance of small substructures.

A selection of our results is presented in evolutionary order in Table 1 on page 2. Here we list a few highlights. A property of compositions holds *with high probability* (*w.h.p.*) if asymptotically the probability of it holding tends to one. We defer the formal presentation of this and other definitions and notational conventions until later.

1. A *gap* is a maximal run of zero terms. As long as $m \ll n^{1/2}$, with high probability the length of the *shortest* gap exceeds any fixed value. However, as soon as $m \gg n^{1/2}$, w.h.p. there are gaps of length 1. (Section 3.2)

Department of Mathematics and Statistics, University of Strathclyde, Glasgow, Scotland (`david.bevan@strath.ac.uk`, `daniel.threlfall@strath.ac.uk`).

m		Prop.
$n^{1/4}$	shortest gap has length $\leqslant \sqrt{n}$	19
$n^{1/2}/\sqrt{\log n}$	shortest gap has length $\leqslant \log n$	19
$n^{1/2}$	longest component has length 2	14
	shortest gap has length 1	18
$n^{1/2}\log n$	longest gap has length $\leqslant \sqrt{n}$	17
$n^{2/3}$	largest term equals 3	27
	increasing run of length 3	36
	exact nonconsecutive patterns with largest term 3	41
	increasing subsequence of length 4	45
$n^{5/6}$	exact consecutive patterns of size 6	20
	increasing run of length 4	36
$\gg n^{1-\varepsilon}$	any exact consecutive pattern	20
$n/\log n$	longest component has length $\geqslant \log n/\log\log n$	15
n	longest component and longest gap have length $\log n$	16
	largest term $\geqslant \log n$	29
	each total ordering of consecutive terms equally likely	35
$n\log n/\log\log n$	largest possible square pattern, length $\geqslant \log n/\log\log n$	24
$n\log n$	every gap has length $\leqslant \log n/\log\log n$	15
$n^{4/3}$	longest gap has length 2	14
	no given exact consecutive pattern of length 4	20
$n^{3/2}/\log n$	longest component has length $\geqslant \sqrt{n}$	17
	largest term $\geqslant \sqrt{n}$	30
$n^{3/2}$	every gap has length 1	14
	length of shortest component exceeds any fixed value	18
	no run of 3 equal terms	40
$n^{3/2}\sqrt{\log n}$	every component has length $\geqslant \log n$	19
$n^{7/4}$	every component has length $\geqslant \sqrt{n}$	19
$n^2/\log n$	largest term $\geqslant n$	30
n^2	no gaps (single component)	14
	smallest term exceeds any fixed value	31
	no balanced peaks or valleys	39
	no adjacent equal terms (Carlitz composition)	40
	no given nonconsecutive exact pattern	41
$n^{5/2}$	no three terms equal	46
n^3	every term $\geqslant n$	32
	every term distinct	46

Table 1: Some thresholds encountered during the evolution of $\mathbf{C}_{n,m}$

2. In contrast, w.h.p. the length of the *longest* gap exceeds any fixed value as long as $m \ll n^{1+\delta}$ for all $\delta > 0$. But once $m \gg n^{1+1/k}$, w.h.p. $\mathbf{C}_{n,m}$ has no gap of length k or greater. In particular, once $m \gg n^{3/2}$, w.h.p. every gap has length 1, and when $m \gg n^2$, w.h.p. $\mathbf{C}_{n,m}$ no longer has any gaps. (Section 3.1)

3. There is a dichotomy between the thresholds for the arrival and for the departure of *exact consecutive patterns* as $\mathbf{C}_{n,m}$ evolves, the former being ordered by *size, smaller* patterns appearing before *larger* ones, and the latter being ordered by *length, longer* patterns disappearing before *shorter* ones. (Proposition 20)

4. If $n^{1-\delta} \ll m \ll n^{1+\delta}$ for every $\delta > 0$, then w.h.p. any given exact consecutive pattern occurs in $\mathbf{C}_{n,m}$. In contrast, once $m \gg n^2$, then w.h.p. any specific exact consecutive pattern is absent from $\mathbf{C}_{n,m}$. (Proposition 20)

5. The largest *square* (run of k terms each equal to k) that we expect to see in the evolution of $\mathbf{C}_{n,m}$ has side length a little greater than $\log n / \log \log n$, and is seen when $m \sim n \log n / \log \log n$. (Proposition 24)

6. As long as $m \ll n$, w.h.p. the *largest term* in $\mathbf{C}_{n,m}$ exhibits *two-point concentration* (that is, it takes one of only two possible values). (Proposition 28)

7. As long as $m \ll n^2$, w.h.p. the *smallest term* in $\mathbf{C}_{n,m}$ equals zero, but $m \sim n^2$ is the threshold for the smallest term to exceed *any* fixed positive value. (Proposition 31)

8. Once $m \gg n$, the relative ordering of any k consecutive terms of $\mathbf{C}_{n,m}$ is asymptotically uniformly distributed over the $k!$ permutations of length k. (Proposition 35)

9. A pattern specifying the relative ordering of k consecutive terms which has a repeated term and has largest term equal to r exhibits a threshold for its disappearance from $\mathbf{C}_{n,m}$ at $m \sim n^{1+1/d}$, where $d = k - 1 - r$. (Proposition 39)

10. Once $m \gg n^2$, w.h.p. $\mathbf{C}_{n,m}$ has no adjacent equal terms (that is, it is a *Carlitz* composition). (Proposition 40)

11. The threshold for the appearance of a *nonconsecutive exact pattern* depends only on its maximum term. In contrast, all such patterns share the same threshold $m \sim n^2$ for their disappearance. (Proposition 41)

12. The threshold for the appearance of a *vincular* pattern depends on the *size* of its *largest* block. In contrast, the threshold for the *disappearance* of such a pattern depends on the *length* of its *longest* block. (Proposition 44)

13. Once $m \gg n^3$, w.h.p. every term of $\mathbf{C}_{n,m}$ is distinct. (Proposition 46)

The random composition doesn't exhibit a spectacular phase transition like that seen in the random graph with the appearance of the giant component. Perhaps the most dramatic change occurs when $m \sim n^2$ with the disappearance of both the last gap and also of the last pair of adjacent equal terms as the smallest term jumps from zero to exceed any fixed value.

There are many questions that we don't address in this introductory paper. One example would be to determine thresholds for the presence of components or gaps of

equal length. Also, we only determine threshold probabilities for certain properties. It would certainly be interesting to establish them for arbitrary nonconsecutive ordering patterns, which we don't investigate in detail. We also raise a general question concerning our models: Does every nontrivial monotone property of compositions have a threshold? (Question 7)
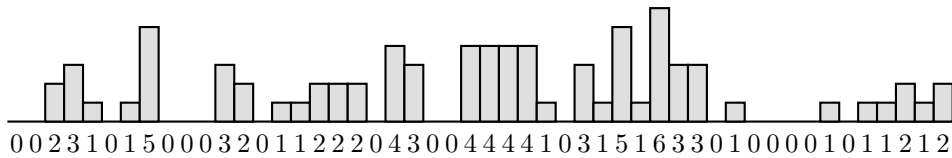
In another direction, in a subsequent paper we build on this work to determine thresholds for the appearance of patterns in the evolution of random *permutations* [7], in a similar manner to the work of Acan and Pittel [1]. A further possibility would be to extend our models with additional structure in order to represent random *multigraphs*, each term in a composition recording the number of edges between a pair of vertices (see [10, Theorems 2.9 and 2.10]). For example, the threshold for a random multigraph to cease being simple and the threshold for the presence of a complete spanning subgraph follow directly from results in this paper.

The evolution of the random *graph* is now a classical topic, with three graduate textbooks available [9, 18, 33]. In contrast, rather surprisingly, other combinatorial objects do not appear to have been investigated from this perspective, with the notable exception of the work of Acan and Pittel [1] on the threshold for connectivity in the evolution of a random permutation. Existing work on the structure of large compositions either has an enumerative flavour (determining generating functions), or else it takes a probabilistic perspective, either considering the uniform distribution over all (non-weak) compositions of n or investigating sequences of geometrically distributed random variables. For the most part our approach is somewhat orthogonal to both of these. Specific intersecting works are cited below in the relevant sections.

In the next section, we define our random models and explore their relationships, and introduce the notions of a property and a threshold in these models. We also present three key tools: the First Moment Method and the Second Moment Method which we use to determine the location of thresholds, and the Chen–Stein Method which gives us the probability of a property holding at its threshold. In Section 3, our focus is on *components* (maximal runs of nonzero terms) and *gaps* (maximal runs of zero terms), establishing thresholds for the length of the longest or shortest component to exceed some value, with analogous results for the length of gaps. Finally, Section 4 concerns the appearance and disappearance of a wide variety of types of pattern, including *exact* patterns (in which terms must take specified values), *upper* and *lower* patterns (in which terms are bounded below or above), and *ordering* patterns (which specify the relative ordering of terms).

If f and g are positive functions of n, then we use the following notation:

$$f \lesssim g \quad \text{if} \quad \limsup_{n \to \infty} f/g < \infty,$$

$$f \asymp g \quad \text{if} \quad 0 < \liminf_{n \to \infty} f/g \quad \text{and} \quad \limsup_{n \to \infty} f/g < \infty,$$

$$f \sim g \quad \text{if} \quad \lim_{n \to \infty} f/g = 1,$$

$$f \sim 0 \quad \text{if} \quad \lim_{n \to \infty} f = 0,$$

$$f \ll g \quad \text{or} \quad g \gg f \quad \text{if} \quad \lim_{n \to \infty} f/g = 0.$$

0 0 2 3 1 0 1 5 0 0 0 3 2 0 1 1 2 2 2 0 4 3 0 0 4 4 4 4 1 0 3 1 5 1 6 3 3 0 1 0 0 0 0 1 0 1 1 2 1 2

Figure 1: Bar-chart representation of a 50-term composition of 80

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 avoid the use of "big O" notation, preferring something more intuitive. We have the following equivalences:

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

Note also that $f \sim 0$ is nonstandard; however, it significantly simplifies the presentation of our results.

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 1 for an example. Alternatively, we can consider such a composition to consist of a sequence of n *boxes*, such that for each $i \in [n] := \{1, 2, \ldots, n\}$, box i contains c_i *balls*. By a simple "stars and bars" argument, it can be seen that the number of distinct n-compositions of m is $\binom{m+n-1}{m}$.

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.

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. In statistical quantum mechanics, this is known as *Bose–Einstein statistics*, modelling the distribution followed by *bosons*. See [32, Example 12.2] for an exposition in the context of a variety of different balls-in-boxes models, and [30] for work more closely related to our concerns.

2.2 The evolutionary random composition $\mathbf{C}_t$

An alternative, evolutionary, perspective comes from taking a dynamic view and considering a process on compositions, namely an infinite sequence of compositions,

$$0^n, \; C_1, \; C_2, \; C_3, \; \ldots,$$

where 0^n denotes the *empty n-composition* $(0, \ldots, 0)$, 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).

The *evolutionary random composition* $(\mathbf{C}_t)_{t \geqslant 0}$ is the Markov chain satisfying $\mathbf{C}_0 = 0^n$ and, for each $t \geqslant 0$ 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:

Proposition 1. *For each $t \geqslant 0$, 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,

$$\mathbb{P}\big[\mathbf{C}_{t+1} = C\big] \;=\; \sum_{j \in [n], \, C(j) \neq 0} \mathbb{P}\big[\mathbf{C}_t = C^{-j}\big] \frac{C(j)}{n + t}$$

$$= \binom{n+t-1}{t}^{-1} \sum_{j \in [n]} \frac{C(j)}{n+t} \;=\; \frac{t! \, (n-1)!}{(n+t-1)!} \frac{t+1}{n+t} \;=\; \binom{n+t}{t+1}^{-1}. \quad \square$$

The evolutionary random composition corresponds to the following multicoloured Pólya urn model: Consider an urn initially containing one ball of each of n different colours. Balls are drawn at random one at a time, and after drawing the ball, it is replaced together with another of the same colour. Let Y_i be the number of balls of colour i in the urn after t draws. Then, $(Y_1 - 1, Y_2 - 1, \ldots, Y_n - 1)$ has the same distribution as $\mathbf{C}_t$. See [29] and [32, Example 12.4].

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, $\mathbb{P}\big[C(i) = k\big] = q p^k$ for each $k \geqslant 0$ and $i \in [n]$. Note that $\mathbf{C}_{n,p}$ is not defined for $p = 1$.

To avoid unnecessary repetition, when considering $\mathbf{C}_{n,p}$ in this work, 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 $\mu_{n,p} \sim np$, and if $q \ll 1$ then $\mu_{n,p} \sim n/q$.

The size $|\mathbf{C}_{n,p}|$ has variance np/q^2, and so exhibits a concentrated distribution as long as $p \gg n^{-1}$. In particular, by Chebyshev's inequality, we have the following:

Observation 2. $\mathbb{P}\big[\big||\mathbf{C}_{n,p}| - np/q\big| \geqslant \alpha\sqrt{np}/q\big] \leqslant \alpha^{-2}$.

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

Proposition 3. *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. If $C_1, C_2 \in \mathcal{C}_{n,m}$, then $\mathbb{P}\big[\mathbf{C}_{n,p} = C_1\big] = p^m q^n = \mathbb{P}\big[\mathbf{C}_{n,p} = C_2\big]$. $\square$

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, $\mathbf{C}_{n,p}$ is more amenable to analysis than $\mathbf{C}_{n,m}$, so we prefer to work with $\mathbf{C}_{n,p}$ and then transfer the results to $\mathbf{C}_{n,m}$ (see Propositions 6, 8 and 21 below).

2.4 Properties

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 $C + C' \in \mathcal{Q}$ for any $C' \in \mathcal{C}_n$, 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, whereas the set of n-compositions with at least one term equal to three is not monotone. Both $\mathbf{C}_{n,m}$ and $\mathbf{C}_{n,p}$ behave monotonically with respect to monotone properties:

Proposition 4. *If $\mathcal{Q}$ is an increasing property and $m_1 < m_2$, then $\mathbb{P}\big[\mathbf{C}_{n,m_1} \in \mathcal{Q}\big] \leqslant \mathbb{P}\big[\mathbf{C}_{n,m_2} \in \mathcal{Q}\big]$.*

Proof. $\mathbb{P}\big[\mathbf{C}_{t+1} \in \mathcal{Q}\big] \geqslant \mathbb{P}\big[\mathbf{C}_{t+1} \in \mathcal{Q} \wedge \mathbf{C}_t \in \mathcal{Q}\big] = \mathbb{P}\big[\mathbf{C}_t \in \mathcal{Q}\big]$, since $\mathcal{Q}$ is increasing. $\square$

Proposition 5. *If $\mathcal{Q}$ is an increasing property and $p_1 < p_2$, then $\mathbb{P}\big[\mathbf{C}_{n,p_1} \in \mathcal{Q}\big] \leqslant \mathbb{P}\big[\mathbf{C}_{n,p_2} \in \mathcal{Q}\big]$.*

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 $i \in [n]$,

$$
\mathbb{P}\big[\mathbf{C}_{n,p_1,p_2}(i) = k\big] = \begin{cases} \frac{q_2}{q_1}, & \text{if } k = 0, \\[2ex] \frac{q_2}{q_1}\big(1 - \frac{p_1}{p_2}\big)p_2^{\,k}, & \text{if } k \geqslant 1. \end{cases}
$$

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

$$
f_p(x) = \sum_{k \geqslant 0} \mathbb{P}\big[\mathbf{C}_{n,p}(i) = k\big] x^k = \frac{q}{1 - px},
$$

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

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 $\mathbf{C}_{n,p_2}$,

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

since $\mathcal{Q}$ is increasing. $\qquad\square$

A property $\mathcal{Q}$ is *convex* if $C \in \mathcal{Q}$ and $C + C_1 + C_2 \in \mathcal{Q}$ implies $C + C_1 \in \mathcal{Q}$. For example, the property of having exactly one zero term is convex. Every convex property is the intersection of an increasing property and a decreasing property.

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

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 $m \sim np/q$, 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 6. *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 $\mathbb{P}\big[\mathbf{C}_{n,p} \in Q\big] \sim \alpha$ for all p for which p/q differs from p_0/q_0 by no more than δ. Then $\mathbb{P}\big[\mathbf{C}_{n,m_0} \in Q\big] \sim \alpha$, 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 $\varepsilon > 0$ and suppose n is sufficiently large that both $\mathbb{P}\big[|\mathbf{C}_{n,p^-}| > m_0\big] \leqslant \varepsilon$ and $\mathbb{P}\big[|\mathbf{C}_{n,p^+}| < m_0\big] \leqslant \varepsilon$. This is possible by Observation 2 given that

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

Then,

$$
\begin{aligned}
\mathbb{P}\big[\mathbf{C}_{n,p^-} \in \mathcal{Q}\big] &= \sum_{k \leqslant m_0} \mathbb{P}\big[\mathbf{C}_{n,k} \in \mathcal{Q}\big] \mathbb{P}\big[|\mathbf{C}_{n,p^-}| = k\big] + \sum_{k > m_0} \mathbb{P}\big[\mathbf{C}_{n,k} \in \mathcal{Q}\big] \mathbb{P}\big[|\mathbf{C}_{n,p^-}| = k\big] \\
&\leqslant \mathbb{P}\big[\mathbf{C}_{n,m_0} \in \mathcal{Q}\big] \mathbb{P}\big[|\mathbf{C}_{n,p^-}| \leqslant m_0\big] + \mathbb{P}\big[|\mathbf{C}_{n,p^-}| > m_0\big] \\
&\leqslant \mathbb{P}\big[\mathbf{C}_{n,m_0} \in \mathcal{Q}\big] + \varepsilon,
\end{aligned}
$$

by Proposition 4, since $\mathcal{Q}$ is increasing. Thus, $\lim_{n \to \infty} \mathbb{P}\big[\mathbf{C}_{n,m_0} \in Q\big] \geqslant \alpha$.

Similarly,

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

and so $\lim_{n \to \infty} \mathbb{P}\big[\mathbf{C}_{n,m_0} \in Q\big] \leqslant \alpha$. $\qquad\square$

Note that, in general, we can't remove the requirement that the property must be increasing. For example, if $\mathcal{Q}$ is the set of n-compositions with no zero terms whose size is not a perfect square, then $\mathbb{P}\big[\mathbf{C}_{n,m} \in \mathcal{Q}\big] = 0$ whenever m is a perfect square, 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 21 enables us to do this for exact consecutive patterns.

2.5 Thresholds

One of the most striking observations regarding the evolution of large random combinatorial objects is the abrupt nature of the appearance and disappearance of many properties. 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

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

Note here that $p_1/q_1 \ll p_2/q_2$ if and only if either $p_1 \ll p_2$ (and $q_1 \sim 1$) or else $q_1 \gg q_2$ (and $p_2 \sim 1$).

That is, a property asymptotically almost never holds below its threshold, but holds asymptotically almost surely above it. *A priori*, there is no reason why a property should have a threshold. Nevertheless, it turns out that thresholds exist for every nontrivial monotone property of the subsets of a set, and hence also of graphs (see [8, 17]). For a recent short exposition, see [43]. Note, however, that this result does not encompass our models.

Question 7. Does every nontrivial monotone property of compositions have a threshold?

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 $\varepsilon > 0$,

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

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

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.

A consequence of Proposition 6 is that a threshold in $\mathbf{C}_{n,p}$ can be transferred to one in $\mathbf{C}_{n,m}$:

Proposition 8. *Let $\mathcal{Q}$ be an increasing property that has a threshold $p^\star \geqslant n^{-1}$ 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 6, $\mathcal{Q}$ 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 6, $\mathcal{Q}$ also asymptotically almost never holds in $\mathbf{C}_{n,m}$. $\qquad \square$

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 9 (First Moment Method). *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, Section 4.3]. 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 10 (Second Moment Method). *Suppose, for each $n \geqslant 1$, 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 $\Delta = \sum\limits_{i \sim j} \mathbb{P}\big[A_i \wedge A_j\big]$, 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 [11]. 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 [5, 6], which is sufficient for our purposes. The following is adapted from [31, Theorem 4].

Proposition 11 (Chen–Stein Method). *Suppose, for each $n \geqslant 1$, 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*

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

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 $e^{-\lambda}$.

3 Components and gaps

In this and the subsequent section we investigate how the structure of the random composition evolves as its size increases. Initially, as long as $p \ll n^{-1}$, the expected number

of nonzero terms in $\mathbf{C}_{n,p}$ equals $np \ll 1$, so, by the First Moment Method, w.h.p. every term is zero and $\mathbf{C}_{n,p}$ is the empty n-permutation 0^n. We are interested in what happens after this.

Our focus in this current section is on components and gaps. 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 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, where p is the probability of a head. For constant p, this has been a topic of study for many years. In particular, the length of the longest run of heads has been investigated in considerable detail [12, 24, 22] (see also [16, pages 308–312]), of particular interest being the tiny fluctuations in its distribution that depend on the fractional part of $\log_2 n$. The asymptotic Gaussian distribution of the number of maximal runs of a fixed length (when p is constant) is established in [42]. Finally, from a statistical mechanics perspective, Huillet [30] investigates the length of both the shortest and longest component in both the constant p regime and also when $p \asymp n^{-1/k}$ for fixed $k \in \mathbb{N}$.

Components and gaps are dual in $\mathbf{C}_{n,p}$, in the sense that any statement about components can be transformed into one about gaps simply by switching the roles of p (the probability that a term is nonzero) and q (the probability that a term is zero). Results concerning gaps thus follow directly from those concerning components. Note however, that there is an asymmetry between $\mathbf{C}_{n,p}$ with small values of p and $\mathbf{C}_{n,p}$ with small values of q. Specifically, if $p \ll n^{-1}$, then w.h.p. there is literally nothing to see, whereas if $q \ll n^{-1}$, then there's a lot of structure to investigate, with each term asymptotically having mean $1/q$ and variance $1/q^2$.

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

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

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 $2 \leqslant j \leqslant n$. Thus the expected number of components equals $p + (n-1)qp = nqp + p^2$. $\qquad\square$

Thus the expected number of components is finite precisely when either $p \lesssim n^{-1}$ or $q \lesssim n^{-1}$. In fact, for any fixed $k \geqslant 2$, we find using the First and Second Moment Methods that $p \asymp n^{-1}$ and $q \asymp n^{-1}$ are the lower and upper thresholds for there being at least k components and at least k gaps:

Proposition 13. *Suppose $k \geqslant 2$ is constant. Then,*

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

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 $i_{j+1} \geqslant i_j + 2$ 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

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

Suppose $\omega \gg 1$. If $p = n^{-1}/\omega$, then $\mathbb{E}\big[X\big] \sim \omega^{-k}(1 + k/n)/k! \ll 1$. Thus, by the First Moment Method (Proposition 9), $X = 0$ a.a.s., or equivalently, w.h.p. $\mathbf{C}_{n,p}$ has fewer than k components.

Similarly, if $q = n^{-1}/\omega$, then $\mathbb{E}\big[X\big] \sim \omega^{-k}(1 + k\omega)/k! \ll 1$, since $k \geqslant 2$, and so, again, w.h.p. $\mathbf{C}_{n,p}$ has fewer than k components.

Finally, suppose that $n^{-1} \ll p$ and $q \gg n^{-1}$. Then, $k/nq \ll 1$ and $qp \gg n^{-1}$, so we have $\mathbb{E}\big[X\big] = E_k \sim (nqp)^k/k! \gg 1$.

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 $|i_r - j_s| \leqslant 1$. 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 $\mathbf{j}$. Thus

$$\Delta \ := \ \sum_{\mathbf{i} \sim \mathbf{j}} \mathbb{P}\big[A_{\mathbf{i}} \wedge A_{\mathbf{j}}\big] \ = \ \sum_{t=1}^{k-1} \binom{k+t}{k} \binom{k}{t} E_{k+t} \ < \ C_k \sum_{t=1}^{k-1} E_{k+t} \ < \ \frac{C_k}{k!} \sum_{t=1}^{k-1} (nqp)^{k+t},$$

for some constant C_k. Thus,

$$\frac{\Delta}{\mathbb{E}\big[X\big]^2} \ < \ C_k \, k! \sum_{t=1}^{k-1} (nqp)^{t-k} \ \leqslant \ C_k \, k! \sum_{s=1}^{k-1} (n/4)^{-s} \ \ll \ 1,$$

since $qp \leqslant \frac{1}{4}$.

So by the Second Moment Method (Proposition 10), 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. $\qquad\square$

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. Note that these are monotone properties, since increasing a term by one can never reduce the length of a component or increase the length of a gap. 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 k, for each $i \in [n + 1 - k]$, let A_i be the event "$\mathbf{C}_{n,p}(i), \dots, \mathbf{C}_{n,p}(i + k - 1)$ are all nonzero". Then $\mathbb{P}\big[A_i\big] = p^k$. So, if X is the total number of runs of k nonzero terms in $\mathbf{C}_{n,p}$, then by linearity of expectation, $\mathbb{E}\big[X\big] = (n + 1 - k)p^k \sim np^k$, as long as $k \ll n$.

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 $\mathbb{P}\big[A_i \wedge A_j\big] = p^{k+t}$. So,

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

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

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

To apply the Chen–Stein Method (Proposition 11), it is sufficient to show that $\Delta \ll 1$ and $\Lambda \ll 1$.

The threshold for the appearance in $\mathbf{C}_{n,p}$ of a component of fixed length k is $p \asymp n^{-1/k}$. We also establish the probability of $\mathbf{C}_{n,p}$ having a component of length k when $p \sim \alpha n^{-1/k}$:

Proposition 14. *Suppose $k \geqslant 1$ is constant. Then, for any positive constant α,*

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

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

Proof. If $p \ll n^{-1/k}$, 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 $\Delta/\mathbb{E}\big[X\big]^2 \lesssim p/np^k \ll p \ll 1$. So by the Second Moment Method, $X > 0$ a.a.s., or equivalently, $\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) \geqslant k$ a.a.s. Since the property of having a component of length at least k is increasing, then by Proposition 5 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 11), the

number of components in $\mathbf{C}_{n,p}$ of length k asymptotically satisfies a Poisson distribution with mean α^k. In particular, the asymptotic probability that no components have length k or greater is $e^{-\alpha^k}$. $\qquad\square$

Thus (using Proposition 8 to transfer the thresholds from $\mathbf{C}_{n,p}$ to $\mathbf{C}_{n,m}$), as m increases, w.h.p. we first see components of length 2 in $\mathbf{C}_{n,m}$ when $m \asymp \sqrt{n}$, first see components of length 3 when $m \asymp n^{2/3}$, and so forth. Hence, if $m \sim \alpha n^c$, for positive constants α and $c \in (0, 1)$, then w.h.p. $\mathsf{comp}_{\max}(\mathbf{C}_{n,m})$ takes only one value unless $c = 1 - 1/k$ for some $k \in \mathbb{N}$ when a.a.s. it takes one of the two values in $\{k - 1, k\}$. However, once m grows faster than $n^{1-\delta}$ for every $\delta > 0$ (for example, $m = n/\log n$), a.a.s. the length of the longest component exceeds any fixed value. Similarly, w.h.p. gaps of length 3 vanish once $m \gg n^{4/3}$, every gap has length 1 when $m \gg n^{3/2}$, and there are no gaps at all once $m \gg n^2$.

We now investigate the presence of components or gaps with lengths that increase with n. Our first result reveals a sharp threshold:

Proposition 15. *Suppose $1 \ll k \ll \log n$. Then, for any $\omega \gg 1$ and constant α,*

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

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

Proof. This proof, and many subsequent ones, follow the same bipartite or tripartite structure as seen above, so, from now on, we abbreviate the argument as far as possible.

If $p = e^{-(\log n + \omega)/k}$, then $\mathbb{E}\big[X\big] \sim e^{-\omega} \ll 1$, so $\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) < k$ a.a.s.

If $p = e^{-(\log n - \omega)/k}$, then $\mathbb{E}\big[X\big] \sim e^{\omega} \gg 1$. If $\omega \ll \log n$, then $p \ll 1$ and $\Delta/\mathbb{E}\big[X\big]^2 \lesssim p e^{-\omega} \ll 1$. So $\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) \geqslant k$ a.a.s.

Finally, if $p \sim e^{-(\log n - \alpha)/k}$, then $\mathbb{E}\big[X\big] \sim e^{\alpha}$ and $\Delta < p e^{\alpha}/q \ll 1$. Moreover, we have $\Lambda < nkp^{2k} \sim e^{2\alpha}kn^{-1} \ll 1$. So, the number of components in $\mathbf{C}_{n,p}$ of length at least k is asymptotically Poisson with mean e^{α}. $\qquad\square$

We first see components of length the order of $\log n$ when p is constant. Again, the threshold is sharp:

Proposition 16. *Suppose $k = c \log n$ for some constant c. Then, for any $\omega \gg 1$,*

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

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

Proof. If $p = e^{-1/c - \omega/\log n}$, then $\mathbb{E}[X] \sim e^{-c\omega} \ll 1$, so $\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) < k$ a.a.s.

If $p = e^{-1/c + \omega/\log n}$, then $\mathbb{E}[X] \sim e^{c\omega} \gg 1$. If $\omega \ll \log n$, then p is asymptotically constant and $\Delta/\mathbb{E}[X]^2 \lesssim pe^{-c\omega}/q \ll 1$. So $\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) \geqslant k$ a.a.s. $\qquad\square$

It is notable (see [30, page 6]) that once p has increased to a constant, w.h.p. the longest component in $\mathbf{C}_{n,p}$ has length of the order of $\log n$, despite the mean length of a component still being asymptotically constant (equal to $1/q$).

The threshold for the appearance of components of length $k \gg \log n$ is at $q = k^{-1} \log n$. For example, w.h.p. we first see a component of length $\sqrt{n}$ in $\mathbf{C}_{n,m}$ once m reaches $n^{3/2}/\log n$.

Proposition 17. *Suppose $k = n^c$ for some $c \in (0, 1)$. Then, for any $\omega \gg 1$,*

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

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

Proof. If $q = (\log n + \omega)/n^c$, then $\mathbb{E}[X] \sim n\big(1 - (\log n + \omega)/k\big)^k \sim e^{-\omega} \ll 1$. Therefore, $\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) < k$ a.a.s.

We need to use an alternative bound on Δ.

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

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

If $q = (\log n - \omega)/n^c$, then $\mathbb{E}[X] \sim e^\omega \gg 1$. If $q = \big((1-c)\log n - \omega\big)/n^c$, then

$$\Delta/\mathbb{E}[X]^2 \lesssim pn^{c-1}\Big(1 - \big((1-c)\log n - \omega\big)/k\Big)^{-k} \sim pe^{-\omega} \ll 1.$$

So $\mathsf{comp}_{\max}(\mathbf{C}_{n,p}) \geqslant k$ a.a.s. $\qquad\square$

Finally, once $q \ll n^{-1}$, the expected number of zero terms in $\mathbf{C}_{n,p}$ equals $nq \ll 1$, so w.h.p. every term is nonzero, and $\mathbf{C}_{n,p}$ consists of a single component of length n with no gaps.

3.2 The shortest component and shortest gap

We now establish thresholds for $\mathbf{C}_{n,p}$ to have a component or gap shorter than a specified length. Note that these are not monotone properties. For example, adding one to the last term of the composition 11200 yields 11201 which reduces the length of the shortest component from 3 to 1. If C is a composition, let $\mathsf{comp}_{\min}(C)$ be the length of the shortest component of 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

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

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

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

Given some $k \ll n$ and assuming $kq \ll 1$ (so $1 - p^k \sim kq$), 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 $A_{j,s}$ $(i \leqslant j)$ are correlated $(i, r \sim j, s)$ in two situations. If $j \leqslant i + r$, 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,

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

$$\sim n\sum_{r=1}^{k}\sum_{s=1}^{k}q^3 p^{r+s} = np^2 q\big(1 - p^k\big)^2 \sim k^2 nq^3.$$

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

$$\Lambda := \sum_{i,\ell} \mathbb{P}\big[A_{i,\ell}\big]^2 + \sum_{i,r \sim j,s} \mathbb{P}\big[A_{i,r}\big]\mathbb{P}\big[A_{j,s}\big]$$

$$\sim n\sum_{r=1}^{k}(r+2)\sum_{s=1}^{k}q^4 p^{r+s} \sim \frac{1}{2}k^2(k+5)nq^4 \lesssim k^3 nq^4.$$

Our first result shows thresholds at $p \asymp n^{-1/2}$, for the appearance of short gaps, and at $q \asymp n^{-1/2}$ for the disappearance of short components:

Proposition 18. *Suppose $k \geqslant 1$ 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 $\omega \gg 1$. If $q = n^{-1/2}/\omega$, then $\mathbb{E}[X] \sim k/\omega^2 \ll 1$, so w.h.p. there are no components of length k or less, and $\mathsf{comp}_{\min}(\mathbf{C}_{n,p}) > k$ a.a.s.

If $q = n^{-1/2}\omega$, then $\mathbb{E}[X] \sim k\omega^2 \gg 1$. Since $\Delta/\mathbb{E}[X]^2 \ll 1$, so $\mathsf{comp}_{\min}(\mathbf{C}_{n,p}) \leqslant k$ a.a.s.

Finally, if $q \sim \alpha n^{-1/2}$, then $\mathbb{E}[X] \sim \alpha^2 k$. Also, we have $\Delta \sim \alpha^3 k^2 n^{-1/2} \ll 1$ and $\Lambda \lesssim \alpha^4 k^3 n^{-1/2} \ll 1$. So, the number of components in $\mathbf{C}_{n,p}$ of length at most k is asymptotically Poisson with mean $\alpha^2 k$. In particular, the asymptotic probability that no components have length k or less is $e^{-\alpha^2 k}$. $\qquad\square$

Thus, as soon as $p \gg n^{-1/2}$, w.h.p. there is a gap of length 1 in $\mathbf{C}_{n,p}$. Prior to this, a.a.s. there is no gap of *any* fixed length. However, w.h.p. components of length 1 remain until $q \asymp n^{-1/2}$. At this point, the *longest* components have length the order of $\sqrt{n}\log n$ (Proposition 17). But once $q \ll n^{-1/2}$, a.a.s. no component of *any* fixed length remains. At these thresholds, we can calculate the asymptotic probability that the shortest component or gap has a given length. For instance, when $p \sim n^{-1/2}$, with probability $1 - e^{-1} \approx 0.63$ a gap of length 1 has already appeared.

The thresholds for the disappearance of longer components are as follows. In particular, components of length n^c vanish a.a.s. once $q \ll n^{-(1+c)/2}$. For example, once $m \gg n^{7/4}$, a.a.s. all components have length $\gg \sqrt{n}$.

Proposition 19. *Suppose $1 \ll k \ll n$. Then, for any positive constant α,*

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

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

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

Suppose $\omega \gg 1$. If $q = \omega^{-1}/\sqrt{kn}$, then $\mathbb{E}[X] \sim \omega^{-2} \ll 1$, so w.h.p. we have $\mathsf{comp}_{\min}(\mathbf{C}_{n,p}) > k$.

If $q = \omega/\sqrt{kn}$, then $\mathbb{E}[X] \sim \omega^2 \gg 1$. Since $\Delta/\mathbb{E}[X]^2 \ll 1$, we have $\mathsf{comp}_{\min}(\mathbf{C}_{n,p}) \leqslant k$ a.a.s.

Finally, if $q \sim \alpha/\sqrt{kn}$, then $\mathbb{E}[X] \sim \alpha^2$. Also, $\Delta \sim \alpha^3\sqrt{k/n} \ll 1$ and $\Lambda \lesssim \alpha^4 k/n \ll 1$. So, the number of components in $\mathbf{C}_{n,p}$ of length at most k is asymptotically Poisson with mean α^2. $\qquad\square$

4 Patterns

The focus of the remainder of this work is on the appearance and disappearance of various types of *pattern*. A pattern is simply a sub-composition, under some notion of
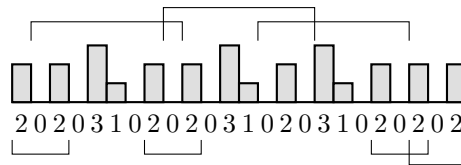
Figure 2: A composition containing four occurrences of the exact consecutive pattern $^{=}\overline{202}$ and three occurrences of $^{=}\overline{02031020}$

containment. Below, we consider a variety of different types of pattern containment. To distinguish between these, we sometimes add a prefix to the pattern (e.g., $^{=}\pi$ or $^{\geqslant}\pi$). Definitions are given below. For the most part we require the terms of a pattern to occur consecutively in a composition, which we signify by using an overline (e.g., $^{=}\overline{123}$). This requirement is eventually relaxed in the final subsection.

We begin in Section 4.1 by considering *exact* patterns, in which terms must take specified values, including an investigation of runs of equal nonzero terms and of *square* patterns (runs of k terms equal to k). Section 4.2 concerns *upper* and *lower* patterns, including a consideration of the largest and smallest terms. We then look at patterns specifying the relative ordering of terms (Section 4.3), including determining the threshold for $\mathbf{C}_{n,p}$ to be a *Carlitz* composition (having no adjacent pair of equal terms). Finally, in Section 4.4, we consider nonconsecutive patterns, including determining the threshold for all the terms of $\mathbf{C}_{n,p}$ to be distinct.

There is quite an extensive literature on patterns in compositions and words, some of which is mentioned below. This includes comprehensive expositions by Heubach and Mansour [28] and Kitaev [35]. However, the enumerative and generatingfunctionological perspective taken in these works is somewhat orthogonal to our interests.

4.1 Exact consecutive patterns

We begin with the simplest notion of pattern. The *exact consecutive pattern* $^{=}\overline{r_1 \dots r_k}$ 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 2 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. It isn't even convex; for example, $^{=}\overline{22}$ occurs in the compositions 221 and 322, but does not occur in 321. However, in $\mathbf{C}_{n,p}$, the presence of a nonzero exact pattern $^{=}\overline{\pi}$ does exhibit both a lower threshold, for its appearance, and an upper threshold, for its disappearance. If $\pi = r_1 \dots r_k$, then the lower threshold depends only on the *size* $|\pi| = \sum_{i=1}^k r_i$ of the pattern, whereas the upper threshold depends only on its *length* k. (Recall that in this work $|\pi|$ always denotes the sum of the terms of π, *not* its length.)

Proposition 20. *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 $\pi = r_1 \ldots r_k$ and $|\pi| = s$. For each $i \in [n+1-k]$, let A_i be the event that $^=\overline{\pi}$ 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 $p = s/(k+s)$.

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, $\mathbb{P}\big[A_i \wedge A_j\big] \leqslant q^{k+1} p^{s+1}$, since π is nonzero. Thus,

$$\Delta := \sum_{i \sim j} \mathbb{P}\big[A_i \wedge A_j\big] \leqslant nkq^{k+1}p^{s+1} \quad \text{and} \quad R := \Delta/\mathbb{E}\big[X\big]^2 \lesssim \frac{k}{nq^{k-1}p^{s-1}}.$$

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

Suppose $p = \omega n^{-1/s} \ll 1$ for some $\omega \gg 1$. Then $\mathbb{E}\big[X\big] \sim \omega^s \gg 1$ and $R \lesssim kp/\omega^s \ll 1$. Similarly, if $q = \omega n^{-1/k} \ll 1$ for some $\omega \gg 1$, then $\mathbb{E}\big[X\big] \sim \omega^k \gg 1$ and $R \lesssim kq/\omega^k \ll 1$. 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. $^=\overline{\pi}$ occurs in $\mathbf{C}_{n,p}$.

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

Thus, we have a dichotomy between the arrival and the departure of exact consecutive patterns as $\mathbf{C}_{n,p}$ evolves, the former being ordered by size, smaller patterns appearing before larger ones, and the latter being ordered by length, longer patterns disappearing before shorter ones.

The following proposition enables us to transfer the thresholds for exact consecutive patterns from $\mathbf{C}_{n,p}$ to $\mathbf{C}_{n,m}$. Since the relevant properties are not monotone, we can't simply apply Proposition 6.

Proposition 21. *If $^=\overline{\pi}$ is an exact consecutive pattern and $m \sim np/q \gg 1$, then*

$$\mathbb{P}\big[\mathbf{C}_{n,m} \text{ contains } {^=\overline{\pi}}\big] \;\sim\; \mathbb{P}\big[\mathbf{C}_{n,p} \text{ contains } {^=\overline{\pi}}\big].$$

Proof. Suppose π has length k and size s. For each $i \in [n+1-k]$, let P_i be the probability that $^=\overline{\pi}$ 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,

$$P_i \;=\; \binom{m_s + n_k}{m_s}\binom{m + n_1}{m}^{-1}$$

$$\sim\; \sqrt{\frac{(m_s + n_k)\, m\, n_1}{m_s\, n_k\, (m + n_1)}}\, \frac{(m_s + n_k)^{m_s + n_k}}{m_s^{m_s}\, n_k^{n_k}}\, \frac{m^m\, n_1^{n_1}}{(m + n_1)^{m + n_1}}$$

$$\sim\; \left(\frac{m_s + n_k}{m + n_1}\right)^{m + n_1} (m_s + n_k)^{-s-k} \left(\frac{m}{m_s}\right)^{m_s} m^s \left(\frac{n_1}{n_k}\right)^{n_k} n_1^k$$

$$=\; \left(1 - \frac{s+k}{m+n_1}\right)^{m+n_1} \left(1 + \frac{s}{m_s}\right)^{m_s} \left(1 + \frac{k}{n_k}\right)^{n_k} \frac{m^s\, n_1^k}{(m_s + n_k)^{s+k}}$$

$$\sim\; e^{-s-k}\, e^s\, e^k\, p^s\, q^k \;=\; p^s q^k \;\sim\; \mathbb{P}\big[{^=\overline{\pi}} \text{ occurs at position } i \text{ in } \mathbf{C}_{n,p}\big].$$

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 (see the proof of Proposition 20). $\qquad\square$

Thus, $m \asymp n^{1-1/s}$ is the threshold for the appearance of each exact consecutive pattern of size s in $\mathbf{C}_{n,m}$. So, if $0 < \gamma < 1$ and $m \sim n^\gamma$, then w.h.p. $\mathbf{C}_{n,m}$ contains every exact consecutive pattern of size less than $1/(1-\gamma)$, but contains no such pattern of size greater than $1/(1-\gamma)$. For example, every fixed length gap of the form $^=\overline{10\ldots01}$ appears when $m \asymp n^{1/2}$ (see Proposition 18), whereas no gap delimited by larger terms appears before $m \asymp n^{2/3}$.

Similarly, $m \asymp n^{1+1/k}$ is the threshold for the disappearance of each exact consecutive pattern of length k in $\mathbf{C}_{n,m}$. So, if $1 < \gamma < 2$ and $m \sim n^\gamma$, then w.h.p. $\mathbf{C}_{n,m}$ contains every exact consecutive pattern of length less than $1/(\gamma-1)$, but contains no such pattern of length greater than $1/(\gamma - 1)$.

These results establish when any given exact consecutive pattern is present. For example, w.h.p. the pattern $^=\overline{314159}$ appears when $m \asymp n^{22/23}$ and has disappeared once $m \gg n^{7/6}$. If π_1 is both shorter and smaller than π_2, then $^=\overline{\pi_1}$ arrives before $^=\overline{\pi_2}$ and leaves after $^=\overline{\pi_2}$. However, since the arrival and departure of patterns depend on

different parameters, the values of which can be chosen independently, it is possible for the departure order of a number of patterns to be any permutation of their arrival order. For example, w.h.p., $=\overline{101}$, $=\overline{3}$, $=\overline{1111}$ and $=\overline{23}$ arrive in that order, but depart in the order $=\overline{1111}$, $=\overline{101}$, $=\overline{23}$, $=\overline{3}$.

Observe that there is a range of values of m for which, a.a.s., *any* given exact consecutive pattern is present in $\mathbf{C}_{n,m}$. Specifically, if $n^{1-\delta} \ll m \ll n^{1+\delta}$ for every $\delta > 0$ (for example, $m \sim n/\log n$ or $m \sim n \log n$), then any given pattern occurs w.h.p.

In contrast, once $m \gg n^2$, then w.h.p. any specific exact consecutive pattern is absent from $\mathbf{C}_{n,m}$. In particular, this holds for patterns of length one: given any value r, a.a.s. no term of $\mathbf{C}_{n,m}$ is equal to r.

4.1.1 Runs of equal terms

In this section we investigate runs of k equal nonzero terms; that is, the occurrence of any of the exact consecutive patterns $=\overline{11\ldots 1}$, $=\overline{22\ldots 2}$, $=\overline{33\ldots 3}$, etc. of length k.

Given $k \in [n]$ and $i \in [n+1-k]$, let A_i be the event that the ith term of $\mathbf{C}_{n,p}$ is the start of a run of k equal nonzero terms. Then

$$\mathbb{P}\big[A_i\big] \;=\; \sum_{r=1}^{\infty} (qp^r)^k \;=\; \frac{p^k q^k}{1-p^k}.$$

Thus, if X is the number of runs of k equal nonzero terms, $\mathbb{E}\big[X\big] \sim np^k q^k/(1-p^k)$, assuming $k \ll n$. Distinct events A_i and A_j are correlated if $t = |j - i| < k$, with

$$\mathbb{P}\big[A_i \wedge A_j\big] \;=\; \sum_{r=1}^{\infty} (qp^r)^{k+t} \;=\; \frac{(pq)^{k+t}}{1-p^{k+t}} \;\leqslant\; \frac{(pq)^{k+1}}{1-p^{k+1}}.$$

Proposition 22. *Let* $P_k = \lim\limits_{n\to\infty} \mathbb{P}\big[\mathbf{C}_{n,p} \text{ contains a run of } k \text{ equal nonzero terms}\big]$. *Then, for any fixed* $k \geqslant 2$ *and positive constant* α,

$$P_k = \begin{cases} 0, & \text{if } p \ll n^{-1/k}, \\ 1 - e^{-\alpha^k}, & \text{if } p \sim \alpha n^{-1/k}, \\ 1, & \text{if } n^{-1/k} \ll p \ \text{ and } \ q \gg n^{-1/(k-1)}, \\ 1 - e^{-\alpha^{k-1}/k}, & \text{if } q \sim \alpha n^{-1/(k-1)}, \\ 0, & \text{if } n^{-1/(k-1)} \gg q. \end{cases}$$

Proof. Suppose $\omega \gg 1$. If $p = n^{-1/k}/\omega$, then $\mathbb{E}\big[X\big] \sim \omega^{-k} \ll 1$. Similarly, if $q = n^{-1/(k-1)}/\omega$, then $\mathbb{E}\big[X\big] \sim \omega^{-(k-1)}/k \ll 1$. So, in either case, w.h.p. $\mathbf{C}_{n,p}$ has no run of k equal nonzero terms.

Let

$$\Delta \;:=\; \sum_{i\sim j} \mathbb{P}\big[A_i \wedge A_j\big] \;\leqslant\; \frac{nk(pq)^{k+1}}{1-p^{k+1}} \quad \text{and} \quad R \;:=\; \frac{\Delta}{\mathbb{E}\big[X\big]^2} \;\lesssim\; \frac{kpq(1-p^k)^2}{np^k q^k(1-p^{k+1})}.$$

If $p = \omega n^{-1/k} \ll 1$, then $\mathbb{E}[X] \sim \omega^k \gg 1$ and $R \lesssim kp\omega^{-k} \ll 1$. Similarly, if we have $q = \omega n^{-1/(k-1)} \ll 1$, then $\mathbb{E}[X] \sim \omega^{k-1}/k \gg 1$ and $R \lesssim k^2 q \omega^{-(k-1)} \ll 1$. Finally, if p is asymptotically bounded away from both 0 and 1, then $\mathbb{E}[X] \asymp n \gg 1$ and $R \asymp n^{-1} \ll 1$. Hence, if $n^{-1/s} \ll p$ and $q \gg n^{-1/(k-1)}$, w.h.p. $\mathbf{C}_{n,p}$ contains a run of k equal nonzero terms.

Let

$$\Lambda := \sum_i \mathbb{P}[A_i]^2 + \sum_{i \sim j} \mathbb{P}[A_i]\mathbb{P}[A_j] \sim \frac{nk(pq)^{2k}}{(1-p^k)^2}.$$

Suppose $p \sim \alpha n^{-1/k}$. Then $\mathbb{E}[X] \sim \alpha^k$ and $\Delta \leqslant k\alpha^{k+1}n^{-1/k} \ll 1$, and $\Lambda \sim k\alpha^{2k}/n \ll 1$. So the number of occurrences of runs of k equal nonzero terms is asymptotically Poisson with mean α^k. Similarly, if $q \sim \alpha n^{-1/(k-1)}$ then $\mathbb{E}[X] \sim \alpha^{k-1}/k$ and $\Delta \leqslant \alpha^k n^{-1/(k-1)} \ll 1$, and $\Lambda \sim \alpha^{2k-2}/kn \ll 1$, so the number of occurrences of runs of k equal nonzero terms is asymptotically Poisson with mean α^{k-1}/k. $\square$

Thus, as one would expect, a.a.s the arrival of a run of k nonzero terms coincides with the appearance of the length k pattern $\overline{=11\ldots1}$. However, the threshold at $q \asymp n^{-1/(k-1)}$ for the departure of all runs of k nonzero terms is later than that at $q \asymp n^{-1/k}$ for the disappearance of any specific length k pattern $\overline{=rr\ldots r}$ (see Proposition 20).

In the case that p is constant, the length of the longest run in $\mathbf{C}_{n,p}$ has been investigated in detail (see [38, 23, 15]), and Gafni [19] has considered the question for a random composition with positive terms. For our variant of the problem, we have the following sharp bounds on the length of the longest run of nonzero terms, which is maximal when $p = \frac{1}{2}$.

Proposition 23. *If $p \in (0,1)$ is constant and $\gamma = 1/pq$, then for any $\omega \gg 1$,*

$$\mathbb{P}[\mathbf{C}_{n,p} \text{ contains a run of } k \text{ equal nonzero terms}] \sim \begin{cases} 0 & \text{if } k = \log_\gamma n + \omega, \\ 1 & \text{if } k = \log_\gamma n - \omega. \end{cases}$$

Proof. If $k = \log_\gamma n + \omega$, then $\mathbb{E}[X] \sim np^k q^k \sim \gamma^{-\omega} \ll 1$ since $\gamma > 1$. So w.h.p. $\mathbf{C}_{n,p}$ has no run of k equal nonzero terms.

To apply the Second Moment Method, we require a stronger bound on Δ than that used in Proposition 22:

$$\Delta := \sum_{i \sim j} \mathbb{P}[A_i \wedge A_j] \leqslant n \sum_{t=1}^{\infty} \frac{(pq)^{k+t}}{1-p^{k+1}} \sim \frac{n(pq)^{k+1}}{1-pq}.$$

If $k = \log_\gamma n - \omega$, then $\mathbb{E}[X] \sim \gamma^\omega \gg 1$, so

$$\frac{\Delta}{\mathbb{E}[X]^2} \lesssim \frac{(pq)^{-(k-1)}}{n(1-pq)} \sim \frac{\gamma^{-\omega}}{\gamma - 1} \ll 1,$$

and w.h.p. $\mathbf{C}_{n,p}$ contains a run of k equal nonzero terms. $\square$

4.1.2 Square patterns

We conclude our investigation of exact consecutive patterns by considering the k-*square* pattern $\overline{=kk\ldots k}$, of length k, an occurrence of which is a run of k terms each equal to k. For example, the composition in Figure 1 on page 5 contains a 4-square and two 2-squares. From Proposition 20, we know that, for any fixed k, a.a.s. k-squares appear when $p \asymp n^{-1/k^2}$ and disappear when $q \asymp n^{-1/k}$.

What is the *largest* square that we can expect to see in the evolution of the random composition? Here we show that w.h.p. it has side length a little greater than $\log n/\log\log n$, as expected from the following non-rigorous heuristic argument: We expect to see the largest k-square when the two thresholds coalesce, that is when both $p = n^{-1/k^2}$ and $q = n^{-1/k}$, so $q = p^k$. Thus $p^k + p = 1$ or $(1-q)^k + (1-q) = 1$. Approximating $(1-q)^k$ with $1 - kq$ (since $k \gg 1$ implies $q \ll 1$) then yields $q = 1/(k+1)$. So $1/(k+1) = n^{-1/k}$ or $k\log(k+1) = \log n$, whose asymptotic solution matches the value of k in the following proposition.

Proposition 24. *Suppose* $q = \dfrac{\theta \log\log n}{\log n}$ *and* $k = \dfrac{\log n}{\log\log n}\left(1 + \dfrac{c\log\log\log n}{\log\log n}\right)$. *Then,*

$$\mathbb{P}\big[\mathbf{C}_{n,p} \text{ contains a } k\text{-square pattern}\big] \sim \begin{cases} 0 & \text{if } c \geqslant 1 \text{ or } \theta \text{ is far,} \\ 1 & \text{if } c < 1 \text{ and } \theta \text{ is near,} \end{cases}$$

where we say that θ *is* near *if* $(\log\log n)^{-\gamma} \ll \theta \ll \log\log\log n$ *for all constant* $\gamma > 0$, *and* θ *is* far *if either* $\theta \ll (\log\log n)^{-\gamma}$ *for all constant* γ *or else* $\log\log\log n \ll \theta$.

Proof. For $i \in [n+1-k]$, let A_i be the event that the ith term of $\mathbf{C}_{n,p}$ is the start of a k-square. Then $\mathbb{P}\big[A_i\big] = q^k p^{k^2}$. Thus, if X is the number of k-squares,

$$\mathbb{E}\big[X\big] \sim q^k p^{k^2} \sim \left(\frac{\theta\log\log n}{\log n}\right)^k \exp\left(-\frac{k^2\,\theta\log\log n}{\log n}\right).$$

Substituting for k and taking the dominant term, if θ is near and $c \neq 1$ then

$$\log\mathbb{E}\big[X\big] \sim (1-c)\frac{\log n\log\log\log n}{\log\log n},$$

and if θ is near and $c = 1$ then

$$\log\mathbb{E}\big[X\big] \sim -\frac{\log n}{\log\log n}.$$

If $\omega \gg 1$ and either $\theta = (\log\log n)^{-\omega}$ or else $\theta = \omega\log\log\log n$, then

$$\log\mathbb{E}\big[X\big] \sim -\frac{\omega\log n\log\log\log n}{\log\log n}.$$

Thus, if $c \geqslant 1$ or θ is far, we have $\mathbb{E}\big[X\big] \ll 1$ and so w.h.p. $\mathbf{C}_{n,p}$ contains no k-square, whereas if $c < 1$ and θ is near then $\mathbb{E}\big[X\big] \gg 1$.

Distinct events A_i and A_j are correlated if $t = |j-i| < k$, with $\mathbb{P}\big[A_i \wedge A_j\big] = q^{k+t}p^{k^2+kt}$. Thus,

$$\Delta \;:=\; \sum_{i \sim j} \mathbb{P}\big[A_i \wedge A_j\big] \;\leqslant\; n\sum_{t=1}^{\infty} q^{k+t}p^{k^2+kt} \;=\; \frac{nq^{k+1}p^{k^2+k}}{1-qp^k} \;\sim\; nq^{k+1}p^{k^2+k},$$

and

$$R \;:=\; \frac{\Delta}{\mathbb{E}\big[X\big]^2} \;\lesssim\; n^{-1}q^{1-k}p^{k-k^2} \;\sim\; n^{-1}\left(\frac{\log n}{\log\log n}\right)^{k-1}\exp\left((k^2-k)\frac{\log\log n}{\log n}\right).$$

Substituting for k and taking the dominant term, if θ is near and $c \neq 1$ then

$$\log R \;\sim\; (c-1)\frac{\log n \log\log\log n}{\log\log n}.$$

So if $c < 1$, we have $R \ll 1$ and thus w.h.p. $\mathbf{C}_{n,p}$ contains a k-square. $\qquad\square$

4.2 Upper and lower consecutive patterns

Although the presence of an exact pattern is not even a convex property, we can weaken our notion of a pattern in two natural ways to yield properties that are monotone. The *upper* consecutive pattern $^{\geqslant}\overline{r_1 \ldots r_k}$ occurs at position i in a composition C if, for each $j \in [k]$, we have $C(i-1+j) \geqslant r_j$, and the *lower* consecutive pattern $^{\leqslant}\overline{r_1 \ldots r_k}$ occurs at position i in C if, for each $j \in [k]$, we have $C(i-1+j) \leqslant r_j$. Thus the presence of an upper pattern is an increasing property, whereas the presence of a lower pattern is decreasing.

The analysis of upper and lower patterns is similar to that for exact patterns in Proposition 20. Indeed, the threshold for the appearance of $^{\geqslant}\overline{\pi}$ is the same as that for $^{=}\overline{\pi}$, and the threshold for the disappearance of $^{\leqslant}\overline{\pi}$ is the same as that for $^{=}\overline{\pi}$, although in the latter case the threshold probabilities differ.

Proposition 25. *If $^{\geqslant}\overline{\pi}$ is a nonzero upper consecutive pattern, then for any positive constant α,*

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

Proof. Suppose $\pi = r_1 \ldots r_k$ and $|\pi| = s$. For each $i \in [n+1-k]$, let A_i be the event that $^{\geqslant}\overline{\pi}$ occurs at position i in $\mathbf{C}_{n,p}$, and let X be the number of occurrences of $^{\geqslant}\overline{\pi}$ in $\mathbf{C}_{n,p}$. Then, $\mathbb{P}\big[A_i\big] = p^s$, and $\mathbb{E}\big[X\big] \sim np^s$. So if $p \ll n^{-1/s}$, then $\mathbb{E}\big[X\big] \ll 1$ and w.h.p. $^{\geqslant}\overline{\pi}$ doesn't occur in $\mathbf{C}_{n,p}$.

Distinct events A_i and A_j are correlated if $t = |j-i| < k$. Then,

$$\mathbb{P}\big[A_i \wedge A_j\big] \;=\; \prod_{\ell=1}^{t} p^{r_\ell} \;\times\; \prod_{\ell=1}^{k-t} p^{\max(r_\ell,\, r_{\ell+t})} \;\times\; \prod_{\ell=k-t+1}^{k} p^{r_\ell},$$

which is at most p^{s+1}, since either $r_\ell > 0$ for some $\ell \leqslant t$ or else $r_{\ell+t} > r_\ell$ for some $\ell \leqslant k-t$. Thus,

$$\Delta := \sum_{i \sim j} \mathbb{P}[A_i \wedge A_j] \leqslant nkp^{s+1} \quad \text{and} \quad R := \Delta/\mathbb{E}[X]^2 \lesssim \frac{k}{np^{s-1}}.$$

Moreover, $\Lambda := \sum_i \mathbb{P}[A_i]^2 + \sum_{i \sim j} \mathbb{P}[A_i]\mathbb{P}[A_j] \sim nkp^{2s}$.

Suppose $p = \omega n^{-1/s} \ll 1$ for some $\omega \gg 1$. Then $\mathbb{E}[X] \sim \omega^s \gg 1$ and $R \lesssim kp/\omega^s \ll 1$. Hence, w.h.p. $^{\geqslant}\overline{\pi}$ occurs in $\mathbf{C}_{n,p}$.

Suppose $p \sim \alpha n^{-1/s}$. Then $\mathbb{E}[X] \sim \alpha^s$ and $\Delta \leqslant \alpha^s kp \ll 1$, and $\Lambda \sim \alpha^{2s}k/n \ll 1$. So the number of occurrences of $^{\geqslant}\overline{\pi}$ is asymptotically Poisson with mean α^s. $\qquad\square$

For a lower consecutive pattern $^{\leqslant}\overline{\pi}$, the probability at the threshold depends on the number of distinct patterns weakly dominated by π of the same length.

Proposition 26. *If $^{\leqslant}\overline{r_1 \dots r_k}$ is a lower consecutive pattern, then for any positive constant α,*

$$\mathbb{P}\big[\mathbf{C}_{n,p} \text{ contains } {}^{\leqslant}\overline{r_1 \dots r_k}\big] \sim \begin{cases} 1, & \text{if } q \gg n^{-1/k}, \\ 1 - e^{-\alpha^k \rho}, & \text{if } q \sim \alpha n^{-1/k}, \\ 0, & \text{if } n^{-1/k} \gg q, \end{cases}$$

where $\rho = \displaystyle\prod_{i \in [k]} (r_i + 1)$.

Proof. For each $i \in [n+1-k]$, let A_i be the event that $^{\leqslant}\overline{\pi}$ occurs at position i in $\mathbf{C}_{n,p}$, and let X be the number of occurrences of $^{\leqslant}\overline{\pi}$ in $\mathbf{C}_{n,p}$. Then, if $q \ll 1$,

$$\mathbb{P}[A_i] = \prod_{\ell \in [k]} (1 - p^{r_\ell+1}) = \prod_{\ell \in [k]} \big(1 - (1-q)^{r_\ell+1}\big) \sim \prod_{\ell \in [k]} (r_\ell + 1)q = \rho q^k,$$

and $\mathbb{E}[X] \sim \rho nq^k$. So if $q \ll n^{-1/k}$, then $\mathbb{E}[X] \ll 1$ and w.h.p. $^{\leqslant}\overline{\pi}$ doesn't occur in $\mathbf{C}_{n,p}$.

Distinct events A_i and A_j are correlated if $t = |j - i| < k$. Then, if $q \ll 1$,

$$\mathbb{P}[A_i \wedge A_j] = \prod_{\ell=1}^{t}(1 - p^{r_\ell+1}) \times \prod_{\ell=1}^{k-t}(1 - p^{\min(r_\ell, r_{\ell+t})+1}) \times \prod_{\ell=k-t+1}^{k}(1 - p^{r_\ell+1}) \leqslant \rho^2 q^{k+1}.$$

Thus,

$$\Delta := \sum_{i \sim j} \mathbb{P}[A_i \wedge A_j] \leqslant \rho^2 nkq^{k+1} \quad \text{and} \quad R := \Delta/\mathbb{E}[X]^2 \lesssim \frac{k}{nq^{k-1}}.$$

Moreover, $\Lambda := \sum_i \mathbb{P}[A_i]^2 + \sum_{i \sim j} \mathbb{P}[A_i]\mathbb{P}[A_j] \sim \rho^2 nkq^{2k}$.

Suppose $q = \omega n^{-1/k} \ll 1$ for some $\omega \gg 1$. Then $\mathbb{E}[X] \sim \rho\omega^k \gg 1$ and we have $R \lesssim kq/\omega^k \ll 1$. Hence, w.h.p. $^{\leqslant}\overline{\pi}$ occurs in $\mathbf{C}_{n,p}$.

Suppose $q \sim \alpha n^{-1/k}$. Then $\mathbb{E}[X] \sim \alpha^k \rho$ and $\Delta \leqslant \alpha^k \rho^2 kq \ll 1$, and we have $\Lambda \sim \alpha^{2k}\rho^2 k/n \ll 1$. So the number of occurrences of $^{\leqslant}\overline{\pi}$ is asymptotically Poisson with mean $\alpha^k \rho$. $\qquad\square$

4.2.1 The largest term

We conclude our investigation of upper and lower consecutive patterns by considering the largest and smallest terms in $\mathbf{C}_{n,p}$. If $\mathsf{max}(C)$ is the largest term in a composition C, then the following is an immediate consequence of Proposition 25.

Proposition 27. *If r is a positive integer, then*

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

Thus $\mathsf{max}(\mathbf{C}_{n,p}) = r$ a.a.s. if $n^{-1/r} \ll p \ll n^{-1/(r+1)}$.

However, the analysis of largest term doesn't require Proposition 25, since we have an exact expression for the probability that all the terms are small: $\mathbb{P}\big[\mathsf{max}(\mathbf{C}_{n,p}) < r\big] = (1 - p^r)^n$. For small p we have the following:

Proposition 28. *Suppose $\omega \gg 1$ and $p = \omega^{-1}$. Then, for any constant c,*

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

Proof. If $r = \frac{\log n + \delta}{\log \omega}$, then

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

Thus, w.h.p. $\mathsf{max}(\mathbf{C}_{n,p}) \sim \log n / \log(1/p)$ if $p \ll 1$, and a.a.s. the largest term exhibits *two-point concentration*, that is it takes one of only two possible values.

This is not true if p is constant. In this case, if $L_p = \log_{1/p} n$ then the probability that the largest term differs from L_p by a constant is bounded away from both 0 and 1:

Proposition 29. *If p is constant, then for any constant c,*

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

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

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

Thus, when p is constant, the distribution of the largest term is concentrated around $\log_{1/p} n$. For a more detailed analysis, including the very slight oscillatory behaviour of its mean (dependent on the fractional part of $\log_{1/p} n$), and the determination of its (constant) variance, see [41, Section 4.1] or the exposition (for $p = \frac{1}{2}$) in [16, Proposition V.1].

When p tends to 1, we have the following bounds on the largest term:

Proposition 30. *If $q \ll 1$, then for any $\varepsilon > 0$,*

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

Proof. 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,

$$L \;<\; n\log\big(1 - \exp(-rq(1+q))\big) \;=\; n\log\big(1 - n^{-(1-\varepsilon)(1+q)}\big) \;<\; -n^{\varepsilon - (1-\varepsilon)q},$$

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

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.

4.2.2 The smallest term

We now turn briefly to consider the smallest term in $\mathbf{C}_{n,p}$. If $\mathsf{min}(C)$ is the smallest term in a composition C, then the following is an immediate consequence of Proposition 26.

Proposition 31. *If r is a positive integer, then for any positive constant α,*

$$\mathbb{P}\big[\mathsf{min}(\mathbf{C}_{n,p}) \geqslant r\big] \;\sim\; \begin{cases} 0, & \text{if } q \gg n^{-1}, \\ e^{-\alpha r}, & \text{if } q \sim \alpha n^{-1}, \\ 1, & \text{if } n^{-1} \gg q. \end{cases}$$

Thus, a.a.s. $\mathsf{min}(\mathbf{C}_{n,p}) = 0$ as long as $q \gg n^{-1}$, but $q \asymp n^{-1}$ is the threshold for $\mathsf{min}(\mathbf{C}_{n,p})$ to exceed *any* fixed positive value. This somewhat paradoxical phenomenon is perhaps a little easier to understand if one recalls that the most likely value taken by any term in $\mathbf{C}_{n,p}$ is *zero* — whatever the values of n and p.

More generally, we have the following.

Proposition 32. *If $r \gg 1$, then for any positive constant α,*

$$\mathbb{P}\big[\min(\mathbf{C}_{n,p}) \geqslant r\big] \;\sim\; \begin{cases} 0, & \text{if } q \gg 1/rn, \\ e^{-\alpha}, & \text{if } q \sim \alpha/rn, \\ 1, & \text{if } 1/rn \gg q. \end{cases}$$

Proof. Suppose $\omega \gg 1$. Then,

$$\mathbb{P}\big[\min(\mathbf{C}_{n,p}) \geqslant r\big] \;=\; (1-q)^{rn} \;=\; \begin{cases} (1-\omega/rn)^{rn} \lesssim e^{-\omega} \ll 1, & \text{if } q = \omega/rn, \\ (1-\alpha/rn)^{rn} \sim e^{-\alpha}, & \text{if } q \sim \alpha/rn, \\ (1-1/\omega rn)^{rn} \sim e^{-1/\omega} \sim 1, & \text{if } q = 1/\omega rn. \end{cases} \qquad \square$$

Hence, once $m \gg n^2$, a.a.s. $\min(\mathbf{C}_{n,m}) \asymp m/n^2$.

4.3 Consecutive patterns specifying relative ordering

A classical notion of pattern containment is that a pattern specifies an ordering. We call such a pattern an *ordering* pattern, to distinguish from other types of pattern in this work, and represent ordering patterns without a prefix. The consecutive ordering pattern $\overline{\pi}$ of length k occurs at position ℓ in a composition C if the relative order of $C(\ell), \ldots, C(\ell-1+k)$ is the same as that of $\pi(1), \ldots, \pi(k)$. That is, if, for every $i, j \in [k]$,

$$\begin{aligned} C(\ell-1+i) &< C(\ell-1+j) &\iff& \pi(i) &<& \pi(j), \\ C(\ell-1+i) &= C(\ell-1+j) &\iff& \pi(i) &=& \pi(j), \\ C(\ell-1+i) &> C(\ell-1+j) &\iff& \pi(i) &>& \pi(j). \end{aligned}$$

For example, $\overline{0211}$ occurs at position ℓ in C if $C(\ell) < C(\ell+2) = C(\ell+3) < C(\ell+1)$. The composition in Figure 1 on page 5 contains a single occurrence of $\overline{0211}$, formed by the consecutive terms 1633.

For uniqueness, we require that the set of numbers used in an ordering pattern is an initial segment of the nonnegative integers. Thus, $\overline{0211}$ is an ordering pattern, whereas $\overline{0322}$, $\overline{1322}$ and $\overline{4977}$ are not valid patterns. For results concerning ordering patterns from an enumerative perspective, see the paper by Kitaev, McAllister and Petersen [36] and the papers by Heubach and Mansour [26, 27].

Let $P(\overline{\pi})$ denote the probability that the ordering pattern $\overline{\pi}$ occurs at some specified position in $\mathbf{C}_{n,p}$. Then, since terms in $\mathbf{C}_{n,p}$ are independent, $P(\overline{\pi})$ does not depend on the order of numbers in π. For example, $P(\overline{0211}) = P(\overline{1021}) = P(\overline{0112})$.

Before we proceed further, we need a technical result relating the probability of a "shifted" occurrence of $\overline{\pi}$ to the value of the "unshifted" probability $P(\overline{\pi})$.

Proposition 33. *Suppose $E_i^h(\overline{\pi})$ is the event that $\overline{\pi}$ occurs at position i in $\mathbf{C}_{n,p}$ with every term of its occurrence at least h. If $\overline{\pi}$ has length k and $i \in [n+1-k]$, then $\mathbb{P}\big[E_i^h(\overline{\pi})\big] = p^{hk}P(\overline{\pi})$.*

Proof. Observe that, by definition,

$$P(\overline{\pi}) = \sum_{\rho \cong \pi} P({}^{=}\overline{\rho}),$$

where the sum is over all the exact patterns that are order-isomorphic to π. Also,

$$\{\rho : \rho \cong \pi \wedge \mathsf{min}(\rho) \geqslant h\} = \{\rho + h : \rho \cong \pi\},$$

where $\rho + h$ is the result of adding h to each term of ρ.

Now $P({}^{=}\overline{\rho + h}) = p^{hk}P({}^{=}\overline{\rho})$. So,

$$\mathbb{P}\big[E_i^h(\overline{\pi})\big] = \sum_{\rho \cong \pi} P({}^{=}\overline{\rho + h}) = \sum_{\rho \cong \pi} p^{hk}P({}^{=}\overline{\rho}) = p^{hk}\sum_{\rho \cong \pi} P({}^{=}\overline{\rho}) = p^{hk}P(\overline{\pi}). \qquad \square$$

With this result, we can determine the probability of an ordering pattern occurring at some specified position in $\mathbf{C}_{n,p}$.

4.3.1 Consecutive ordering patterns with distinct terms

We begin by considering patterns in which every term is distinct, such as $\overline{102}$ or $\overline{3120}$. These patterns are permutations of an initial segment of the nonnegative integers, and an occurrence of such a pattern induces a total order on the relevant terms, so we call an ordering pattern in which every term is distinct a *total ordering pattern.*

Proposition 34. *Suppose $\overline{\pi}$ is a total ordering pattern of length k. Then, for each $i \in [n + 1 - k]$,*

$$\mathbb{P}\big[\overline{\pi} \text{ occurs at position } i \text{ in } \mathbf{C}_{n,p}\big] = P(\overline{\pi}) = \prod_{j=1}^{k} \frac{qp^{j-1}}{1 - p^j} = p^{k(k-1)/2} \prod_{j=1}^{k} \frac{q}{1 - p^j}.$$

Proof. Without loss of generality, let $\pi = 01 \ldots (k - 1)$. We proceed by induction on the length of the pattern. Trivially, $P(\overline{0}) = 1$.

Suppose $k > 1$. Let $\pi_1 = 01 \ldots (k - 2)$, and assume that $P(\overline{\pi_1})$ satisfies the statement of the proposition. Then, by Proposition 33,

$$P(\overline{\pi}) = \sum_{h=0}^{\infty} \mathbb{P}\big[(\mathbf{C}_{n,p}(i) = h) \wedge E_{i+1}^{h+1}(\overline{\pi_1})\big] = \sum_{h=0}^{\infty} qp^h p^{(h+1)(k-1)} P(\overline{\pi_1}) = \frac{qp^{k-1}}{1 - p^k} P(\overline{\pi_1}),$$

as required. $\qquad \square$

If $\overline{\pi}$ is a total ordering pattern of length $k \geqslant 2$, then by Proposition 20, $p \asymp n^{-1/|\pi|} = n^{-2/k(k-1)}$ is the threshold in $\mathbf{C}_{n,p}$ for the arrival of the exact pattern ${}^{=}\overline{\pi}$, with any other exact pattern order-isomorphic to π arriving later. Thus $p \asymp n^{-2/k(k-1)}$ is also the threshold in $\mathbf{C}_{n,p}$ for the appearance of $\overline{\pi}$. However, unlike the situation with exact patterns, a.a.s. a total ordering pattern $\overline{\pi}$ does not disappear as p tends to 1. Indeed, once $q \ll 1$, the relative ordering of k consecutive terms of $\mathbf{C}_{n,p}$ is asymptotically uniformly distributed over the $k!$ permutations of length k.

Proposition 35. *If $\overline{\pi}$ is a total ordering pattern of length k, and $q \ll 1$, then, for each $i \in [n+1-k]$,*

$$\mathbb{P}\big[\overline{\pi} \text{ occurs at position } i \text{ in } \mathbf{C}_{n,p}\big] \sim \frac{1}{k!}.$$

Proof.

$$P(\overline{\pi}) = \prod_{j=1}^{k} \frac{qp^{j-1}}{1-p^j} \sim \prod_{j=1}^{k} \frac{1}{1+p+\ldots+p^{j-1}} \sim \prod_{j=1}^{k} \frac{1}{j} = \frac{1}{k!}. \qquad \square$$

A natural question in the context of total ordering patterns is to determine the length of the longest increasing run in $\mathbf{C}_{n,p}$. If $p \ll n^{-\delta}$ for some $\delta > 0$, then Proposition 20 together with Proposition 35 gives us the following:

Proposition 36. *If $\overline{\pi}$ is a total ordering pattern of length $k \geqslant 2$, then*

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

Once p grows faster than any negative exponential function, then the asymptotic length of the longest increasing run is given by the following proposition:

Proposition 37. *Suppose $n^{-\delta} \ll p \ll 1$ for every positive δ. Then, for any $\varepsilon > 0$,*

$$\mathbb{P}\big[\mathbf{C}_{n,p} \text{ contains an increasing run of length } k\big] \sim \begin{cases} 0, & \text{if } k \geqslant (1+\varepsilon)\sqrt{2\log_{1/p} n}, \\ 1, & \text{if } k \leqslant (1-\varepsilon)\sqrt{2\log_{1/p} n}. \end{cases}$$

Proof. For each $i \in [n+1-k]$, let A_i be the event that $\overline{01\ldots(k-1)}$ occurs at position i in $\mathbf{C}_{n,p}$, and let X be the number of increasing runs of length k in $\mathbf{C}_{n,p}$.

Then, $\mathbb{P}\big[A_i\big] \sim q^k p^{k(k-1)/2}$, and so $\mathbb{E}\big[X\big] \sim nq^k p^{k(k-1)/2}$ if $k \ll n$.

Suppose $\alpha > 1$ and $k = \alpha\sqrt{2\log_{1/p} n}$. Since $\mathbb{E}\big[X\big] \lesssim np^{k(k-1)/2}$, we have

$$\log \mathbb{E}\big[X\big] \lesssim \log n + \alpha^2 \log p \log_{1/p} n - \alpha \log p \sqrt{2\log_{1/p} n}$$

$$= -(\alpha^2 - 1)\log n + \alpha\sqrt{2\log(1/p)\log n}.$$

Now $\log(1/p) \ll \log n$, since $p \gg n^{-\delta}$ for every $\delta > 0$. Thus, $\lim_{n\to\infty} \log \mathbb{E}\big[X\big] = -\infty$. Hence, $\mathbb{E}\big[X\big] \ll 1$ and w.h.p. there are no increasing runs of length k in $\mathbf{C}_{n,p}$.

Now suppose $0 < \alpha < 1$ and $k = \alpha\sqrt{2\log_{1/p} n}$. Since $\mathbb{E}\big[X\big] \gtrsim nq^k p^{k^2/2}$, we have

$$\log \mathbb{E}\big[X\big] \gtrsim \log n + \alpha \log(1-p)\sqrt{2\log_{1/p} n} + \alpha^2 \log p \log_{1/p} n$$

$$\gtrsim (1-\alpha^2)\log n - 2\alpha p\sqrt{2\log_{1/p} n} \gg 1.$$

Distinct events A_i and A_j are correlated if $t = |j - i| < k$. Then,

$$\mathbb{P}\big[A_i \wedge A_j\big] \ \sim \ q^{k+t} p^{(k+t)(k+t-1)/2} \ < \ q^{k+1} p^{k(k+1)/2}.$$

So $\Delta := \sum\limits_{i \sim j} \mathbb{P}\big[A_i \wedge A_j\big] \lesssim nk q^{k+1} p^{k(k+1)/2}$, and

$$R \ := \ \Delta/\mathbb{E}\big[X\big]^2 \ \lesssim \ \frac{k}{nq^{k-1} p^{k(k-3)/2}} \ < \ \frac{k}{nq^k p^{k^2/2}}.$$

Thus,

$$\log R \ \lesssim \ \log k \ - \ \log n \ - \ \alpha \log(1-p) \sqrt{2\log_{1/p} n} \ - \ \alpha^2 \log p \, \log_{1/p} n$$

$$\lesssim \ \log k \ - \ (1 - \alpha^2) \log n \ + \ 2\alpha p \sqrt{2\log_{1/p} n}.$$

So $\lim\limits_{n \to \infty} \log R = -\infty$. Hence, $R \ll 1$ and w.h.p. $\mathbf{C}_{n,p}$ contains an increasing run of length k. $\qquad\square$

Thus, for example, when $p \sim 1/\log n$, a.a.s. the longest increasing run in $\mathbf{C}_{n,p}$ has length close to $\sqrt{2 \log n / \log\log n}$. For the length of the longest increasing run when p is constant or $q \ll 1$, in which case the contribution from $\prod_{j=1}^{k}(1 - p^j) = (p;p)_k$ in the denominator of $P(\overline{01\ldots(k-1)})$ is nontrivial, see the work of Louchard and Prodinger [40].

4.3.2 Consecutive ordering patterns with repeated terms

Let us now move on to ordering patterns, such as $\overline{12201}$, in which at least one term is repeated. Since $P(\overline{\pi})$ is independent of the order of the terms in π, when calculating $P(\overline{\pi})$ we may assume that the terms of π are weakly increasing and that π has the form $0^{\ell_0} 1^{\ell_1} \ldots r^{\ell_r}$, in which the exponents record the number of occurrences of each distinct term (for example, $0^3 1^1 2^2 = 000122$).

Proposition 38. *Suppose $\pi = 0^{\ell_0} 1^{\ell_1} \ldots r^{\ell_r}$. If π has length k, then, for each $i \in [n+1-k]$,*

$$\mathbb{P}\big[\overline{\pi} \text{ occurs at position } i \text{ in } \mathbf{C}_{n,p}\big] \ = \ P(\overline{\pi}) \ = \ \prod_{j=0}^{r} \frac{q^{\ell_j} p^{s_{j+1}}}{1 - p^{s_j}} \ = \ q^{k-(r+1)} p^{|\pi|} \prod_{j=0}^{r} \frac{q}{1 - p^{s_j}},$$

where $s_j = \sum\limits_{i=j}^{r} \ell_i$ for each $j = 0, \ldots, r+1$ (with $s_{r+1} = 0$).

Proof. The proof is by induction on r, and is very similar to that for Proposition 34. If $r = 0$, then $s_0 = \ell_0 = k$ and

$$P(\overline{\pi}) \ = \ \sum_{h=0}^{\infty} (qp^h)^k \ = \ \frac{q^{\ell_0}}{1 - p^{s_0}}.$$

Suppose now that $r > 1$. Let $\pi_1 = 0^{\ell_1} 1^{\ell_2} \ldots (r-1)^{\ell_r}$, and assume that $P(\overline{\pi_1})$ satisfies the statement of the proposition. Then, using Proposition 33,

$$P(\overline{\pi}) \;=\; \sum_{h=0}^{\infty} \mathbb{P}\left[\,^=\overline{h^{\ell_0}} \text{ occurs at position } i \text{ in } \mathbf{C}_{n,p} \text{ and } E_{i+\ell_0}^{h+1}(\overline{\pi_1}) \,\right]$$

$$= \sum_{h=0}^{\infty} (qp^h)^{\ell_0} \, p^{(h+1)s_1} P(\overline{\pi_1}) \;=\; \frac{q^{\ell_0} p^{s_1}}{1 - p^{s_0}} P(\overline{\pi_1}),$$

as required. $\qquad\square$

If π is nonzero, then by Proposition 20, $p \asymp n^{-1/|\pi|}$ is the threshold in $\mathbf{C}_{n,p}$ for the arrival of the exact pattern $^=\overline{\pi}$, with any other exact pattern order-isomorphic to π arriving later. Thus $p \asymp n^{-1/|\pi|}$ is also the threshold in $\mathbf{C}_{n,p}$ for the appearance of the ordering pattern $\overline{\pi}$. However, unlike in the case of total ordering patterns, an ordering pattern with a repeated term also exhibits a threshold for its disappearance:

Proposition 39. *Suppose the multiset of terms in π is $\{0^{\ell_0}, 1^{\ell_1}, \ldots, r^{\ell_r}\}$, where $\ell_j > 1$ for at least one term j. If π has length k and $d = k - (r+1)$, then for any positive constant α,*

$$\mathbb{P}\left[\mathbf{C}_{n,p} \text{ contains } \overline{\pi}\right] \;\sim\; \begin{cases} 1, & \text{if } 1 \gg q \gg n^{-1/d}, \\ 1 - e^{-\alpha^d/\lambda}, & \text{if } q \sim \alpha n^{-1/d}, \\ 0, & \text{if } n^{-1/d} \gg q, \end{cases}$$

where $\displaystyle \lambda = \prod_{j=0}^{r} \sum_{i=j}^{r} \ell_i$.

Proof. For each $i \in [n+1-k]$, let A_i be the event that $\overline{\pi}$ occurs at position i in $\mathbf{C}_{n,p}$, and let X be the number of occurrences of $\overline{\pi}$ in $\mathbf{C}_{n,p}$. If $q \ll 1$, then

$$\mathbb{P}\left[A_i\right] \;\sim\; q^d \prod_{j=0}^{r} \frac{1-p}{1-p^{s_j}} \;=\; q^d \prod_{j=0}^{r} \frac{1}{1 + p + \ldots + p^{s_j - 1}} \;\sim\; q^d \prod_{j=0}^{r} \frac{1}{s_j} \;=\; \frac{q^d}{\lambda},$$

where $s_j = \displaystyle\sum_{i=j}^{r} \ell_i$ for each $j = 0, \ldots, r+1$. So $\mathbb{E}\left[X\right] \sim nq^d/\lambda$.

If $q \ll n^{-1/d}$, then $\mathbb{E}\left[X\right] \ll 1$ and w.h.p. $\overline{\pi}$ doesn't occur in $\mathbf{C}_{n,p}$.

Distinct events A_i and A_j are correlated if $t = |j - i| < k$. There may be several ways in which A_i and A_j may consistently overlap, corresponding to a variety of longer ordering patterns. For example, if $\pi = 010$, then there are three possible arrangements: 01010, 01020 and 02010 (each with $t = 2$). However, for each t, there are certainly never more than $(k+t)^t < (2k)^k$ possibilities.

Thus,

$$\Delta := \sum_{i \sim j} \mathbb{P}[A_i \wedge A_j] \leqslant \kappa n q^{d+1} \quad \text{and} \quad R := \Delta / \mathbb{E}[X]^2 \lesssim \frac{\kappa \lambda^2}{n q^{d-1}},$$

for some constant κ, dependent only on π.

Moreover, $\Lambda := \sum_i \mathbb{P}[A_i]^2 + \sum_{i \sim j} \mathbb{P}[A_i] \mathbb{P}[A_j] \sim n k q^{2d}/\lambda^2$.

Suppose $q = \omega n^{-1/d} \ll 1$ for some $\omega \gg 1$, then $\mathbb{E}[X] \sim \omega^d/\lambda \gg 1$ and $R \lesssim \kappa \lambda^2 q/\omega^d \ll 1$. So w.h.p. $\overline{\pi}$ occurs in $\mathbf{C}_{n,p}$.

Finally, suppose $q = \alpha n^{-1/d}$. Then $\mathbb{E}[X] \sim \alpha^d/\lambda$ and $\Delta \leqslant \alpha^d \kappa q \ll 1$, and we have $\Lambda \sim \alpha^{2d} k/\lambda^2 n \ll 1$, so the number of occurrences of $\overline{\pi}$ is asymptotically Poisson with mean α^d/λ. $\qquad \square$

Thus, for example, the threshold for the disappearance from $\mathbf{C}_{n,p}$ of *balanced peaks* $\overline{010}$ is at $q \asymp n^{-1}$, as is the threshold for the disappearance of *balanced valleys* $\overline{101}$. However, at this threshold, $\mathbf{C}_{n,p}$ is more likely to contain a balanced valley than it is to contain a balanced peak.

A composition in which no pair of adjacent terms are equal (that is, avoiding the ordering pattern $\overline{00}$) is called a *Carlitz composition*. These have been well-studied [21, 34, 37, 39]. More generally, the distribution of the lengths of runs of consecutive equal terms in compositions has been investigated [23, 46]. From our evolutionary perspective, we have the following threshold for the disappearance of runs of equal terms (a direct consequence of Proposition 39):

Proposition 40. *For any fixed $k \geqslant 2$ and positive constant α,*

$$\mathbb{P}[\mathbf{C}_{n,p} \text{ contains a run of } k \text{ equal terms}] \sim \begin{cases} 1, & \text{if } 1 \gg q \gg n^{-1/(k-1)}, \\ 1 - e^{-\alpha^{k-1}/k}, & \text{if } q \sim \alpha n^{-1/(k-1)}, \\ 0, & \text{if } n^{-1/(k-1)} \gg q. \end{cases}$$

Comparison with Proposition 22 shows that this threshold is the same as for the disappearance of runs of equal *nonzero* terms. In particular, the threshold for $\mathbf{C}_{n,p}$ to become a Carlitz composition is at $q \asymp n^{-1}$.

4.4 Nonconsecutive patterns

In this concluding section we finally remove the restriction that the occurrence of a pattern must be consecutive. Nonconsecutive patterns are represented without an overline. We begin with exact patterns. The pattern ${}^{=}r_1 \ldots r_k$ occurs in a composition C if there exists a sequence of indices $i_1 < i_2 < \ldots < i_k$ such that for each $j \in [k]$, we have $C(i_j) = r_j$. The threshold for the appearance of a nonconsecutive exact pattern depends only on its maximum term. In contrast, all such patterns share the same threshold for their disappearance.

Proposition 41. *If $^=\pi$ is a nonzero exact pattern whose largest term is r, then*

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

Proof. If $p \ll n^{-1/r}$ then, by Proposition 20, w.h.p. $\mathbf{C}_{n,p}$ has no term equal to r and thus contains no occurrence of $^=\pi$. Also, by Proposition 31, if $q \ll n^{-1}$ then w.h.p. the smallest term of $\mathbf{C}_{n,p}$ exceeds any fixed value, so there is no occurrence of $^=\pi$.

Suppose π has length k. For each $j \in [k]$, let $i_j = \lfloor jn/k \rfloor$ and $\mathbf{C}_j$ be the sub-composition $\big(\mathbf{C}_{n,p}(i_{j-1}+1),\ldots,\mathbf{C}_{n,p}(i_j)\big)$, so that $\mathbf{C}_1,\ldots,\mathbf{C}_k$ is a partition of the terms of $\mathbf{C}_{n,p}$. If $n^{-1/r} \ll p$ and $q \gg n^{-1}$, then, for each $j \in [k]$, by Proposition 20, w.h.p. $\mathbf{C}_j$ has a term equal to $\pi(j)$, so $\mathbf{C}_{n,p}$ contains an occurrence of $^=\pi$. $\square$

Nonconsecutive upper and lower patterns exhibit analogous behaviour. The proofs are very similar to that for Proposition 41, so are omitted.

Proposition 42. *If $^\geqslant\pi$ is a nonzero upper pattern whose largest term is r, then*

$$\mathbb{P}\big[\mathbf{C}_{n,p} \ contains \ ^\geqslant\pi\big] \ \sim \ \begin{cases} 0, & if \ p \ll n^{-1/r}, \\ 1, & if \ n^{-1/r} \ll p. \end{cases}$$

Proposition 43. *If $^\leqslant\pi$ is a lower pattern, then*

$$\mathbb{P}\big[\mathbf{C}_{n,p} \ contains \ ^\leqslant\pi\big] \ \sim \ \begin{cases} 1, & if \ q \gg n^{-1}, \\ 0, & if \ n^{-1} \gg q. \end{cases}$$

We can generalise these results to exact *vincular* patterns, in which there are some adjacency requirements, represented by an overline (or *vinculum*). For example, C contains the exact vincular pattern $^=\overline{12}3$ if there exist indices i and $j > i+1$ such that $C(i) = 1$, $C(i+1) = 2$ and $C(j) = 3$. We consider a vincular pattern to consist of a number of consecutive *blocks*, the terms of which must be adjacent in any occurrence. For example, $^=\overline{12}0\,4\overline{003}$ consists of four blocks, of lengths 2, 1, 1 and 3, and sizes 3, 0, 4 and 3, respectively.

The threshold for the appearance of an exact vincular pattern depends on the *size* of its *largest* block, whereas the threshold for its disappearance depends on the *length* of its *longest* block. This is analogous to the situation for exact consecutive patterns (Proposition 20). We omit the proof since it is exactly analogous to that for Proposition 41 (with one part in the partition of $\mathbf{C}_{n,p}$ for each block in the pattern).

Proposition 44. *If $^=\pi$ is an exact vincular pattern whose largest block has size s and whose longest block has length ℓ, then*

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

Let's now consider nonconsecutive total ordering patterns (ordering patterns with distinct terms, such as 102 or 3120). The threshold for their appearance depends only on their length (or equivalently on their largest term, which is one less than their length). In the case of constant p, the distribution of the number of inversions (occurrences of 10) has previously been investigated by Prodinger [44] and by Heubach, Knopfmacher, Mays and Munagi [25].

Proposition 45. *If π is a total ordering pattern of length k, then*

$$\mathbb{P}\big[\mathbf{C}_{n,p} \text{ contains } \pi\big] \ \sim \ \begin{cases} 0, & \text{if } p \ll n^{-1/(k-1)}, \\ 1, & \text{if } n^{-1/(k-1)} \ll p. \end{cases}$$

Proof. By Proposition 41, $p \asymp n^{-1/(k-1)}$ is the threshold for the appearance of the exact pattern ${}^{=}\pi$, which doesn't disappear until $q \asymp n^{-1}$. And by Proposition 36, the consecutive total ordering pattern $\overline{\pi}$ a.a.s. appears when $p \asymp n^{-2/k(k-1)}$, and doesn't disappear. The result follows because an occurrence of ${}^{=}\pi$ is an occurrence of π, as is an occurrence of $\overline{\pi}$. $\qquad\square$

Our last result concerns equal terms in $\mathbf{C}_{n,p}$. In particular, Proposition 46 gives a threshold at $q \asymp n^{-2}$ for the disappearance of (nonconsecutive) ordering pattern 00. Thus, when $q \ll n^{-2}$ a.a.s. no two terms of $\mathbf{C}_{n,p}$ are equal, and so nonconsecutive total ordering patterns satisfy the same asymptotic distribution in $\mathbf{C}_{n,p}$ as they do in a random permutation. Hence, when $q \ll n^{-2}$ the distribution of the length of the longest increasing subsequence in $\mathbf{C}_{n,p}$ is as described in the celebrated work of Baik, Deift and Johansson [4] (see Romik [45] for an extended expository presentation).

Proposition 46. *For any fixed $k \geqslant 2$ and positive constant α,*

$$\mathbb{P}\big[\mathbf{C}_{n,p} \text{ has at least } k \text{ equal terms}\big] \ \sim \ \begin{cases} 1, & \text{if } q \gg n^{-k/(k-1)}, \\ 1 - e^{-\alpha^{k-1}/(k \times k!)}, & \text{if } q \sim \alpha n^{-k/(k-1)}, \\ 0, & \text{if } n^{-k/(k-1)} \gg q. \end{cases}$$

Proof. By Proposition 14, if $q \gg n^{-1/k}$ then $\mathbf{C}_{n,p}$ contains a gap of length at least k (an occurrence of ${}^{=}\overline{0^k}$) and thus has at least k equal (zero) terms. Suppose now that $q \ll 1$.

Given a vector $\mathbf{i} := (i_1, i_2, \ldots, i_k) \in [n]^k$ such that $i_{j+1} \geqslant i_j$ for each $j \in [k-1]$, let $A_{\mathbf{i}}$ be the event that $\mathbf{C}_{n,p}(i_1) = \mathbf{C}_{n,p}(i_2) = \ldots = \mathbf{C}_{n,p}(i_k)$. Then

$$\mathbb{P}\big[A_{\mathbf{i}}\big] \ = \ \sum_{r=0}^{\infty} q^k p^{kr} \ = \ \frac{q^k}{1 - p^k} \ \sim \ \frac{q^{k-1}}{k} \qquad (q \ll 1).$$

If X is the total number of k-tuples of equal-valued terms, then

$$\mathbb{E}\big[X\big] \ = \ \binom{n}{k} \mathbb{P}\big[A_{\mathbf{i}}\big] \ \sim \ \frac{n^k q^{k-1}}{k \times k!}.$$

So if $q \ll n^{-k/(k-1)}$, then $\mathbb{E}[X] \ll 1$ and w.h.p. there is no occurrence of 0^k in $\mathbf{C}_{n,p}$.

Distinct events $A_{\mathbf{i}}$ and $A_{\mathbf{j}}$ are correlated ($\mathbf{i} \sim \mathbf{j}$) if the indices in $\mathbf{i}$ and $\mathbf{j}$ have nonempty intersection. Thus,

$$\Delta := \sum_{\mathbf{i} \sim \mathbf{j}} \mathbb{P}[A_{\mathbf{i}} \wedge A_{\mathbf{j}}] \sim \sum_{t=k+1}^{2k-1} \binom{n}{t}\binom{t}{k}\frac{q^{t-1}}{t} \lesssim \sum_{t=k+1}^{2k-1} n^t q^{t-1}.$$

Hence,

$$R := \Delta/\mathbb{E}[X]^2 \lesssim \sum_{t=1}^{k-1} n^{-t} q^{-(t-1)}.$$

Suppose $q = \omega n^{-k/(k-1)}$ for some $\omega \gg 1$. Then $\mathbb{E}[X] \gg 1$ and

$$R \lesssim \sum_{t=1}^{k-1} n^{-\frac{t-k}{t-1}} \omega^{-(t-1)} \ll 1.$$

Hence, a.a.s. there is an occurrence of 0^k in $\mathbf{C}_{n,p}$.

Finally,

$$\Lambda := \sum_{\mathbf{i}} \mathbb{P}[A_{\mathbf{i}}]^2 + \sum_{\mathbf{i} \sim \mathbf{j}} \mathbb{P}[A_{\mathbf{i}}]\mathbb{P}[A_{\mathbf{j}}] \sim \sum_{t=k}^{2k-1} \binom{n}{t}\binom{t}{k}\left(\frac{q^{k-1}}{k}\right)^2 \lesssim q^{2k-2}\sum_{t=k}^{2k-1} n^t.$$

Suppose $q \sim \alpha n^{-k/(k-1)}$. Then $\mathbb{E}[X] \sim \alpha^{k-1}/(k \times k!)$, and

$$\Delta \lesssim \sum_{t=1}^{k-1} n^{-t/(k-1)} \ll 1 \qquad \text{and} \qquad \Lambda \lesssim \sum_{t=1}^{k-1} n^{-t} \ll 1.$$

Thus the number of occurrences of 0^k is seen to be asymptotically Poisson with mean $\alpha^{k-1}/(k \times k!)$. $\square$

Acknowledgements

References

[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] Jinho Baik, Percy Deift, and Kurt Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.*, 12(4):1119–1178, 1999.

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

[6] 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.

[7] 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.

[8] B. Bollobás and A. Thomason. Threshold functions. *Combinatorica*, 7(1):35–38, 1987.

[9] Béla Bollobás. *Random Graphs*. Cambridge University Press, second edition, 2001.

[10] Ai Lian Chen, Fu Ji Zhang, and Hao Li. The degree distribution of the random multigraphs. *Acta Math. Sin. (Engl. Ser.)*, 28(5):941–956, 2012.

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

[12] 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.

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

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

[15] Serkan Eryilmaz. A note on runs of geometrically distributed random variables. *Discrete Math.*, 306(15):1765–1770, 2006.

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

[17] Ehud Friedgut and Gil Kalai. Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.*, 124(10):2993–3002, 1996.

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

[19] Ayla Gafni. Longest run of equal parts in a random integer composition. *Discrete Math.*, 338(2):236–247, 2015.

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

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

[22] 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.

[23] Peter J. Grabner, Arnold Knopfmacher, and Helmut Prodinger. Combinatorics of geometrically distributed random variables: run statistics. *Theoret. Comput. Sci.*, 297(1–3):261–270, 2003.

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

[25] S. Heubach, A. Knopfmacher, M. E. Mays, and A. Munagi. Inversions in compositions of integers. *Quaest. Math.*, 34(2):187–202, 2011.

[26] Silvia Heubach and Toufik Mansour. Counting rises, levels, and drops in compositions. *Integers*, 5(1): A11, 2005.

[27] Silvia Heubach and Toufik Mansour. Enumeration of 3-letter patterns in compositions. In B. Landman, M. Nathanson, J. Nešetřil, R. Nowakowski, and C. Pomerance, editors, *Combinatorial number theory*, pages 243–264. de Gruyter, 2007.

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

[29] Lars Holst. A unified approach to limit theorems for urn models. *J. Appl. Probab.*, 16(1):154–162, 1979.

[30] Thierry E Huillet. A Bose–Einstein approach to the random partitioning of an integer. *J. Statistical Mechanics: Theory and Experiment*, 2011(08): P08021, 2011.

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

[32] Svante Janson. Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probab. Surv.*, 9:103–252, 2012.

[33] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random Graphs*. Wiley, 2000.

[34] 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.

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

[36] Sergey Kitaev, Tyrrell B. McAllister, and T. Kyle Petersen. Enumerating segmented patterns in compositions and encoding by restricted permutations. *Integers*, 6: A34, 2006.

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

[38] Guy Louchard. Runs of geometrically distributed random variables: a probabilistic analysis. *J. Comput. Appl. Math.*, 142(1):137–153, 2002.

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

[40] Guy Louchard and Helmut Prodinger. Ascending runs of sequences of geometrically distributed random variables: a probabilistic analysis. *Theoret. Comput. Sci.*, 304(1–3):59–86, 2003.

[41] Guy Louchard, Helmut Prodinger, and Mark Daniel Ward. The number of distinct values of some multiplicity in sequences of geometrically distributed random variables. In *2005 International Conference on Analysis of Algorithms*, Discrete Math. Theor. Comput. Sci. Proc., AD, pages 231–256. 2005.

[42] Frosso S. Makri and Zaharias M. Psillakis. On success runs of a fixed length in Bernoulli sequences: exact and asymptotic results. *Comput. Math. Appl.*, 61(4):761–772, 2011.

[43] Jinyoung Park. Threshold phenomena for random discrete structures. *Notices Amer. Math. Soc.*, 70(10):1615–1625, 2023.

[44] Helmut Prodinger. Combinatorics of geometrically distributed random variables: inversions and a parameter of Knuth. *Ann. Comb.*, 5(2):241–250, 2001.

[45] Dan Romik. *The surprising mathematics of longest increasing subsequences*. Cambridge University Press, 2015.

[46] Herbert S. Wilf. The distribution of run lengths in integer compositions. *Electron. J. Combin.*, 18(2):#P23, 2011.