Asymptotic Distributions for Block Statistics on Non-crossing Partitions

by

Boyu Li

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Pure Mathematics

Waterloo, Ontario, Canada, 2013

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The set of non-crossing partitions was first studied by Kreweras in 1972 and was known to play an important role in combinatorics, geometric group theory, and free probability. In particular, it has a natural embedding into the symmetric group, and there is an extensive literature on the asymptotic cycle structures of random permutations. This motivates our study on analogous results regarding the asymptotic block structure of random non-crossing partitions.

We first investigate an analogous result of the asymptotic distribution for the total number of cycles of random permutations due to Goncharov in 1940's: Goncharov showed that the total number of cycles in a random permutation is asymptotically normally distributed with mean $\log(n)$ and variance $\log(n)$. As a analog of this result, we show that the total number of blocks in a random non-crossing partition is asymptotically normally distributed with mean $\frac{n}{2}$ and variance $\frac{n}{8}$.

We also investigate the outer blocks, which arise naturally from non-crossing partitions and has many connections in combinatorics and free probability. It is a surprising result that among many blocks of non-crossing partitions, the expected number of outer blocks is asymptotically 3. We further computed the asymptotic distribution for the total number of blocks, which is a shifted negative binomial distribution.

Acknowledgements

I would like to thank Professor Alexandru Nica for opening this interesting topic and supervising me over this research. I would like to thank Professor Bruce Richmond and Professor Kevin Hare for their constructive comments on this thesis.

I would also like to thank Betty Xing for being so supportive of me during the past few years.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Non-crossing Partitions

The collection $NC(n)$ of all non-crossing partitions of the set $\{1, 2, \cdots, n\}$ was first studied by Kreweras [9] in 1972, as an interesting example of partially ordered set. An element $\pi \in NC(n)$ is a partition $\pi = \{V_1, V_2, \cdots, V_k\}$ of $\{1, 2, \cdots, n\}$, where $V_1, V_2, \cdots, V_k$ (called the blocks of π) are non-empty pairwise disjoint sets such that $V_1 \bigcup V_2 \bigcup \cdots \bigcup V_k = \{1, 2, \cdots, n\}$. Moreover, for any $i \neq j$, we require that V_i, V_j do not cross; in other words, it is not possible to have $a < b < c < d$ where $a, c \in V_i$ and $b, d \in V_j$. An example of a non-crossing partition $\pi \in NC(9)$ is shown in Figure 1.1.1.

The point of Kreweras's paper is that $NC(n)$ has an interesting structure as a poset (short for partially ordered set), where the partial order is given by reverse refinement: we say $\pi \leq \rho$ whenever every block of ρ is a union of some blocks in π. Kreweras studies several properties of this partially ordered set, including the Möbius function on the line of Rota's program started in [15] (where Rota considered the Möbius function on different posets, including the poset $\mathcal{P}(n)$ of all partitions on $\{1, 2, \cdots, n\}$). In certain respect, $NC(n)$ turns

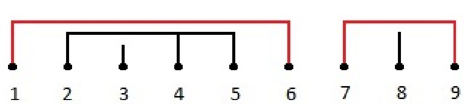

Figure 1.1.1: An Example of $\pi \in NC(9)$

out to be a nicer poset than $\mathcal{P}_n$, because it has a nice symmetric structure in contrary to the set of partitions. The symmetry is given by an order-reserving automorphism on $NC(n)$, which is known as the Kreweras complement. The enumerative properties of non-crossing partitions was studied by several other authors after Kreweras, and one may refer to the survey paper [17] by R. Simion in 2000.

More recently, $NC(n)$ has caught the attention of people studying geometric group theory. This is due to the fact that $NC(n)$ can be naturally embedded into the Cayley graph of the symmetric group $\mathcal{S}_n$, where the set of generators for $\mathcal{S}_n$ is the set of all transpositions (i, j) with $1 \leq i < j \leq n$. The embedding can be described by regarding each block V_i of a non-crossing partition π as a cycle, and writing the numbers of V_i in increasing order. For example, the non-crossing partition π shown in Figure 1.1.1 can be regarded as the permutation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 6 & 4 & 3 & 5 & 2 & 1 & 9 & 8 & 7 \end{pmatrix}$$

There is a natural partial order on the Cayley graph of $\mathcal{S}_n$, and this embedding of $NC(n)$ turns out to be an order preserving isomorphism between $NC(n)$ and an interval in the symmetric group. People in geometric group theory have studied the analog of such an interval in other Weyl groups, and (starting with Reiner [13]) have found other posets which can be thought of as non-crossing partitions. For this topic, one can refer to J. McCammond's survey paper [10] on the connection between non-crossing partitions and geometric group.

Another area where $NC(n)$ has found a role recently is the development of free probability, and in connection to that, the one of random matrix theory. In free probability, $NC(n)$ plays a critical role in the description of the R-transform, which is the free probability counterpart of the usual Fourier transform. In the study of random matrices, $NC(n)$ appears naturally in the leading terms in various trace formulas. A more detailed description on the role of $NC(n)$ in free probability can be found in the Lecture 16 and 23 in the book [12] by A. Nica and R. Speicher.

1.2 Asymptotic Results on Random Non-crossing Partitions

The motivation of this thesis comes from the fact that $NC(n)$ embed canonically into the symmetric group $\mathcal{S}_n$, and that on the other hand, there is a rather extensive literature (going back at least to the 1940's) on the cycle structure of random permutation. For a clear and concise presentation of this topic, we refer to Section XIV.4 of the book of B. Bollobas [2]. By a random permutation, we simply understand an element of $\mathcal{S}_n$ is assigned atomic measure of $\frac{1}{n!}$. When $NC(n)$ embeds into $\mathcal{S}_n$, the blocks of non-crossing partitions become cycles of permutations. It stands to reason that there should be analogous results concerning the block structure of a random $\pi \in NC(n)$. The results about random non-crossing partition will not, however, follow directly from the known results about $\mathcal{S}_n$ because the relative size of $NC(n)$ inside $\mathcal{S}_n$ goes to 0 as $n \to \infty$.

The first test question that one can ask is about asymptotic distribution of the total number of blocks in a random $\pi \in NC(n)$. For random permutations, Goncharov [7, 8] showed that the total number of cycles in a random permutation is asymptotically normally distributed with mean $\log(n)$ and variance $\log(n)$. It is clear that we cannot have $\log(n)$ as the asymptotic expectation, because the symmetry given by the Kreweras complement immediately results in an asymptotic expectation around $\frac{n}{2}$. In fact, we shall prove, in Chapter 3:

Theorem 1.2.1. *Let $X_n : NC(n) \to \mathbb{N}$ count the total number of blocks in a non-crossing partition. Then*

$$\frac{X_n - n/2}{\sqrt{n/8}} \xrightarrow{d} \mathcal{N}(0,1)$$

We also investigate an object arise naturally from non-crossing partition, which we we call them the outer blocks. For $\pi = \{V_1, V_2, \cdots, V_k\} \in NC(n)$, a block V_i is called an *inner block* if there exists $a, b \in V_j,\ j \neq i$ such that for all $v \in V_i,\ a < v < b$. If a block is not inner, we call it an *outer block*. The word 'outer' comes from the following observation: If we draw a non-crossing partition $\pi \in NC(n)$ on a line, the inner block are those enclosed by some other blocks; the outer blocks are those blocks not enclosed by any other blocks. For example, the $\pi \in NC(9)$ shown in Figure 1.1.1 has two outer blocks: $\{1, 6\}$ and $\{7, 9\}$ (also colored in red). Outer block arise naturally as we study the block structure of non-crossing partitions, and are related to some other combinatorial objects as well. They also play a special role in various summation formulas used in free probability [3, 11]. It is quite surprising that a random non-crossing partition cannot have too many outer blocks as the asymptotic expectation of the total number of outer blocks is 3. We established the following result regarding the asymptotic distribution of the total number of outer blocks:

Theorem 1.2.2. *Let $Y_n : NC(n) \to \mathbb{N}$ count the total number of outer blocks in a non-crossing partition. Let ν be a probability measure with mass $\frac{k}{2 \cdot 2^k}$ at point $k = 1, 2, \cdots$. Then $Y_n \xrightarrow{d} \nu$.*

Beside the present introduction, the thesis is divided into 4 chapters. In Chapter 2, we shall give a review of background on non-crossing partition and asymptotic distributions. Then in Chapter 3 and 4, we develop the asymptotic results on the total number of blocks and total number of outer blocks that are announce in Theorem 1.2.1 and 1.2.2 respectively. Finally, as this study also opens the door for many interesting topics on non-crossing partitions, we shall briefly discuss some topics for further research in Chapter 5.

Chapter 2

Background

2.1 Non-crossing Partitions

This section will discuss the concept of non-crossing partition in more detail. Let us first introduce the formal definition and some notations.

Definition 2.1.1. Fix an positive integer n,

1. A partition π of the set $\{1, 2, \cdots, n\}$ is a collection $\{V_1, V_2, \cdots, V_k\}$, where $V_i \subseteq \{1, 2, \cdots, n\}$ are pairwise disjoint non-empty sets with $\bigcup_{i=1}^{k} V_i = \{1, 2, \cdots, n\}$. The sets V_i are called the blocks of π.

2. A non-crossing partition of $\{1, 2, \cdots, n\}$ is a partition $\pi = \{V_1, V_2, \cdots, V_k\}$ of $\{1, 2, \cdots, n\}$, where there is no $a < b < c < d$, such that $a, c \in V_i$, $b, d \in V_j$, and $i \neq j$.

3. The set of non-crossing partition for $\{1, \cdots, n\}$ is denoted by $NC(n)$.

The set of non-crossing partitions is one of the many structures that can be enumerated by Catalan numbers (one may refer to Exercise 6.19 in [19], where Stanley gave 66 such structures). Catalan number is a well-known number in combinatorics, and we shall use the following facts later on.

Lemma 2.1.2. *Let C_n be the n-th Catalan number, where $C_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{(n+1)!n!}$*

1. $C_n = \sum_{i=0}^{n-1} C_i C_{n-i-1}$

2. *Let* $C(z) = \sum_{n=0}^{\infty} C_n z^n$*, then* $C(z)$ *satisfies* $1 + zC(z)^2 = C(z)$*. As a result,* $C(z) = \frac{1-\sqrt{1-4z}}{z}$ *with radius of convergence* $1/4$*.*

The Catalan number appears in various enumeration problems, and it can be shown that there are bijections between the non-crossing partitions and some other objects which are known to be enumerated by Catalan numbers. Here, we pick two of them as an illustrative example.

Definition 2.1.3. A Dyck path of length $2n$ is a path on $\mathbb{Z}^2$, starting at $(0,0)$ and ending at $(2n,0)$. Each step of the path is either $(1,1)$ (a up move), or $(1,-1)$ (a down move). Moreover, the path is always above the x-axis. The collection of all these paths is D_n.

It is a well-known result that the total number of Dyck path of length $2n$ is exactly the n-th Catalan number C_n.

Definition 2.1.4. A Non-crossing pairing π of the set $\{1, 2, \cdots, 2n\}$ is a non-crossing partition $\{(a_i, b_i)\}_{i=1}^{n}$ of $\{1, 2, \cdots, 2n\}$. In other words, it is a non-crossing partition where every block has exactly two elements. The set of all non-crossing pairings for $\{1, 2, \cdots, 2n\}$ is denoted as $NCP(n)$.

Proposition 2.1.5. *There exists bijections between* D_n*,* $NCP(n)$*, and* $NC(n)$*. In particular, all of these three sets have cardinality* C_n*, the* n*-th Catalan number*

Proof. To go from D_n to $NCP(n)$, let us denote a up move by U and a down move by D, so that each Dyck path can be represented as a sequence of U, D where any subsequence starting from the first step must have at least as many U as D.

Define the map from D_n to $NCP(n)$ as follow: For each $i = 1, 2, 3, ..., 2n$, if the i-th step is U, then i is placed inside a pair with $j > i$ where j-th step is D and the sub-string from i to j has the equal number of U and D. One can easily check this is indeed a bijection. In fact, we can describe its inverse explicitly: given any non-crossing pairing $P = \{(a_i, b_i)\}_{i=1}^{n}$ where $a_i < b_i$, map it to a Dyck path where each step at a_i is U and each step at b_i is D.

Now to go from $NCP(n)$ to $NC(n)$, for a non-crossing pairing P on $1, 2, \cdots, 2n$, define a non-crossing partition π using the following procedure: for any $i, j \in \{1, 2, \cdots, n\}$, i, j is in the same block of π if and only if there is no pair (a, b) in P such that $a \leq 2i - 1$ and $b \geq 2i$. It is not hard to check this is a bijection.

□

One can make $NC(n)$ into a partially ordered set via the following definition:

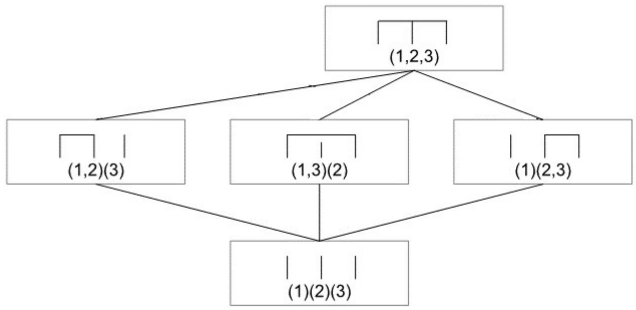

Figure 2.1.1: The Hasse Diagram of $NC(3)$

Definition 2.1.6. For $\pi, \tau \in NC(n)$, we say π is an refinement of τ if for every block V in π, V is also contained in some block of τ. This is denoted by $\pi \prec \tau$.

A quick observation is that this refinement defines a partial order on $NC(n)$.

Remark 2.1.7. Once the partial order is defined, we can define the notion of cover: for two non-crossing partitions $\pi, \sigma \in NC(n)$, we call π covers σ if there is no other $\tau \in NC(n)$ such that $\pi \prec \tau \prec \sigma$. We can consider the Hasse diagram of this partially ordered set, which is a undirected graph G whose vertex set is $NC(n)$ and (π, σ) is an edge of G if π covers σ or vice versa. For example, the Hasse diagram of $NC(3)$ is shown in Figure 2.1.1:

Remark 2.1.8. The notion of cover allows us to define the rank of the non-crossing partitions: we denote the non-crossing partition $e = \{\{1\}, \{2\}, \cdots, \{n\}\}$, the minimum element in this poset, to have rank 0; for any other $\pi \in NC(n)$, define the rank of π to be the length of the shortest path from e to π in the Hasse diagram. In other words, the rank of π is the number of covers required to reach π from e.

Remark 2.1.9. Recall for $\pi \in NC(n)$, $X_n(\pi)$ denotes the total number of blocks in π. One can observe that the rank of a non-crossing partition π is exactly $n - X_n(\pi)$.

The Hasse diagram of non-crossing partitions is particularly nice in the sense that there exists a nice symmetry given by the so-called Kreweras complement map.

Definition 2.1.10. (Kreweras Complement) Let $\pi \in NC(n)$ be a non-crossing partition on points $\{1, 2, \cdots, n\}$. The Kreweras complement $K(\pi)$ is defined to be the biggest non-crossing partition on points $\{1', 2', \cdots, n'\}$, such that $\pi \bigcup K(\pi)$ is also a non-crossing partition on points $1, 1', 2, 2', \cdots, n, n'$.

Example 2.1.11. Let us consider the non-crossing partition $\pi \in NC(6)$ where $\pi = \{\{1,4\},\{2\},\{3\},\{5,6\}\}$:

To compute its Kreweras complement, we first put nodes 1,1′,2,2′,$\cdots$,6,6′ in order, and connect $\{1,2,\cdots,6\}$ according to π. Now we notice that the points $1',2',3'$ can be placed in one block without creating any crossing with blocks in π, but none of the other three points can be added to this block. Hence $\{1',2',3'\}$ is one block of $K(\pi)$. Similarly, we can verify that $K(\pi)$ contains two other blocks $\{4,6\}$ and $\{5\}$. Figure 2.1.2 gives an illustration of π (black) and $K(\pi)$ (red).

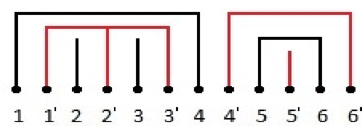

Figure 2.1.2: The Kreweras Complement of π

Remark 2.1.12. It can be shown that the Kreweras complement map K is an order reversing automorphism on $NC(n)$. As a result, for any $\pi \in NC(n)$, $rank(\pi) + rank(K(\pi)) = n-1$. One can immediately deduce that $X_n(\pi) + X_n(K(\pi)) = n+1$.

2.2 Narayana Number and Lukasiewicz Path

Kreweras, when he first investigated non-crossing partitions, also enumerates the number of non-crossing partitions with a given number of blocks [9]. We are going to present a proof of this result based on the method discussed in [12] using the Lukasiewicz path.

Definition 2.2.1.

1. An almost-rising path is a lattice path in $\mathbb{Z}^2$, starting at $(0,0)$. Each step, the path moves in the direction $(1,i)$ where $i \geq -1$ is an integer. We call a step of the path falling if $i=-1$.

2. A Lukasiewicz path is an almost-rising path γ, which ends on the x-axis and never goes below the x-axis. We denote the set of all the Lukasiewicz path ending at point $(n,0)$ by $Luk(n)$.

Definition 2.2.2. For almost-rising paths γ of n-steps, one can completely describe this path using a vector $(\lambda_1,\lambda_2,\cdots,\lambda_n)$, $\lambda_i \geq -1$. Here, the i-th step in γ is in the direction $(1,\lambda_i)$. This vector is called the rise-vector of γ.

Remark 2.2.3. For an almost-rising path γ with rise-vector $(\lambda_1, \lambda_2, \cdots, \lambda_n)$, $\gamma \in Luk(n)$ if and only if $\sum_{k=1}^{n} \lambda_k = 0$ and $\sum_{k=1}^{i} \lambda_i \geq 0$ for all $i = 1, 2, \cdots, n$.

It turns out that $Luk(n)$ is also in one-to-one correspondence with the set of non-crossing partitions $NC(n)$ in the way described below:

Proposition 2.2.4. *Fix a positive integer n. Define $\Lambda : NC(n) \to Luk(n)$ as follow: given $\pi \in NC(n)$ where $\pi = \{V_1, V_2, \cdots, V_k\}$. Let the left most element of V_i be a_i (i.e. $a_i = \min V_i$). Then let*

$$\lambda_m = \begin{cases} |V_i| - 1 & \text{if } m = a_i \\ -1 & \text{otherwise} \end{cases}$$

There is a unique Lukasiewicz path, defined as $\Lambda(\pi)$, that corresponds to the rise-vector $(\lambda_1, \lambda_2, \cdots, \lambda_n)$. Moreover, Λ is a bijection.

Remark 2.2.5. It is clear from our construction of Λ that a non-crossing partition $\pi \in NC(n)$ has exactly k blocks if and only if $\Lambda(\pi)$ has exactly $n-k$ falling steps: exactly $n-k$ points out of $\{1, 2, \cdots, n\}$ are not the left-most point of some block, which give rise to $n-k$ falling steps. The rest k points corresponds to steps of $(1, |V_i| - 1)$ where $|V_i| - 1 \geq 0$.

Remark 2.2.6. The enumeration on Lukasiewicz paths is easier due to the *cyclic permutation* trick: Consider a Lukasiewicz path γ with rise-vector $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_n)$. By appending an -1 be the end of the rise vector λ, we get a rise vector for an almost-rising path ended at $(n+1, -1)$.

Recall that cyclic permutation on a vector of n variables is essentially applying the cycle permutation $(1, 2, \cdots, n)$ repeatedly: for example, applying cyclic permutations the vector $(a_1, a_2, \cdots, a_n)$ once gives $(a_2, a_3, \cdots, a_n, a_1)$; twice gives $(a_3, a_3, \cdots, a_n, a_1, a_2)$, etc..

Now observe that if we apply cyclic permutations on the new rise-vector $(\lambda_1, \cdots, \lambda_n, -1)$, we always get another rise-vector of some almost-rising path from $(0, 0)$ to $(n+1, -1)$: indeed, no matter how we permute this vector, $-1 + \sum_{k=1}^{n} \lambda_k = -1$. Hence we may define a map:

$$\Gamma : Luk(n) \times \{0, 1, \cdots, n\} \to \{\tau : \tau \text{ is an almost-rising path ended at (n+1,-1)}\}$$

where given $\gamma \in Luk(n)$ with rise-vector λ and $0 \leq i \leq n$, $\Gamma(\gamma, i)$ be the unique almost-rising path from $(0, 0)$ to $(n+1, -1)$ whose rise-vector is the same as what we get by apply cyclic permutation on $(\lambda_1, \cdots, \lambda_n, -1)$ i times. In other words, for $0 \le i \le n-1$, $\Gamma(\gamma, i)$ has rise vector $(\lambda_{i+1}, \cdots, \lambda_n, -1, \lambda_1, \cdots, \lambda_i)$; and when $i = n$, $\Gamma(\gamma, i)$ has rise vector $(-1, \lambda_1, \cdots, \lambda_n)$.

Proposition 2.2.7. *The map Γ defined in Remark 2.2.6 is a bijection.*

For the detailed proof of 2.2.7, please refer to the Proposition 9.11 in the book [12].

Remark 2.2.8. For a Lukasiewicz path γ and any $i = 0, 1, \cdots, n$, $\Gamma(\gamma, i)$ contains exactly one more falling step than γ. Hence for each $\pi \in NC(n)$ with k blocks, $\Gamma(\Lambda(\pi), i)$ contains $n+1-k$ falling steps; Moreover, since Γ, Λ are bijections, each almost-rising path with $n+1-k$ falling steps is the image of $\Gamma(\Lambda(\pi), i)$ for some $\pi \in NC(n)$ with k blocks and $0 \le i \le n$.

Now we can prove the following result on the enumeration of non-crossing partitions with certain size.

Proposition 2.2.9. *Fix n and $1 \le k \le n$. Then the number of $\pi \in NC(n)$ with $X_n(\pi) = k$ is given by the Narayana number $N(n,k) = \frac{1}{n}\binom{n}{k}\binom{n}{k-1}$. In particular, $P(X_n = k) = \frac{N(n,k)}{C_n}$.*

Proof. Let $\Lambda : NC(n) \to Luk(n)$ be the bijection defined in Proposition 2.2.4, and Γ be the bijection defined in Proposition 2.2.7. It follows from Remark 2.2.8 that the number of $\pi \in NC(n)$ with k blocks is exactly

$$\frac{1}{n+1}|\{\gamma : \gamma \text{ almost rising path ended at (n+1,-1), with n+1-k falling steps}\}|$$

Now consider γ almost rising path ended at $(n+1, -1)$ with $n+1-k$ falling steps: let its rise-vector be $(\lambda_1, \lambda_2, \cdots, \lambda_{n+1})$. Among these $n+1$ numbers, only k of them are non-negative. Let them be $l_1, l_2, \cdots, l_k$ (in the same order as they appear in λ_i. Then l_i satisfies $\sum_{i=1}^{k} l_i - (n+1-k) = -1$, i.e. $\sum_{i=1}^{k} l_i = n-k$.

Since $l_i \ge 0$, the equation $\sum_{i=1}^{k} l_i = n-k$ contains exactly $\binom{n-1}{k-1}$ solutions. To choose these k numbers out of $n+1$ λ_i's, there are $\binom{n+1}{k}$ choices. Hence, in total we get:

$$
\begin{aligned}
& |\{\pi \in NC(n) : \pi \text{ has k blocks}\}| \\
= \;& \frac{1}{n+1} |\{\gamma : \gamma \text{ almost rising path ended at (n+1,-1), with n+1-k falling steps}\}| \\
= \;& \frac{1}{n+1}\binom{n+1}{k}\binom{n-1}{k-1} \\
= \;& \frac{1}{n+1}\frac{(n+1)!}{k!(n+1-k)!}\frac{(n-1)!}{(k-1)!(n-k)!} \\
= \;& \frac{1}{n}\binom{n}{k}\binom{n}{k-1}
\end{aligned}
$$

□

2.3 Embed Non-crossing partitions in the Symmetric Group

One might notice the cycle structure of permutations also gives a partition of the set $\{1, 2, \cdots, n\}$, though most of the time, this partition has crossings. One can regard a non-crossing partition as a permutation in the following manner: for a non-crossing partition $\pi = \{V_1, V_2, \cdots, V_k\} \in NC(n)$, it corresponds to an element in $P_\pi \in \mathcal{S}_n$ by arranging each V_i in the increasing order and treating them as a cycle permutation. In other words, for each $V_i = \{v_{i_1} < v_{i_2} < \cdots < v_{i_k}\}$, the permutation $P_\pi(v_{i_j}) = v_{i_{j+1}}$ for $j = 1, 2, \cdots, k-1$ and $P_\pi(v_{i_k}) = v_{i_1}$. For example, for a non-crossing partition $\pi = \{\{1, 5, 7\}, \{2, 3, 4\}, \{6\}\} \in NC(7)$, the corresponding P_π, when written in cyclic form, is equal to $(1, 5, 7)(2, 3, 4)(6)$. This map $\pi \mapsto P_\pi$ seems like an ad-hoc definition. However, it has been shown, starting with [1], that this map is quite meaningful. We shall start with Cayley graph structure of the symmetric group.

Definition 2.3.1. (Cayley graph) Let G be a group with a set of generators T. Assume T satisfies the following assumptions:

1. $e \notin T$ (e is the identity in G)
2. If $t \in T$, t^{-1} is also in T

3. If $t \in T, g \in G$, then $gtg^{-1} \in T$.

Then the Cayley graph $\Gamma = \Gamma(G, T)$ is defined as follows:

1. The vertex set $V = G$
2. The edge set $E = \{(x, y) : x^{-1}y \in T\}$. Here (x, y) is an unordered pair.

Remark 2.3.2. In the group framework, $x^{-1}y$ is a substitute for the difference between y and x. At first glance, it might look arbitrary that $x^{-1}y$ is preferred over $xy^{-1}, yx^{-1}, y^{-1}x$. It turns out the assumptions on T rule out this ambiguity on E. Indeed, if $x^{-1}y \in T$, the by T closed under inversion, $y^{-1}x \in T$. Moreover, if we conjugate $x^{-1}y$ by x, we get $yx^{-1} \in T$ so that $xy^{-1} \in T$.

Since we required T to be a set of generators, the Cayley graph Γ must be a connected graph. We may define the distance between two group elements by the distance to the Cayley graph.

Definition 2.3.3. For $x \neq e \in G$, define $\ell(x) = \min_n\{\exists t_1, \cdots, t_n \in T, x = t_1 t_2 \cdots t_n\}$. In addition, define $\ell(e) = 0$.

Recall that in order for an edge (x, y) to appear in the Cayley graph, there have to exist $t \in T$ such that $y = xt$. Because of this, the above definition of $\ell(x)$ is giving precisely the length of the shortest path from x to e. This allows us to define the distance on G:

Definition 2.3.4. Let $x, y \in G$, define their distance to be $d(x, y) = \ell(x^{-1}y)$.

One can check that d defines a metric on the group G. Given this distance, we can define a partial order on G:

Definition 2.3.5. For any $x, y \in G$, define $x \leq y$ whenever $\ell(y) = \ell(x) + d(x, y)$. In other words, $x \leq y$ whenever x is on a shortest path from e to y.

Definition 2.3.6. For any $x, y \in G$ with $x \leq y$, define the interval $[x, y] = \{z \in G : x \leq z \leq y\}$.

Example 2.3.7. Let $G = \mathcal{S}_n$ be the symmetric group. Consider $T = \{(i, j) \in G\}$ be the set of transpositions. It is clear that T is a set of generators for G. Let us verify T satisfies the requirement for Cayley graph:

Indeed, every transposition is the inverse of itself. If we have $\pi \in G, (i, j) \in T$, one can verify that $\pi \cdot (i, j) \cdot \pi^{-1}$ is the transposition $(\pi(i), \pi(j))$.

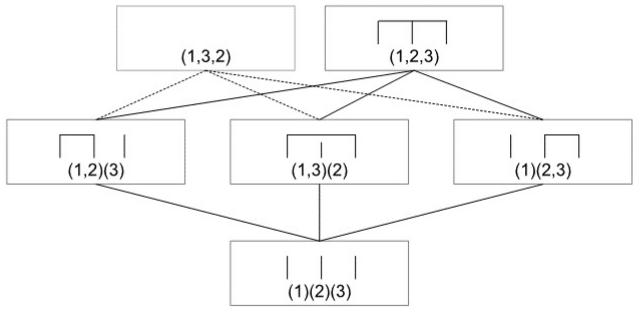

Figure 2.3.1: Embedding of $NC(3)$ into $\mathcal{S}_3$

The symmetric group $\mathcal{S}_n$ is a finite group so that as a partially ordered set, it has maximal elements. One can verify the maximal elements in $\mathcal{S}_n$ are exactly those long cycles, i.e., a cycle with n elements. Meanwhile, it has a unique minimal element, namely, e the identity. Let $\gamma_n = (1, 2, \cdots, n) \in \mathcal{S}_n$ be a long cycle. Biane [1], in 1995, showed this map $\pi \mapsto P_\pi$ (discussed in the previous section) identifies $NC(n)$ with the interval $[e, \gamma_n]$.

Theorem 2.3.8. *The map $P : NC(n) \to [\epsilon, \gamma_m]$ is an order preserving isomorphism.*

Figure 2.3.1 illustrate the case where $NC(3)$ is embedded as a maximal interval in $\mathcal{S}_3$.

As an immediate observation, since all the maximal elements in $\mathcal{S}_n$ is a long cycle of length n, every interval between the minimal element ϵ and a maximal element γ' is isomorphic to $NC(n)$. In terms of the Cayley graph, theorem 2.3.8 implies that the Hasse diagram of $NC(n)$ is a sub-graph in the Cayley graph of $(\mathcal{S}_n, T)$. This tells us $NC(n)$ is a quite important chunk of $\mathcal{S}_n$, which motivates our study of their block structures.

2.4 Some Approximation Lemmas

Recall from Proposition 2.2.9 that the total number of non-crossing partitions with a fixed number of blocks is enumerated using the Narayana number. The goal of this section is to provide some estimates on the Narayana number, which can be used as a tool to approximate the asymptotic distribution of the total number of blocks. Since the binomial coefficients $\binom{n}{k}$ appear in the Narayana number, let us provide some approximations on them.

We start off with the following well-known result due to Chernoff:

Lemma 2.4.1. *(Chernoff Bound). Let Z follow a binomial distribution with n trials and $p = 1/2$. Then for any $\alpha > 0$, we have:*

$$\begin{aligned} P(Z \geq \frac{n}{2} + \alpha) &\leq e^{-\frac{\alpha^2}{n}} \\ P(Z \leq \frac{n}{2} - \alpha) &\leq e^{-\frac{\alpha^2}{n}} \end{aligned}$$

As an immediate corollary:

Corollary 2.4.2. *Fix $\epsilon > 0$. For each non-negative integer $0 \leq k \leq n$, if $|k - \frac{n}{2}| \geq n^{\frac{1+\epsilon}{2}}$, then*

$$\frac{1}{2^n}\binom{n}{k} \leq e^{-n^{\epsilon}}$$

Proof. Suppose Z follows binomial distribution with n and $p = 1/2$. Then apply Chernoff's bound:

$$\begin{aligned} \frac{1}{2^n}\binom{n}{k} &= P(Z = k) \\ &\leq e^{-\frac{\alpha^2}{n}} \\ &= e^{-n^{\epsilon}} \end{aligned}$$

□

The good thing about the estimation $e^{-n^{\epsilon}}$ is that it is $o(n^p)$ for all $p > 0$. So when it is multiplied by any polynomial in n, the limit will goes to 0. As a result, we may just focus on the approximation when $|k - \frac{n}{2}|$ is no bigger than $\alpha = n^{\frac{1+\epsilon}{2}}$. The following version of Stirling's approximation can be found in the book [6].

Lemma 2.4.3. *(Stirling's approximation) For all positive integer n,*

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

where $\frac{1}{12n+1} < \lambda_n < \frac{1}{12n}$.

As an immediate application, one can give the following approximation for the Catalan numbers:

Corollary 2.4.4. $\lim_{n\to\infty} \frac{C_n \cdot n^{3/2}}{\sqrt{\pi}4^n} = 1$

Before we start the key approximations, it is useful to define the 'Big-O' notation.

Definition 2.4.5. Let $f, g : \mathbb{N} \to \mathbb{R}$ be two functions. We call $f = O(g)$ whenever $\limsup_{n\to\infty} \left|\frac{f(n)}{g(n)}\right| = c < \infty$. Equivalently, $f = O(g)$ whenever there exists a constant K such that for all $n \in \mathbb{N}$, $|f(n)| < K|g(n)|$.

Now let us approximate $\frac{1}{2^n}\binom{n}{\frac{n}{2}+j}$ when j is small:

Lemma 2.4.6. *Fix n and $\epsilon > 0$, for each integer j with $|j| < n^{\frac{1+\epsilon}{2}}$, let $r_{n,j}$ satisfies*

$$\frac{1}{2^n}\binom{n}{\frac{n}{2}+j} = \sqrt{\frac{2}{\pi n}} e^{-\frac{2j^2}{n}}(1 + r_{n,j})$$

Then there exists $K > 0$ constant such that for all such j and n, $|r_{n,j}| \leq K \cdot n^{3\epsilon/2-1/2}$.

Proof. Apply Stirling's formula (Lemma 2.4.3). We have:

$$
\begin{aligned}
\frac{1}{2^n}\binom{n}{\frac{n}{2}+j} &= \frac{1}{2^j}\frac{n!}{(n/2+j)!(n/2-j)!} \\
&= \frac{1}{2^n}\frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\alpha_n-\beta_n-\gamma_n}}{\sqrt{2\pi(n/2+j)}\left(\frac{(n/2+j)}{e}\right)^{n/2+j}\sqrt{2\pi(n/2-j)}\left(\frac{(n/2-j)}{e}\right)^{n/2-j}} \\
&= \frac{1}{2^n}\sqrt{\frac{n}{2\pi(n/2-j)(n/2+j)}}\frac{n^n}{(n/2+j)^{n/2+j}(n/2-j)^{n/2-j}}e^{\alpha_n-\beta_n-\gamma_n} \\
&= \frac{1}{2^n}\sqrt{\frac{2}{\pi n}}\sqrt{\frac{n^2}{n^2-4j^2}}\frac{n^n}{(n+2j)^{n/2+j}(n-2j)^{n/2-j}2^{-(n/2+j)-(n/2-j)}}e^{\alpha_n-\beta_n-\gamma_n} \\
&= \sqrt{\frac{2}{\pi n}}(1+\frac{2j}{n})^{-(n/2+j)}(1-\frac{2j}{n})^{-(n/2-j)}e^{\alpha_n-\beta_n-\gamma_n} \\
&= \sqrt{\frac{2}{\pi n}}e^{-(n/2+j)\log(1+\frac{2j}{n})-(n/2-j)\log(1-\frac{2j}{n})}\sqrt{\frac{n^2}{n^2-4j^2}}e^{\alpha_n-\beta_n-\gamma_n} \\
&= \sqrt{\frac{2}{\pi n}}e^{-2j^2/n}e^{2j^2/n-(n/2+j)\log(1+\frac{2j}{n})-(n/2-j)\log(1-\frac{2j}{n})}\sqrt{\frac{n^2}{n^2-4j^2}}e^{\alpha_n-\beta_n-\gamma_n}
\end{aligned}
$$

Hence the error term becomes

$$
1+r_{n,j} = e^{-2j/n}e^{2j/n-(n/2+j)\log(1+\frac{2j}{n})-(n/2-j)\log(1-\frac{2j}{n})}\sqrt{\frac{n^2}{n^2-4j^2}}e^{\alpha_n-\beta_n-\gamma_n}
$$

Let us break it into three pieces:

First of all, $e^{\alpha_n-\beta_n-\gamma_n}$, $\alpha_n, \beta_n, \gamma_n$ comes from the Stirling approximation. We have from Lemma 2.4.3 that

$$
\begin{aligned}
\frac{1}{12n+1} <& \quad \alpha_n \quad < \frac{1}{12n} \\
\frac{1}{12(n/2-j)+1} <& \quad \beta_n \quad < \frac{1}{12(n/2-j)} \\
\frac{1}{12(n/2+j)+1} <& \quad \gamma_n \quad < \frac{1}{12(n/2+j)}
\end{aligned}
$$

One can check $\limsup_{n\to\infty} \left|n(e^{\alpha_n-\beta_n-\gamma_n}-1)\right| = c < \infty$, and thus one may find a constant K_1 such that for all n, j, $|e^{\alpha_n-\beta_n-\gamma_n}-1| < K_1 \cdot \frac{1}{n}$.

Denote $\alpha = n^{\frac{1+\epsilon}{2}}$. The second part $\sqrt{\frac{n^2}{n^2-4j^2}}$ can be bounded as follow:

$$\begin{aligned} 1 &\leq \sqrt{\frac{n^2}{n^2-4j^2}} \\ &\leq \sqrt{\frac{n^2}{n^2-4\alpha^2}} \\ &= \left(1+\frac{4\alpha^2}{n^2-4\alpha^2}\right)^{1/2} \end{aligned}$$

Now apply Taylor expansion, we have:

$$\left(1+\frac{4\alpha^2}{n^2-4\alpha^2}\right)^{1/2} = 1+\frac{2\alpha^2}{n^2-4\alpha^2}+O(\frac{\alpha^4}{n^4})$$

The error term $\frac{2\alpha^2}{n^2-4\alpha^2}+O(\frac{\alpha^4}{n^4})$ satisfies

$\lim_{n\to\infty} \frac{\frac{2\alpha^2}{n^2-4\alpha^2}+O(\frac{\alpha^4}{n^4})}{n^{\epsilon-1}} = 2$ and thus we can pick a constant $K_2 > 2$ such that for all n, $|\frac{2\alpha^2}{n^2-4\alpha^2}+O(\frac{\alpha^4}{n^4}| < K_2 \cdot n^{\epsilon-1}$

The last piece comes from $e^{2j^2/n-(n/2+j)\log(1+\frac{2j}{n})-(n/2-j)\log(1-\frac{2j}{n})}$. Let us consider its exponent:

$$\begin{aligned} & 2j^2/n-(n/2+j)\log(1+\frac{2j}{n})-(n/2-j)\log(1-\frac{2j}{n}) \\ = \; & 2j^2/n-(n/2+j)(\frac{2j}{n}-\frac{2j^2}{n^2}+O(\frac{j^3}{n^3}))-(n/2-j)(-\frac{2j}{n}-\frac{2j^2}{n^2}+O(\frac{j^3}{n^3})) \\ = \; & O(\frac{j^3}{n^2}) \end{aligned}$$

Now take the exponential,

$$e^{2j^2/n-(n/2+j)\log(1+\frac{2j}{n})-(n/2-j)\log(1-\frac{2j}{n})} = 1 + O(\frac{j^3}{n^2})$$

The error term $O(\frac{j^3}{n^2})$ can be bounded above by

$$\begin{aligned} O(\frac{j^3}{n^2}) &= O(\frac{\alpha^3}{n^2}) \\ &= O(n^{\frac{3\epsilon}{2}-1/2}) \end{aligned}$$

We can thus find constant K_3 such that $e^{2j^2/n-(n/2+j)\log(1+\frac{2j}{n})-(n/2-j)\log(1-\frac{2j}{n})} < 1 + K_3 n^{\frac{3\epsilon}{2}-1/2}$ for all n, j.

When we multiply these three error terms together, the overall error term $r_{n,j}$ is bounded by $K_1 \cdot \frac{1}{n} + K_2 \cdot n^{\epsilon-1} + K_3 n^{\frac{3\epsilon}{2}-1/2} + O(n^{\frac{5\epsilon}{2}-3/2})$, for all valid choices of n, j. One can check this bound can be reduced to $K_3' n^{\frac{3\epsilon}{2}-1/2}$ when $\epsilon < 1/3$.

□

2.5 Asymptotic Distribution and the Method of Moments

Given a sequence of random variables, we often want to study their behavior asymptotically. In other words, we want to know how a sequence of random variables, which in some sense, converges to a particular probability measure. One natural way to define this convergence is the weak convergence.

Definition 2.5.1. Let $(\Omega, \mathcal{A})$ be a measurable space where Ω is a metric space, and ν_n, ν are probability measure on it. We say ν_n converges to ν weakly if for all bounded continuous function $f : \Omega \to \mathbb{R}$,

$$\lim_{n\to\infty} \int_\Omega f d\nu_n = \int_\Omega f d\nu$$

Fix a measure space $(\Omega, \mathcal{A}, P)$. Recall that each real-valued Borel random variable $X : \Omega \to \mathbb{R}$ defines a associated Borel measure P_X on $\mathbb{R}$ via $P_X(A) = P(X^{-1}(A))$. This allows us to define the weak convergence of random variables.

Definition 2.5.2. Let X_n be a sequence of real-valued Borel random variables on measure spaces $(\Omega_n, \mathcal{A}_n, P_n)$ and X be a real-values Borel random variable on a measure space $(\Omega, \mathcal{A}, P)$. Let the associated probability measures of X_n, and X be ν_n, and ν respectively. Then we say X_n converges to X weakly whenever ν_n converges to ν weakly.

Equivalently, since $\int_{\mathbb{R}} f(t) d\nu_n(t) = \int_{\Omega_n} f(X_n) dP_n$, we have X_n converges to X weakly whenever for all bounded continuous function $f \in C_b(\mathbb{R}, \mathbb{R})$,

$$\lim_{n\to\infty} \int_{\Omega_n} f(X_n) dP_n = \int_{\Omega} f(X) dP$$

When $(\Omega, \mathcal{A}) = (\mathbb{R}, \mathcal{B})$, the real line equipped with the Borel σ-algebra, one can define the distribution function (also called the cumulative distribution function) $F(t)$ for any Borel measure ν by $F(t) = \nu((-\infty, t]))$. The distribution function can be easily seen to be right continuous and left limit exists everywhere. Moreover, whenever $\nu(\{t\}) = 0$, F is continuous at t.

In such cases, one can define the convergence in terms of the distribution function:

Definition 2.5.3. Let ν_n, and ν be Borel measures on $\mathbb{R}$, with distribution functions F_n, and F respectively. We call ν_n converges to ν in distribution, if for any t at which F is continuous,

$$\lim_{n\to\infty} F_n(t) = F(t)$$

It turns out that these two definition are the same for probability measures on $\mathbb{R}$.

Theorem 2.5.4. *(Helly Bray theorem) Let ν_n, ν be Borel measures on $\mathbb{R}$. Then ν_n converges to ν weakly if and only if ν_n converges to ν in distribution.*

This theorem is a special case of a more general type of theorems called the Portmanteau theorem, which involves the equivalent conditions of weak convergence. Here we provide one version of it:

Theorem 2.5.5. *(Portmanteau Theorem) Let ν_n, ν be probability measures on a measurable metric space $(\Omega, \mathcal{A})$. Then the following are equivalent:*

1. *ν_v converges to ν weakly*

2. *For all open set* $U \subset \Omega$, $\liminf_{n\to\infty} \nu_n(U) \geq \nu(U)$

3. *For all closed set* $F \subset \Omega$, $\limsup_{n\to\infty} \nu_n(F) \leq \nu(F)$

4. *For all* $A \subset \Omega$ *with* $P(\partial A) = 0$, $\lim_{n\to\infty} \nu_n(A) = \nu(A)$. *Here* ∂A *is the boundary of* A.

For detailed proof and more background on these results, please refer to R.M.Dudley's book [4].

The definition of weak convergence is not easy to apply directly, since one has to check all possible bounded continuous function f if $\int_\Omega f d\nu_n$ converges to $\int_\Omega f d\nu$. A natural question to ask is whether it is possible to check a smaller collection of functions and still get the weak convergence. There are many approaches to this question, and we will focus on one of them: the method of moments. Recall, for a Borel probability measure μ on $\mathbb{R}$, the k-th moment of μ is defined as $\int_\mathbb{R} x^k d\mu$. As an illustrating example:

Proposition 2.5.6. *Let* ν_n, ν *be Borel measures on* $\mathbb{R}$. *Assume that their supports are all included in a compact interval* $[-N, N]$. *Then* $\nu_n \to \nu$ *weakly if and only if for every* $k \geq 0$, $\lim_{n\to\infty} \int_\mathbb{R} x^k d\nu_n = \int_\mathbb{R} x^k d\nu$

Proof. Since ν_n, ν are all supported within the compact interval $[-N, N]$, $\nu_n \to \nu$ weakly if and only if for every continuous function $f : [-N, N] \to \mathbb{R}$, $\lim_{n\to\infty} \int_\mathbb{R} f d\nu_n = \int_\mathbb{R} f d\nu$. Now the space of continuous functions $C([-N, N], \mathbb{R})$ can be approximated uniformly by polynomials due to the Stone-Weierstrass theorem. This uniform convergence is strong enough to guarantee that $\lim_{n\to\infty} \int_\mathbb{R} f d\nu_n = \int_\mathbb{R} f d\nu$ if and only if $\lim_{n\to\infty} \int_\mathbb{R} x^k d\nu_n = \int_\mathbb{R} x^k d\nu$ for all integer $k \geq 0$.

□

In general, we do not have the result above if the condition of compactly supported is dropped. The case when ν_n, and ν are not compactly supported is more complicated and requires much more care. We need some extra conditions on ν_n, and ν to make this theorem holds.

Definition 2.5.7. Let ν be a Borel measure on $\mathbb{R}$. We call ν is *determined by its moments* if for any Borel measure μ on $\mathbb{R}$ with $\int_\mathbb{R} x^k d\nu = \int_\mathbb{R} x^k d\mu < \infty$ for all integer $k \geq 0$, it must be the case that $\nu = \mu$.

Theorem 2.5.8. *Let* ν_n, *and* ν *be Borel measures on* $\mathbb{R}$ *with* $\lim_{n\to\infty} \int_\mathbb{R} x^k d\nu_n = \int_\mathbb{R} x^k d\nu$ *for all integer* $k \geq 0$. *Assume furthermore that* ν *is determined by its moments. Then* $\nu_n \to \nu$ *weakly.*

Theorem 2.5.8 gives us a condition that only depends on the limiting distribution ν, instead of all ν_n. Now what kind of measures is determined by its moments?

Definition 2.5.9. Let ν be a Borel measure on $\mathbb{R}$.

1. The *characteristic function* of ν is defined as $\phi_\nu(t) = \int_{\mathbb{R}} e^{itx} d\nu(x)$, when the integral exists.

2. The *moment generating function* of ν is defined as $M_\nu(t) = \int_{\mathbb{R}} e^{tx} d\nu(x)$, when the integral exists.

Rosenthal's book [14] provide a great introduction on this material. The following two results comes from section 11.1 to 11.4 from his book. We will omit their detailed proofs.

Theorem 2.5.10. *Let ν, and μ be Borel measures on $\mathbb{R}$. Then $\nu = \mu$ if and only if $\phi_\nu = \phi_\mu$.*

Proposition 2.5.11. *Let ν be a Borel measure on $\mathbb{R}$. Assume there exists $\epsilon > 0$ such that $M_\nu(t)$ is finite for $|t| < \epsilon$, then ν is determined by its moments.*

Example 2.5.12. Let ν be the measure of standard normal distribution (i.e. $d\nu(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$). Then the moment generating function for ν is $M_\nu(t) = e^{\frac{t^2}{2}}$ for all $t \in \mathbb{R}$. In particular, it follows from Theorem 2.5.11 that the Gaussian distribution is determined by its moments.

Notation 2.5.13. Denote the k-th falling factorial of n by $[n]_k = n(n-1)\cdots(n-k+1)$.

Corollary 2.5.14. *Let ν_n, and ν be Borel measures on $\mathbb{R}$ with $\lim_{n\to\infty} \int_{\mathbb{R}} [x]_k d\nu_n = \int_{\mathbb{R}} [x]_k d\nu$ for all integer $k \geq 0$. Assume furthermore that ν is determined by its moments. Then $\nu_n \to \nu$ weakly.*

Chapter 3

The Number of Blocks of a Random Non-crossing Partition

3.1 Introduction

Let us denote $X_n : NC(n) \to \mathbb{R}$ be the random variable that counts the total number of blocks of a random non-crossing partition. Our goal is to understand the asymptotic distribution of X_n.

First of all, let us recall from Remark 2.1.12 that the Kreweras complement K, a order reversing automorphism defined on $NC(n)$, give raise to the relation $X_n(\pi) + X_n(K(\pi)) = n+1$. Since K is an automorphism on $NC(n)$ where $NC(n)$ is assigned with the uniform measure, $E(X_n) = E(K(X_n))$. Immediately, one can deduce that $E(X_n) = \frac{n+1}{2}$. However, the calculation of higher moments are not easy in general. Recall from Proposition 2.2.9 that the total number of non-crossing partitions in $NC(n)$ with exactly k blocks is enumerated by the Narayana numbers $N(n,k) = \frac{1}{n}\binom{n}{k}\binom{n}{k-1}$. One can explicitly write down a formula for the k-th moments of X_n in terms of Narayana number:

$$E(X_n^k) = \sum_{i=1}^{n} \frac{N(n,i)}{C_n} i^k$$

Even though we have this explicit expression for the k-th moment, its direct computation is not easy. We shall discuss its computation in Section 3.3, where we are able to give an explicit formula for the factorial moments $E([X_n - 1]_k) = E((X_n-1)(X_n-2)\cdots(X_n-k))$.

Based on these results, we can compute the following table that shows the results for the first few moments:

Moments	Explicit Formula
$E(X_n)$	$\frac{n+1}{2}$
$E(X_n^2)$	$\frac{(n+1)(n^2+n-1)}{2(2n-1)}$
$E(X_n^3)$	$\frac{(n+1)^2(n^2+2n-2)}{4(2n-1)}$
$E(X_n^4)$	$\frac{(n+1)(n^5+4n^4-3n^3-12n^2+2n+6)}{4(2n-1)(2n-3)}$
$E(X_n^5)$	$\frac{(n+1)^2(n^5+6n^4-n^3-24n^2+4n+12)}{8(2n-1)(2n-3)}$

Table 3.1: Moments of X_n

Since $E(X_n) = \frac{n+1}{2}$, one might want to normalize X_n by subtracting its mean. It is not hard to compute this centralized moments based on table 3.1.

Moments	Explicit Formula	Asymptotic Formula
$E(X_n - \frac{n+1}{2})$	0	0
$E(\left(X_n - \frac{n+1}{2}\right)^2)$	$\frac{(n+1)(n-1)}{4(2n-1)}$	$\frac{n}{8}$
$E(\left(X_n - \frac{n+1}{2}\right)^3)$	0	0
$E(\left(X_n - \frac{n+1}{2}\right)^4)$	$\frac{(n+1)(n-1)(3n^2-4n-3)}{16(2n-1)(2n-3)}$	$\frac{3n^2}{64}$
$E(\left(X_n - \frac{n+1}{2}\right)^5)$	0	0

Table 3.2: Centralized Moments of X_n

From table 3.2, one may notice the centralized moments matches the Gaussian distribution asymptotically. However, since we only have a nice formula for the falling factorials of $X_n - 1$, the conversion to centralized moments is not easy and there are some mysterious reductions that we do not understand. Fortunately, Section 2.4 provides powerful tools for us to approximate the centralized moments which leads to the main results in Section 3.2

3.2 Main Result

We obtained the following result on the asymptotic distribution for the total number of blocks in random non-crossing partitions.

Theorem 1.2.1. *Let $X_n : NC(n) \to \mathbb{N}$ count the total number of blocks in a non-crossing partition. Then*

$$\frac{X_n - n/2}{\sqrt{n/8}} \xrightarrow{d} \mathcal{N}(0,1)$$

First, recall from Proposition 2.2.9, one can explicitly write down the p-th moment of this normalized X_n:

$$E\left(\left(\sqrt{\frac{8}{n}}(X_n - n/2)\right)^p\right) = \frac{1}{C_n}\sum_{k=1}^{n} \frac{1}{n}\binom{n}{k}\binom{n}{k-1}\left(\sqrt{\frac{8}{n}}(k - \frac{n}{2})\right)^p \tag{3.2.1}$$

We first reduce the summation to those k close enough to $n/2$:

Lemma 3.2.1. *Fix $\epsilon > 0$. We have:*

$$\lim_{n\to\infty} \frac{1}{nC_n} \sum_{|k-\frac{n}{2}|>n^{\frac{1+\epsilon}{2}}+1} \binom{n}{k}\binom{n}{k-1}\left(\sqrt{\frac{8}{n}}(k - \frac{n}{2})\right)^p = 0$$

Proof. Let $S_n = \frac{1}{nC_n}\sum_{|k-\frac{n}{2}|>n^{\frac{1+\epsilon}{2}}+1} \binom{n}{k}\binom{n}{k-1}\left(\sqrt{\frac{8}{n}}(k - \frac{n}{2})\right)^p$.

$$\begin{aligned} S_n &= \frac{1}{nC_n} \sum_{|k-\frac{n}{2}|>n^{\frac{1+\epsilon}{2}}+1} \binom{n}{k}\binom{n}{k-1}\left(\sqrt{\frac{8}{n}}(k - \frac{n}{2})\right)^p \\ &= \frac{4^n}{nC_n} \sum_{|k-\frac{n}{2}|>n^{\frac{1+\epsilon}{2}}+1} \frac{1}{2^n}\binom{n}{k}\frac{1}{2^n}\binom{n}{k-1}\left(\sqrt{\frac{8}{n}}(k - \frac{n}{2})\right)^p \end{aligned}$$

Now apply the approximation of Catalan numbers (Corollary 2.4.4),

$$\lim_{n\to\infty} \frac{\sqrt{\frac{n}{\pi}}\sum_{|k-\frac{n}{2}|>n^{\frac{1+\epsilon}{2}}+1} \frac{1}{2^n}\binom{n}{k}\frac{1}{2^n}\binom{n}{k-1}\left(\sqrt{\frac{8}{n}}(k - \frac{n}{2})\right)^p}{S_n} = 1$$

Now whenever $|k-\frac{n}{2}| > n^{\frac{1+\epsilon}{2}}+1$, both k and $k+1$ are at least $n^{\frac{1+\epsilon}{2}}$ away from $n/2$, so that Corollary 2.4.2 applies to both $\binom{n}{k}$ and $\binom{n}{k-1}$ to obtain an upper bound of $e^{-n^{\epsilon}}$. Moreover, for any $1 \le k \le n$,

$$\begin{aligned} \left(\sqrt{\frac{8}{n}}(k-\frac{n}{2})\right)^p &\le \left(\sqrt{\frac{8}{n}}(\frac{n}{2})\right)^p \\ &= (2n)^{p/2} \end{aligned}$$

Hence we have:

$$\begin{aligned} 0 &\le \sqrt{\frac{n}{\pi}} \sum_{|k-\frac{n}{2}|>n^{\frac{1+\epsilon}{2}}+1} \frac{1}{2^n}\binom{n}{k}\frac{1}{2^n}\binom{n}{k-1}\left(\sqrt{\frac{8}{n}}(k-\frac{n}{2})\right)^p \\ &\le \sqrt{\frac{n}{\pi}} \sum_{|k-\frac{n}{2}|>n^{\frac{1+\epsilon}{2}}+1} e^{-2n^{\epsilon}}(2n)^{p/2} \\ &\le \sqrt{\frac{n}{\pi}} \cdot n \cdot e^{-2n^{\epsilon}}(2n)n^{p/2} \end{aligned}$$

Take $n \to \infty$, we get:

$$\sqrt{\frac{n}{\pi}} \sum_{|k-\frac{n}{2}|>n^{\frac{1+\epsilon}{2}}+1} \frac{1}{2^n}\binom{n}{k}\frac{1}{2^n}\binom{n}{k-1}\left(\sqrt{\frac{8}{n}}(k-\frac{n}{2})\right)^p \to 0$$

and thus the same holds for S_n.

□

Now it suffices the evaluate the sum only when $|k-n/2| \le n^{\frac{1+\epsilon}{2}}$. The following lemma simplifies the summation for computational convenience:

Lemma 3.2.2. *The limit*

$$\lim_{n\to\infty}\frac{1}{nC_n}\sum_{|j|\le n^{\frac{1+\epsilon}{2}}}\binom{n}{\frac{n}{2}+j}\binom{n}{\frac{n}{2}+j-1}\left(\sqrt{\frac{8}{n}}j\right)^p$$

exists if and only if

$$\lim_{n\to\infty}\frac{1}{nC_n}\sum_{|j|\le n^{\frac{1+\epsilon}{2}}}\binom{n}{\frac{n}{2}+j}^2\left(\sqrt{\frac{8}{n}}j\right)^p$$

exists. Moreover, the limit are the same (if they exist).

Proof. Notice $\binom{n}{k}=\frac{n-k}{k}\binom{n}{k-1}$. So that for each $|j|\le n^{\frac{1+\epsilon}{2}}$,

$$\binom{n}{\frac{n}{2}+j-1}=\frac{\frac{n}{2}+j}{\frac{n}{2}-j}\binom{n}{\frac{n}{2}+j}$$

Now j is small compared to n, we may have the following estimate:

$$\begin{aligned}\frac{\frac{n}{2}+j}{\frac{n}{2}-j} &= 1+\frac{2j}{\frac{n}{2}-j}\\ &= 1+\frac{2}{\frac{n}{2j}-1}\\ &\le 1+\left|\frac{2}{\frac{n}{2n^{\frac{1+\epsilon}{2}}}-1}\right|\\ &= 1+O(n^{-(1-\epsilon)/2})\end{aligned}$$

The bound $O(n^{-(1-\epsilon)/2})$ provides an uniform bound of the error term, independent of j. Apply squeeze theorem, we get the desired result.

□

Now we are well equipped for the proof of the main theorem:

Theorem 3.2.3.

$$\lim_{n\to\infty} E\left(\left(\sqrt{\frac{8}{n}}(X_n - \frac{n}{2})\right)^p\right) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} t^p dt$$

Proof. Recall Equation 3.2.1 where we can write $E\left(\left(\sqrt{\frac{8}{n}}(X_n - \frac{n}{2})\right)^p\right)$ explicitly as:

$$E(\left(\sqrt{\frac{8}{n}}(X_n - n/2)\right)^p) = \frac{1}{C_n}\sum_{k=1}^{n}\frac{1}{n}\binom{n}{k}\binom{n}{k-1}\left(\sqrt{\frac{8}{n}}(k-\frac{n}{2})\right)^p$$

Lemma 3.2.1 reduces this summation over all k with $|k - n/2| \le n^{\frac{1+\epsilon}{2}}$. Lemma 3.2.2 further shows that this summation is the same as the summation where $\binom{n}{k-1}$ is replaced by $\binom{n}{k}$. We are now arrived at the following equation:

$$\lim_{n\to\infty} E\left(\left(\sqrt{\frac{8}{n}}(X_n - \frac{n}{2})\right)^p\right) = \lim_{n\to\infty}\frac{4^n}{nC_n}\sum_{|j|\le n^{\frac{1+\epsilon}{2}}}\frac{1}{4^n}\binom{n}{\frac{n}{2}+j}^2\left(\sqrt{\frac{8}{n}}j\right)^p \tag{3.2.2}$$

Now apply Lemma 2.4.6, for each j with $|j| \le n^{\frac{1+\epsilon}{2}}$, we have the equation:

$$\frac{1}{2^n}\binom{n}{\frac{n}{2}+j} = \sqrt{\frac{2}{\pi n}}e^{-\frac{2j^2}{n}}(1 + r_{n,j}) \tag{3.2.3}$$

where there exists a constant K such that $|r_{n,j}| \le K \cdot n^{3\epsilon/2-1/2}$ for all $|j| \le n^{\frac{1+\epsilon}{2}}$. Substitute Equation 3.2.3 into the Equation 3.2.2,

$$
\begin{aligned}
&\lim_{n\to\infty} E\left(\left(\sqrt{\frac{8}{n}}(X_n-\frac{n}{2})\right)^p\right)\\
=\ &\lim_{n\to\infty}\frac{4^n}{nC_n}\sum_{|j|\le n^{\frac{1+\epsilon}{2}}}\left(\sqrt{\frac{2}{\pi n}}e^{-\frac{2j^2}{n}}(1+r_{n,j})\right)^2\left(\sqrt{\frac{8}{n}}j\right)^p\\
=\ &\lim_{n\to\infty}\frac{4^n}{nC_n}\sum_{|j|\le n^{\frac{1+\epsilon}{2}}}\frac{2}{\pi n}e^{-\frac{4j^2}{n}}(1+r_{n,j})^2\left(\sqrt{\frac{8}{n}}j\right)^p\\
=\ &\lim_{n\to\infty}\frac{4^n}{\sqrt{\pi}n^{3/2}C_n}\sum_{|j|\le n^{\frac{1+\epsilon}{2}}}\frac{1}{\sqrt{2\pi}}\sqrt{\frac{8}{n}}e^{-\frac{4j^2}{n}}(1+r_{n,j})^2\left(\sqrt{\frac{8}{n}}j\right)^p
\end{aligned}
$$

Apply the asymptotic formula of Catalan number as suggested in Corollary 2.4.4,

$$
\begin{aligned}
&\lim_{n\to\infty} E\left(\left(\sqrt{\frac{8}{n}}(X_n-\frac{n}{2})\right)^p\right)\\
=\ &\lim_{n\to\infty}\sum_{|j|\le n^{\frac{1+\epsilon}{2}}}\frac{1}{\sqrt{2\pi}}\sqrt{\frac{8}{n}}e^{-\frac{4j^2}{n}}(1+r_{n,j})^2\left(\sqrt{\frac{8}{n}}j\right)^p\\
=\ &\lim_{n\to\infty}\sum_{|j|\le n^{\frac{1+\epsilon}{2}}}\frac{1}{\sqrt{2\pi}}\sqrt{\frac{8}{n}}e^{-\frac{\left(\sqrt{\frac{8}{n}}j\right)^2}{2}}(1+r_{n,j})^2\left(\sqrt{\frac{8}{n}}j\right)^p
\end{aligned}
$$

The factor $(1+r_{n,j})^2$ can be bounded uniformly using by $(1-K\cdot n^{3\epsilon/2-1/2})^2\le(1+r_{n,j})^2\le(1+K\cdot n^{3\epsilon/2-1/2})^2$. For ϵ small enough, both $(1-K\cdot n^{3\epsilon/2-1/2})^2$, $(1+K\cdot n^{3\epsilon/2-1/2})^2$ converges to 1 as $n\to\infty$. Hence it suffices to remove the factor of $(1+r_{n,j})^2$ and consider the following summation:

$$
\sum_{|j|\le n^{\frac{1+\epsilon}{2}}}\frac{1}{\sqrt{2\pi}}\sqrt{\frac{8}{n}}e^{-\frac{\left(\sqrt{\frac{8}{n}}j\right)^2}{2}}\left(\sqrt{\frac{8}{n}}j\right)^p
$$

Observe that this can be thought as a Riemann sum with sampling points at $\sqrt{\frac{8}{n}}j$, where $|j| \leq n^{\frac{1+\epsilon}{2}}$. $\sqrt{\frac{8}{n}}j$ will lie in an interval of width in the order of $n^{\epsilon/2}$, which goes to ∞ as $n \to \infty$. We deduce immediately that this summation converges to the following Riemann Integral:

$$\sum_{|j| \leq n^{\frac{1+\epsilon}{2}}} \frac{1}{\sqrt{2\pi}} \sqrt{\frac{8}{n}} e^{-\frac{\left(\sqrt{\frac{8}{n}}j\right)^2}{2}} \left(\sqrt{\frac{8}{n}}j\right)^p = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} t^p dt$$

This is exactly the desired result.

$\square$

Now Theorem 1.2.1 follows from 3.2.3:

Proof. Recall from 3.2.3, the p-th moment of $\sqrt{\frac{8}{n}}(X_n - n/2)$ is asymptotically equal to the integral

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} t^p dt$$

Recall the probability density function of $\mathcal{N}(0,1)$ is $\rho(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$. Hence the p-th moment of $\sqrt{\frac{8}{n}}(X_n - n/2)$ is asymptotically equal to the p-th moment of $\mathcal{N}(0,1)$.

We have shown that normal distributions are uniquely determined by its moments in Example 2.5.12, and thus Theorem 2.5.8 can be applied to obtain $\sqrt{\frac{8}{n}}(X_n - n/2)$ converges weakly to standard normal distribution.

$\square$

3.3 Moments of the Total Number of Blocks

We are able to compute the moments of X_n explicitly. Recall the notation $[n]_l$ denotes the l-th falling factorial which is $[n]_l = n(n-1)\cdots(n-l+1)$. The following theorem computes the falling factorial moments of X_n:

Proposition 3.3.1. *For every positive integer l,*

$$E([X_n - 1]_l) = \frac{[n]_l[n-1]_l}{[2n]_l}$$

Proof. First we notice from the distribution of X_n, we have:

$$\begin{aligned} E([X_n - 1]_l) &= \frac{1}{nC_n} \sum_{k=1}^{n} \binom{n}{k} \binom{n}{k-1} [k-1]_l \\ &= \frac{1}{nC_n} \sum_{k=1}^{n} \binom{n}{n-k} \binom{n}{k-1} [k-1]_l \end{aligned}$$

Let us denote

$$\begin{aligned} g(z) &= (1+z)^n = \sum_{k=0}^{n} \binom{n}{k} z^k \\ f(z) &= \sum_{k=1}^{n} \binom{n}{k-1} z^{k-1} = (1+z)^n - z^n \end{aligned}$$

Then,

$$\begin{aligned} f_l(z) &= \frac{d^l}{dz^l} f(z) \\ &= [n]_l((1+z)^{n-l} - z^{n-l}) \\ &= \sum_{k=1}^{n} \binom{n}{k-1} [k-1]_l z^{k-l-1} \end{aligned}$$

Consider $f_l(z) \cdot g(z)$: on one hand,

$$\begin{aligned}
f_l(z)\cdot g(z) &= \left(\sum_{k=1}^{n}\binom{n}{k-1}[k-1]_l z^{k-l-1}\right)\cdot\left(\sum_{k=0}^{n}\binom{n}{k}z^k\right)\\
&= \sum_{k=1}^{n}\sum_{j=0}^{n}\binom{n}{k-1}[k-1]_l z^{k-l-1}\binom{n}{j}z^j\\
&= \sum_{d=1}^{2n}\sum_{k=0}^{n}\binom{n}{k-1}[k-1]_l\binom{n}{d-k}z^{d-l-1}
\end{aligned}$$

So consider the coefficient of z^{n-l-1} in $f_l(z)\cdot g(z)$, it is exactly:

$$\sum_{k=0}^{n}\binom{n}{k-1}[k-1]_l\binom{n}{n-k}$$

On the other hand,

$$\begin{aligned}
f_l(z)\cdot g(z) &= [n]_l((1+z)^{n-l}-z^{n-l})\cdot(1+z)^n\\
&= [n]_l((1+z)^{2n-l}-z^{n-l}(1+z)^n)
\end{aligned}$$

One can check the coefficient for z^{n-l-1} is exactly $[n]_l\binom{2n-l}{n-l-1}$

Now plug everything back, we have:

$$\begin{aligned}
E([X_n-1]_l) &= \frac{1}{C_n}\frac{1}{n}\sum_{k=1}^{n}\binom{n}{k}\binom{n}{k-1}[k-1]_l\\
&= \frac{1}{nC_n}[n]_l\binom{2n-l}{n-l-1}\\
&= \frac{(n+1)!n!}{(2n)!}\frac{1}{n}[n]_l\frac{(2n-l)!}{(n-l-1)!(n+1)!}\\
&= [n]_l\frac{(n-1)!}{(n-l-1)!}\frac{(2n-l)!}{(2n)!}\\
&= \frac{[n]_l[n-1]_l}{[2n]_l}
\end{aligned}$$

$\square$

Based on Theorem 3.3.1, one can explicitly compute the first few moments of X_n as shown in table 3.1. However, it is hard to use this method to handle more generalized problems. Since we are interested in the moments of centralized X_n, let us try to compute $E([X_n - \frac{n}{2} - 1]_l)$.

Corollary 3.3.2.

$$E([X_n - \frac{n+1}{2} - 1]_l) = \frac{1}{nC_n} \sum_{j=0}^{l} \binom{l}{j} [-n/2]_j [n]_{l-j} \binom{2n-l+j}{n-l+j-1}$$

Proof. As before, we can explicitly write

$$E([X_n - \frac{n}{2} - 1]_l) = \frac{1}{nC_n} \sum_{k=1}^{n} \binom{n}{n-k} \binom{n}{k-1} [k - n/2 - 1]_l \tag{3.3.1}$$

Using the same trick, if we let

$$\begin{aligned} g(z) &= (1+z)^n = \sum_{k=0}^{n} \binom{n}{k} z^k \\ f(z) &= \sum_{k=1}^{n} \binom{n}{k-1} z^{k-1} = (1+z)^n - z^n \end{aligned}$$

However, since we have the factor $[k - n/2 - 1]_l$ instead of $[k-1]_l$, we need to find the l-th derivative

$$\begin{aligned} F_l(z) &= \frac{d^l}{dz^l} z^{-n/2} f(z) \\ &= \sum_{k=1}^{n} \binom{n}{k-1} [k - n/2 - 1]_l z^{k-l-1-n/2} \end{aligned}$$

But in this case, we need to apply the product rule to find $F_l(z)$:

$$\begin{aligned} F_l(z) &= \frac{d^l}{dz^l} z^{-n/2} f(z) \\ &= \sum_{j=0}^{l} \binom{l}{j} \frac{d^j}{dz^j} z^{-n/2} \frac{d^{l-j}}{dz^{l-j}} f(z) \\ &= \sum_{j=0}^{l} \binom{l}{j} [-n/2]_j z^{-n/2-j} f_{l-j}(z) \end{aligned}$$

Here $f_i(z)$ are the same as those in the proof of Proposition 3.3.1.

Then consider $F_l(z) \cdot g(z)$: a similar argument gives us that $\sum_{k=1}^{n} \binom{n}{n-k}\binom{n}{k-1}[k-n/2-1]_l$ is the coefficient of $z^{n/2-l-1}$ in $F_l(z)g(z)$.

Now we break $F_l(z)$ into a sum of $l+1$ pieces, consider the coefficient of $z^{n/2-l-1}$ in the j-th piece $[-n/2]_j z^{-n/2-j} f_{l-j}(z)g(z)$. One can check it is equal to $[-n/2]_l[n]_{l-j}\binom{2n-l+j}{n-l+j-1}$.

Combine everything back, we get

$$E([X_n - n/2 - 1]_l) = \frac{1}{nC_n} \sum_{j=0}^{l} \binom{l}{j} [-n/2]_j [n]_{l-j} \binom{2n-l+j}{n-l+j-1}$$

□

In some sense, Corollary 3.3.2 is a big step towards our main result in the sense that it can express the l-th factorial moments of centralized X_n in terms of a sum of $l+1$ pieces, where each piece is a rational function of n of order n^l

However, due to Theorem 1.2.1, we know $E([X_n - n/2 - 1]_l)$ should have order $n^{l/2}$ when l is even and $n^{(l-1)/2}$ when l is odd. This suggests that this summation of $l+1$ rational functions of order n^l should have extensive cancellations so that the order is reduced by about a half after the summation. It is not known how these cancellations happen.

Chapter 4

The Number of Outer Blocks of a Random Non-crossing Partition

4.1 Introduction

Recall an outer block V_i from a non-crossing partition $\pi \in NC(n)$ is a block which is not enclosed by any other blocks. We denote $Y_n : NC(n) \to \mathbb{N}$ to be the random variable that sends each $\pi \in NC(n)$ to the number of outer blocks in π. In this chapter, we study the asymptotic distribution for the random variable Y_n.

Remark 4.1.1. The number of outer blocks turns to be connected with many combinatorical objects. Let $Z_n : D_n \to \mathbb{N}$ count the number of zeros in a Dyck path of length $2n$ (Here D_n contains all the Dyck paths of length $2n$, and it is assigned with the uniform measure). Then $Z_n - 1$ and Y_n (the number of outer blocks) have the same distribution. In fact, let $\tau : D_n \to NC(n)$ be the bijection discussed discussed in Proposition 2.1.5, we have $Z_n - 1 = Y_n \circ \tau$.

Remark 4.1.2. Outer blocks also contains important information on the structures of non-crossing partition. For example, consider the non-crossing partition $\pi = \{\{1,6\},\{2,4,5\},\{3\},\{7,9\},\{8\}\} \in NC(9)$. It has two outer blocks $\{1,6\}$ and $\{7,9\}$. These two outer blocks break the points $\{1,2,\cdots,9\}$ into two chunks, namely a chunk from 2 to 5 and another chunk of single point 8. Now focus on the chunk of from 2 to 5, π restricted to this chunk is another non-crossing partition in $NC(4)$. On this section there is only one outer block $\{2,4,5\}$, which further divides this chunk into a chunk of one point 3. In general, we can obtain any non-crossing partition using the a similar procedure of placing layers of

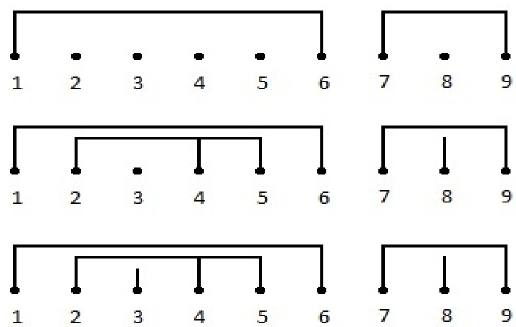

Figure 4.1.1: An Example of Outer Block Layers

outer blocks. It starts with the first layer of outer blocks, which divides $\{1, 2, \cdots, n\}$ into smaller chunks. We then repeatedly place the next layer of outer blocks consisting of those outer blocks for each of these small chunks, and each layer further divides the points into smaller chunks. Figure 4.1.1 illustrates this process for π defined as above.

In essence, outer blocks breaks the non-crossing partition into self-similar pieces of smaller non-crossing partition.

Remark 4.1.3. In the preceding chapter, we saw that the expected total number of blocks of $\pi \in NC(n)$ is $\frac{n+1}{2}$. In view of that, it is rather surprising to find out that, as it will follow from the main theorem of the present chapter, the expected number of outer blocks of $\pi \in NC(n)$ is $\frac{C_n - C_{n-1}}{C_{n-1}}$, which is asymptotically 3. Notice that the block containing 1 or n is automatically an outer block. Thus, when $1, n$ are in different blocks, we already have two outer blocks on the left-most and right-most end of the non-crossing partition. Furthermore, if we only count those outer blocks with 1 elements (which we call it outer singletons), it is shown in Proposition 5.3.1 that the expected number of outer singletons is exactly 1 for all $n \in \mathbb{N}$.

There is no immediate analog of outer blocks in terms of random permutations. However, the following observation, which we shall prove as Lemma 4.3.1, relates the number of outer blocks with the size of the block containing a certain vertex:

Observation 4.1.4. For $\pi \in NC(n)$, define $Y_n'(\pi)$ to be the size of the block that the node n belongs to. Then, for any $\pi \in NC(n)$, $Y_n(\pi) = Y'(K(\pi))$. Here $K(\pi)$ is the Krewewas complement of π. One can immediately deduce that Y_n has the same distribution as the random variable on $NC(n)$ that counts the size of the block containing 1.

Remark 4.1.5. Let the random variable θ_n be the size of the cycle containing 1. This can be seen by Observation 4.1.4 as an analog of the number of outer block in the context of random permutations. Consider the probability that $\theta_n = k$: to have 1 contained in a cycle of length k, we need to first choose and arrange the rest $k-1$ elements for the cycle. This gives us $\binom{n-1}{k-1} \cdot (k-1)!$ choices. The rest $n-k$ elements has $(n-k)!$ ways to arrange, so in total, we have $P(\theta_n = k) = \binom{n-1}{k-1} \cdot (k-1)! \cdot (n-k)! = \frac{1}{n}$. This tells us θ_n is a uniform distribution on $\{1, 2, \cdots, n\}$.

The goal of this chapter is to derive the asymptotic distribution of Y_n, the total number of outer blocks. Our main result is:

Theorem 1.2.2. *Let ν be a probability measure with mass $\frac{k}{2 \cdot 2^k}$ at point $k = 1, 2, \cdots$. Then $Y_n \xrightarrow{d} \nu$.*

The proof requires a deeper understanding of the asymptotic behavior of coefficients in function composition, which we shall develop in Section 4.2.

4.2 Coefficients in the Composition of Functions

Recall the notation $[z^n]f(z) = a_n$, where $f(z) = \sum_{n=0}^{\infty} a_n z^n$. In the study of the asymptotic distribution of Y_n, the following result is handy at multiple occasions:

Proposition 4.2.1. *Let $C(z)$ be the generation function of the Catalan numbers. Then,*

$$\lim_{n\to\infty} \frac{[z^n]C(z)^k}{[z^n]C(z)} = k \cdot 2^{k-1}$$

This section gives a proof for a more general problem: consider two analytical functions $f(z)$ and $g(z)$, what is the asymptotic behavior of the coefficients in the composition of g and f (i.e. $[z^n]g(f(z))$) in terms of f and g? One can quickly observe that Proposition 4.2.1 is an special case of this question where $f(z) = C(z)$ is the generating function of Catalan numbers, and $g(z) = z^k$. The main result we obtained in this special case is the following Proposition:

Proposition 4.2.2. *$f(z) = \sum_{n=0}^{\infty} a_n z^n$ where $\lim_{n\to\infty} \frac{a_n}{a_{n+1}} = r$ and $a_n \geq 0$. Now let g be another analytic function with non-negative coefficients (i.e. $[z^n]g(z) \geq 0$ for all $n = 0, 1, \cdots$) such that $g'(f(r))$ exists. Assume furthermore that $\frac{a_n}{a_{n+1}} \leq r$ for all n. Then,*

$$\lim_{n\to\infty} \frac{[z^n]g(f(z))}{[z^n]f(z)} = g'(f(r))$$

Remark 4.2.3. Let $f(z) = C(z)$ be the generating function of Catalan numbers and $g(z) = z^k$. Then one can check that $\lim_{n\to\infty} \frac{C_n}{C_{n+1}} = \frac{1}{4}$ and moreover $\frac{C_n}{C_{n+1}} = \frac{n+2}{4n+2} \leq \frac{1}{4}$. We also have $C(1/4) = 2$ and $g'(2) = k \cdot 2^{k-1}$ exists. Thus Proposition 4.2.2 applies, which gives the Proposition 4.2.1

Now observe that for any function $h(z)$, $[z^n]h'(z) = (n+1)[z^{n+1}]h(z)$. Hence the ratio $\frac{[z^n]g(f(z))}{[z^n]f(z)}$ remains the same when we take derivatives on both the top and the bottom. In other words,

$$\frac{[z^{n+1}]g(f(z))}{[z^{n+1}]f(z)} = \frac{[z^n]g'(f(z))f'(z)}{[z^n]f'(z)} \tag{4.2.1}$$

Set $u(z) = f'(z)$ and $v(z) = g'(f(z))$, and observe that for $u(z) = \sum_{n=0}^{\infty} u_n z^n$, the condition $\lim_{n\to\infty} \frac{u_n}{u_{n+1}} = r$ is still satisfied. This leads to the following equivalent result of Proposition 4.2.2:

Proposition 4.2.4. *Let $u(z) = \sum_{n=1}^{\infty} u_n z^n$ and $v(z) = \sum_{n=0}^{\infty} v_n z^n$ be such that $u_i, v_i \geq 0$ for all $i \geq 0$ and $\lim_{n\to\infty} \frac{u_n}{u_{n+1}} = r$. Assuming further that $v(r) = \sum_{n=0}^{\infty} v_n r^n$ converges and $\frac{u_n}{u_{n+1}} \leq r$ for all n, then*

$$\frac{[z^n]v(z)u(z)}{[z^n]u(z)} = v(r)$$

Proof. Define a positive measure on $\mathbb{N}$ where $\nu = \sum_{n=0}^{\infty} v_i \delta_i$, where δ_i is the Dirac measure at point i. Denote $f_n : \mathbb{R} \to \mathbb{R}$ be such that $f_n(i) = \frac{u_{n-i}}{u_n}$ for all $i = 0, 1, \cdots, n$ and 0 otherwise. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(i) = r^i$ for all positive integer $i \geq 0$ and 0 otherwise. Then by our condition on u_n, we know $0 \leq f_n \leq f$, and f_n converges to f pointwisely.

Now notice

$$\begin{aligned}\int f_n d\nu &= \frac{\sum_{i=0}^{n} v_i u_{n-i}}{u_n} \\ &= \frac{[z^n]v(z)u(z)}{[z^n]u(z)}\end{aligned}$$

Moreover, $\int f d\nu = \sum_{n=0}^{\infty} v_n r^n$ exists, f is an integrable function that dominates each f_n. Apply the dominated convergence theorem to obtain

$$\begin{aligned}
\lim_{n\to\infty} \frac{[z^n]v(z)u(z)}{[z^n]u(z)} &= \lim_{n\to\infty} \int f_n d\nu \\
&= \int f d\nu \\
&= v(r)
\end{aligned}$$

□

4.3 Asymptotic Distribution of Outer Blocks

Let us start this section with the following Lemma, which plays an important role in the proof:

Lemma 4.3.1. *For $\pi \in NC(n)$, define $B_n^{(i)}(\pi)$ be the size of the block containing node i where $i = 1, 2, \cdots, n$. Then,*

1. *For any $\pi \in NC(n)$, $Y_n(\pi) = B_n^{(n)}(K(\pi))$. Here $K(\pi)$ is the Krewewas complement of π.*
2. *For a fixed integer $n > 0$, $B_n^{(i)}$ has the same distribution for all $i = 1, 2, \cdots, n$. Moreover, Y_n has the same distribution as $B_n^{(1)}$.*

Now we have all the tools to prove the Theorem 1.2.2. Lemma 4.3.1 suggests that it suffices to compute the asymptotic distribution of $B_n^{(1)}$.

Let $D_{n,k}$ be the cardinality of the set $\{\pi \in NC(n) | 1 \text{ belongs to a block of size } k\}$. Define its joint generating function

$$D(t, z) = \sum_{n=0}^{\infty} \sum_{k=0}^{n} D_{n,k} t^k z^n$$

The following Lemma gives an explicit formula for $D(t, z)$:

Lemma 4.3.2. $D(t, z) = \frac{1}{1-tzC(z)}$*, where $C(z)$ is the generating function for Catalan numbers.*

Proof. Let us fix a k for which we compute $D_k(z) = \sum_{n=0}^{\infty} D_{n,k} z^n$. Consider the set $A_{n,k} = \{\pi \in NC(n) | 1 \text{ belongs to a block of size } k\}$, the block V containing node 1 divides the set $\{1, 2, \cdots, n\} \backslash V$ into k disjoint chunks. We can thus partition this set $A_{n,k}$ according to the number of points in each chunk: let there are d_1 points in the first chunk, d_2 in the second, and so on. Then $d_1, d_2, \cdots, d_k$ are non-negative integers satisfies $\sum_{i=1}^{k} d_i = n - k$.

Now let us count how many $\pi \in NC(n)$ satisfies that 1 is in a block of size k that breaks $\{1, 2, \cdots, n\}$ into chunks of given sizes $d_1, d_2, \cdots, d_k$. In each chunk, we can fill in any non-crossing partition, and two different chunks will not interfere with each other. One can check this implies there is a total of $C_{d_1} C_{d_2} \cdots C_{d_k}$ ways to fill π.

Sum over all possible $\{d_1, d_2, \cdots, d_k\}$, we obtain that

$$
\begin{aligned}
D_k(z) &= \sum_{n=0}^{\infty} D_{n,k} z^n \\
&= \sum_{n=0}^{\infty} \sum_{d_1+d_2+\cdots+d_k=n-k} C_{d_1} C_{d_2} \cdots C_{d_k} z^n \\
&= \sum_{n=0}^{\infty} [z^{n-k}] C(z)^k z^n \\
&= (zC(z))^k
\end{aligned}
$$

Hence,

$$
\begin{aligned}
D(t, z) &= \sum_{k=0}^{\infty} D_k(z) t^k \\
&= \sum_{k=0}^{\infty} (tzC(z))^k \\
&= \frac{1}{1 - tzC(z)}
\end{aligned}
$$

□

We are able to compute the asymptotic factorial moments based given this explicit formula of $D(t, z)$:

Theorem 4.3.3. $\lim_{n\to\infty} E([B_n^{(1)}]_p) = (2p+1) \cdot p!$

Proof. Observe that $P(B_n^{(1)} = k) = \frac{D_{n,k}}{C_n}$. Let $\lambda_n^{(p)} = \sum_{k=1}^{n} D_{n,k}[k]_p$. We have that $E([B_n^{(1)}]_p) = \frac{\lambda_n^{(p)}}{C_n}$.

Consider $\frac{\partial^p D(t,z)}{\partial t^p}$: on one hand, we have:

$$\frac{\partial^p D(t,z)}{\partial t^p} = \sum_{n=0}^{\infty}\sum_{k=p}^{n}[k]_p D_{n,k} t^{k-p} z^n$$

Set $t = 1$, it becomes:

$$\begin{aligned}\frac{\partial^p D(t,z)}{\partial t^p}|_{t=1} &= \sum_{n=0}^{\infty}\left(\sum_{k=p}^{n}[k]_p D_{n,k}\right) z^n \\ &= \sum_{n=0}^{\infty}\lambda_n^{(p)} z^n\end{aligned}$$

This is exactly the generating function for $\lambda_n^{(p)}$.

On the other hand, we may compute this p-th partial derivative explicitly by using the formula obtained in Lemma 4.3.2:

$$\begin{aligned}D(t,z) &= (1 - tzC(z))^{-1} \\ \frac{\partial D(t,z)}{\partial t} &= zC(z)(1 - tzC(z))^{-2} \\ \frac{\partial^2 D(t,z)}{\partial t^2} &= 2(zC(z))^2(1 - tzC(z))^{-3} \\ &\vdots \\ \frac{\partial^p D(t,z)}{\partial t^p} &= p!(zC(z))^p(1 - tzC(z))^{-p-1}\end{aligned}$$

Set $t = 1$, and according to Lemma 2.1.2, $C(z) = 1 + zC(z)^2$. We have $1 - zC(z) = 1/C(z)$. Hence

$$\frac{\partial^p D(t,z)}{\partial t^p}|_{t=1} = p!z^p C(z)^{2p+1}$$

Combine these facts together, we reach the following:

$$\begin{aligned} E([\hat{Y}_n]_p) &= \frac{\lambda_n^{(p)}}{C_n} \\ &= \frac{[z^n]p!z^pC(z)^{2p+1}}{[z^n]C(z)} \\ &= p!\frac{[z^{n-p}]C(z)^{2p+1}}{[z^n]C(z)} \\ &= p!\frac{[z^{n-p}]C(z)^{2p+1}}{[z^{n-p}]C(z)}\frac{[z^{n-p}]C(z)}{[z^n]C(z)} \end{aligned}$$

By Lemma 2.1.2, $C(z)$ has convergence radius $1/4$, so $\lim_{n\to\infty}\frac{[z^{n-p}]C(z)}{[z^n]C(z)} = \frac{1}{4^p}$

Apply the Proposition 4.2.1 discussed in the previous section:

$$\lim_{n\to\infty}\frac{[z^{n-p}]C(z)^{2p+1}}{[z^{n-p}]C(z)} = f'(C(1/4)) = (2p+1)2^{2p}$$

Hence

$$\lim_{n\to\infty} E([B_n^{(1)}]_p) = p! \cdot (2p+1)2^{2p} \cdot \frac{1}{4^p} = (2p+1)p!$$

□

Recall Corollary 2.5.14 that if the factorial moments of a sequence of probability measures ν_n converges to a probability measure ν which is uniquely determined by its moments, then ν_n converges weakly to ν. Theorem 4.3.3 computes the asymptotic factorial moments of these random variables $B_n^{(1)}$. This raises a question: which distribution does $B_n^{(1)}$ converges to?

Now fix n, and consider $B_n^{(1)}$:

Lemma 4.3.4. $P(B_n^{(1)} = k) = \frac{\sum_{d_1+\cdots+d_k=n-k} C_{d_1}C_{d_2}\cdots C_{d_k}}{C_n}$. *Moreover,* $\lim_{n\to\infty} P(B_n^{(1)} = k) = \frac{k}{2\cdot 2^k}$.

Proof. Consider what happens when the vertex 1 is in a block of size k: this block divides $\{1, \cdots, n\}$ into k buckets. Let d_i be the number of points in each bucket. For each bucket, the points inside cannot connect to any other buckets, so there is C_{d_i} ways to partition them in a non-crossing way. Clearly, $\sum_{i=1}^{k} d_i = n - k$. This gives the first part of the lemma.

Observe that $\sum_{d_1+\cdots+d_k=n-k} C_{d_1}C_{d_2}\cdots C_{d_k} = [z^{n-k}]C(z)^k$. Apply Lemma 4.2.1 to get the second part of the result.

□

Lemma 4.3.4 gives a good candidate for us to verify, namely a distribution ν where $\nu(k) = \frac{k}{2\cdot 2^k}$ for all $k = 1, 2, \cdots$. It remains to verify the factorial moments of ν are indeed what we had in Theorem 4.3.3 and ν is uniquely determined by its moments, after which we can conclude that ν is the limiting distribution of the total number of Outer blocks.

Proposition 4.3.5. *Let ν be a positive measure on $\mathbb{N}$ determined by the requirement that $\nu(\{k\}) = \frac{k}{2\cdot 2^k}$ for all $k = 1, 2, \cdots$. Then ν is a probability measure with finite moments of all orders, where the p-th factorial moments of ν is $(2p+1)p!$ for all $p \in \mathbb{N}$.*

Proof. We first check the following equation, from which we may prove ν is a probability measure with the desired moments.

$$\sum_{k=1}^{\infty} \frac{k}{2\cdot 2^k}[k]_p = (2p+1)p! \tag{4.3.1}$$

Let $f(z) = \frac{1}{1-z} = \sum_{k=0}^{\infty} z^k$. Then consider its p-th derivative. On one hand, one can easily check $\frac{d^p}{dz^p} f(z) = p!(1-z)^{-p-1}$. On the other hand,

$$\frac{d^p}{dz^p} f(z) = \sum_{k=0}^{\infty} [k]_p z^{k-p}$$

Hence put $z = 1/2$,

$$\sum_{k=0}^{\infty} [k]_p \frac{1}{2^{k-p}} = p! 2^{p+1}$$

Multiply both sides by $\frac{1}{2^{p+1}}$, we obtain

$$\sum_{k=1}^{\infty} \frac{1}{2 \cdot 2^k} [k]_p = p! \tag{4.3.2}$$

Now $k \cdot [k]_p = (k-p) \cdot [k]_p + p \cdot [k]_p$, and thus we may apply Equation 4.3.2 to obtain:

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{k}{2 \cdot 2^k} [k]_p &= \sum_{k=1}^{\infty} \frac{1}{2 \cdot 2^k} [k]_{p+1} + p \cdot \sum_{k=1}^{\infty} \frac{1}{2 \cdot 2^k} [k]_p \\ &= (p+1)! + p \cdot p! \\ &= (2p+1)p! \end{aligned}$$

Now to verify ν is uniquely determined by its moments, consider its moment generating function $\int_{\mathbb{N}} e^{tx} d\nu(x)$:

$$\begin{aligned} \int_{\mathbb{N}} e^{tx} d\nu(x) &= \sum_{k=1}^{\infty} \frac{k}{2 \cdot 2^k} e^{tk} \\ &= \frac{1}{2} \sum_{k=1}^{\infty} k \cdot \left(\frac{e^t}{2} \right)^k \end{aligned}$$

Recall $\frac{1}{1-z} = \sum_{k=0}^{\infty} z^k$ for all $|z| < 1$. Taking the derivative we get

$$\frac{1}{(1-z)^2} = \sum_{k=0}^{\infty} k \cdot z^{k-1}$$

Hence

$$\begin{aligned} \int_{\mathbb{N}} e^{tx} d\nu(x) &= \frac{1}{2} \sum_{k=1}^{\infty} k \cdot \left(\frac{e^t}{2} \right)^k \\ &= \frac{e^t}{4} \frac{1}{(1-e^t/2)^2} \end{aligned}$$

which converges for all t where $|t| < ln(2)$. It follows from the Proposition 2.5.11 that ν is uniquely determined by its moments.

□

Remark 4.3.6. The probability measure ν where $\nu(k) = \frac{k}{2 \cdot 2^k}$ comes from shifting the negative binomial distribution $NB(2, 1/2)$ by 1. Recall a negative binomial distribution $NB(n, p)$ assigns probability $\binom{k+n-1}{k}(1-p)^n p^k$ at each $k = 0, 1, 2, \cdots$.

Chapter 5

Further Discussions

5.1 Further Research

This section discusses some further questions regarding the block structure of random non-crossing partitions.

First of all, there are many other statistics considered for random permutations that have natural analog in the context of non-crossing partitions. For example, if $L_n : \mathcal{S}_n \to \mathbb{N}$ be the random variable that gives the length of the longest cycle, then the limit $\lim_{n\to\infty} \frac{E(L_n)}{n}$ converges to an integral $\int_0^\infty \exp(-x - \int_x^\infty e^{-y}/y dy)dx$, which is approximately equal to $0.62432965\cdots$ [16]. We can ask the same question for non-crossing partitions:

Question 5.1.1. Let $L_n : NC(n) \to \mathbb{N}$ be the random variable which gives the size of the largest block (i.e., for $\pi = \{V_1, V_2, \cdots, V_k\} \in NC(n)$, $L(\pi) = \max |V_i|$). Then, what is the asymptotic behavior of L_n when it is restricted to non-crossing partition?

Another question we would like to ask comes from our motivation of studying outer blocks. Recall, one of the motivations to study outer blocks is that it breaks a non-crossing partitions into layers of self similar parts. One can define the depth of a non-crossing partition to be the number of outer block layers in a non-crossing partition. For example, the non-crossing partition shown in Figure 4.1.1 contains three layers of outer blocks, so we define it to have depth 3. A natural question to ask is:

Question 5.1.2. What is the asymptotic distribution of the "depth" of a non-crossing partition?

As a generalization of our main results, one natural question to ask is the joint distribution of the total number of blocks or outer blocks sorted by size. Denote $X_n^{(k)} : NC(n) \to \mathbb{N}$ the random variable that counts the total number of blocks of size k, and $Y_n^{(k)} : NC(n) \to \mathbb{N}$ the one that counts the total number of outer blocks of size k. The more general question would be:

Question 5.1.3. Fix a positive integer k, and consider the joint distribution of the random vectors $(X_n^{(1)}, X_n^{(2)}, \cdots, X_n^{(k)})$ and $(Y_n^{(1)}, Y_n^{(2)}, \cdots, Y_n^{(k)})$. What are their asymptotic joint distributions?

For random permutations, it is not hard to check for any k, the number of k-cycles is asymptotically Poisson distributed with mean $\frac{1}{k}$. We have numerical evidence that, in contrary to that for random permutations, the total number of blocks of size k in a random non-crossing partition will have asymptotically a normal distribution. However, no proof is known so far for general k, and even for the simplest case when $k = 1$, we shall show in Section 5.2 that we do not have the tool to solve its asymptotic distribution yet.

The rest of this chapter will focus on the simplest case of Question 5.1.3: Section 5.2 discusses the asymptotic distribution of the total number of singletons (i.e. blocks of size 1), and Section 5.3 proves a result regarding the asymptotic distribution of the total number of outer singletons.

5.2 Asymptotic Distribution of Singleton Blocks

Let us denote by $S_n : NC(n) \to \mathbb{N}$ the random variable that counts the total number of singletons. It turns out that the factorial moments of S_n has a nice formula:

Proposition 5.2.1. *Let p be a fixed positive integer, then for every $n \geq p$, we have $E([S_n]_p) = [n]_p \frac{C_{n-p}}{C_n}$. If we denote $\gamma_n = \frac{nC_{n-1}}{C_n}$, then we can write this as $E([S_n]_p) = \gamma_n \gamma_{n-1} \cdots \gamma_{n-p+1}$*

Proof. Consider $E\left(\binom{S_n}{p}\right) = \frac{1}{p!}E([S_n]_p) = \sum_{\pi \in NC(n)} \frac{1}{C_n}\binom{S_n(\pi)}{p}$. For each $\pi \in NC(n)$, $\binom{S_n(\pi)}{p}$ counts the total number of $i_1 < i_2 < \cdots < i_p$ where $\{i_j\}$ is a block of π for each $j = 1, 2, \cdots, p$. So instead of summing over $\pi \in NC(n)$, we may switch it to a sum over all such indices of $i_1 < i_2 < \cdots < i_p$ in the following way:

$$\sum_{\pi \in NC(n)} \binom{S_n(\pi)}{p} = \sum_{i_1 < i_2 < \cdots < i_p} |\{\pi \in NC(n) : \text{each } \{i_j\} \text{ is a block of } \pi\}|$$

Now fix $i_1 < i_2 < \cdots < i_p$, the total number of π that contains $\{i_j\}$ as a block for all $j = 1, 2, \cdots, p$ can be easily seen to be $|NC(n-p)| = C_{n-p}$. There is a total of $\binom{n}{p}$ ways to fix such p distinct indices. Hence,

$$\begin{aligned} E([S_n]_p) &= p! E\left(\binom{S_n}{p}\right) \\ &= \frac{p!}{C_n} \sum_{i_1 < i_2 < \cdots < i_p} |\{\pi \in NC(n) : \text{each } \{i_j\} \text{ is a block of } \pi\}| \\ &= p! \binom{n}{p} \frac{C_{n-p}}{C_n} \\ &= [n]_p \frac{C_{n-p}}{C_n} \end{aligned}$$

One can check by setting $\gamma_n = \frac{nC_{n-1}}{C_n}$, this is exactly the product $\gamma_n \gamma_{n-1} \cdots \gamma_{n-p+1}$

$\square$

Remark 5.2.2. It is immediate that the expectation of S_n is asymptotically $n/4$, and its variance is $o(n^2)$. By Cheybeshev's inequality, one can check S_n concentrate in an interval of size $o(n)$ around its mean $n/4$. However, we are still interested in the behavior of S_n around its mean value.

Using computer programs, we can easily generate random non-crossing partitions for simulation purposes. Based on our simulation of 10000 randomly selected non-crossing partitions from $NC(1000)$, the observed S_n is "close" to a normal distribution, in the sense that its quantile matches with that of the normal distributions. In Figure 5.2.1, we used the software R to make histogram and QQ-plot (quantile of the sampled S_v v.s. the quantile of the standard normal distribution). One may observe in the QQ-plot, the sample points are close to a straight line. Roughly speaking, this implies the normalized S_n has its distribution function similar to the distribution function of the standard normal distribution.

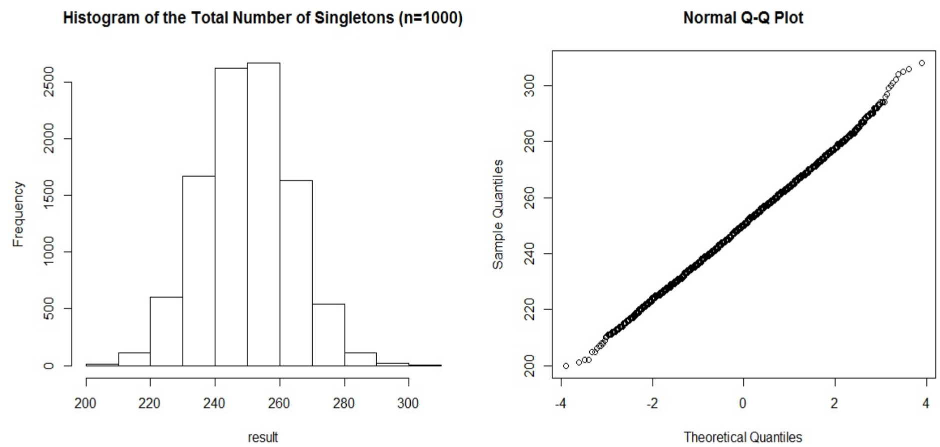

Figure 5.2.1: Histogram and QQ-plot for simulated S_n

Moments	Asymptotic Formula
$E(S_n - \frac{n}{4})$	$\frac{3}{4}$
$E\left(\left(S_n - \frac{n}{4}\right)^2\right)$	$\frac{3n}{16}$
$E\left(\left(S_n - \frac{n}{4}\right)^3\right)$	$\frac{39n}{128}$
$E\left(\left(S_n - \frac{n}{4}\right)^4\right)$	$\frac{27n^2}{256}$
$E\left(\left(S_n - \frac{n}{4}\right)^5\right)$	$\frac{765n^2}{2048}$
$E\left(\left(S_n - \frac{n}{4}\right)^6\right)$	$\frac{405n^3}{4096}$

Table 5.1: Centralized Moments of S_n

We can explicitly compute the first few 'centralized' moments, where S_n is shifted by its asymptotic mean value $n/4$. Here we record the asymptotic behavior of the first few centralized moments in Table 5.1

If we scale this centralized S_n by $\left(\frac{3n}{16}\right)^{-1/2}$, the resulting normalized random variable has first six moments matching the standard normal distribution. We have the following conjecture regarding the asymptotic distribution of S_n.

Conjecture 5.2.3. *Let $S_n : NC(n) \to \mathbb{N}$ count the total number of singletons in a non-crossing partition. Then*

$$\frac{S_n - n/4}{\sqrt{3n/16}} \xrightarrow{d} \mathcal{N}(0,1)$$

One could hope that Conjecture 5.2.3 can be proved using the same types of methods as we used for proving the main theorem of Chapter 3. However, recall that the proof in Chapter 3 was based on Proposition 2.2.9, which showed that the number of non-crossing partitions with a certain number of blocks can be enumerated by the Narayana number. Then the formula for Narayana numbers can be approximated with a quite sharp bound using the Lemma 2.4.6. The sharp bound on Narayana numbers directly leads to the proof of the main result (Theorem 1.2.1). Two natural questions to ask is

1. What is the analog of Proposition 2.2.9 to enumerate the total number of singletons?
2. Can this enumeration be well approximated similar to Lemma 2.4.6?

We have a analog of the Narayana number in the context of singletons which solves question (1), but we cannot solve question (2) yet due to lack of sharp approximations on this number. We solve the first question by using the Motzkin sum

Definition 5.2.4. The Motzkin sum M_n (sometimes called the Riordan number) is equal to the total number of $\pi \in NC(n)$ where π has no singleton block.

Proposition 5.2.5. *There exists a sequence of non-negative integers M_n, such that*

1. *The total number of $\pi \in NC(n)$ with k singletons is exactly $\binom{n}{k}M_{n-k}$*
2. *M_n satisfies the recursive formula $M_n = C_n - \sum_{k=1}^{n}\binom{n}{k}M_{n-k}$.*

Proof. Consider the procedure of constructing a $\pi \in NC(n)$ with k singletons: we first choose k points from $\{1, 2, \cdots, n\}$ which gives a total of $\binom{n}{k}$ choices. The remaining $n-k$ points needs to be connected using a non-crossing partition in $NC(n-k)$. Since we require π to have exactly k singletons, the non-crossing partition on $n-k$ points cannot have any singletons, which is exactly enumerated by M_{n-k}.

Now enumerate $\pi \in NC(n)$ based on its total number of singletons, we can obtain the desired recursive formula of M_n.

□

It is known that Motzkin sum satisfies the following properties [18].

Fact 5.2.6. 1. It has an explicit formula given by

$$M_n = C_n - \binom{n}{1}C_{n-1} + \binom{n}{2}C_{n-2} + \cdots + (-1)^n\binom{n}{n}C_0$$

where C_n is the n-th Catalan number

2. Its moment generating function $\sum_{n=0}^{\infty} M_n z^n = \frac{1+x-\sqrt{1-2x-3x^2}}{2x(1+x)}$, with convergence radius $\frac{1}{3}$.

However, we cannot find a nice approximation of the Motzkin sum that is similar to Lemma 2.4.6. Hence, the asymptotic behavior of the total number singletons is still unknown.

5.3 Asymptotic Distribution of Outer Singleton Blocks

Let us denote $O_n : NC(n) \to \mathbb{N}$ the random variable that counts the total number of outer singletons. The technique used in the Proposition 5.2.1 can be borrowed here to immediately obtain the following result:

Proposition 5.3.1. *For each positive integer p,* $\lim_{n\to\infty} E([O_n]_p) = \frac{(p+1)!}{2^p}$

Proof. Similar to the proof of Proposition 5.2.1,

$$E\left(\binom{O_n}{p}\right) = \frac{1}{p!}E([O_n]_p) = \sum_{\pi\in NC(n)} \frac{1}{C_n}\binom{O_n(\pi)}{p}$$

and rewrite the summation as:

$$\sum_{\pi\in NC(n)} \binom{O_n(\pi)}{p} = \sum_{i_1<i_2<\cdots<i_p} |\{\pi \in NC(n) : \text{each } \{i_j\} \text{ is an outer block of } \pi\}|$$

Now $i_1, i_2, \cdots, i_p$ divides the set $\{1, 2, \cdots, n\}$ into $p+1$ chunks. Let d_j be the number of points strictly contained in the j-th chunk. One can check $d_1 = i_1 - 1$, $d_2 = i_2 - i_1 - 1$, $d_3 = i_3 - i_2 - 1$, $\cdots$, $d_{p+1} = n - i_p$, and $\sum_{j=1}^{p+1} d_j = n - p$.

Now instead of summing over $i_1, i_2, \cdots, i_p$, it is equivalent of summing over all such d_j where $\sum_{j=1}^{p+1} d_j = n - p$.

Now fix a particular choice of d_j and corresponding i_k for which we compute

$$|\{\pi \in NC(n) : \text{each } \{i_j\} \text{ is an outer block of } \pi\}|$$

The points $i_1, i_2, \cdots, i_p$ are all outer blocks, so that no other block can connect two points from two different chunks. This immediately leads to a total of $C_{d_1}C_{d_2}\cdots C_{d_{p+1}}$ choices of π.

Hence, we have

$$\begin{aligned}
E([O_n]_p) &= p! \sum_{\pi \in NC(n)} \frac{1}{C_n} \binom{O_n(\pi)}{p} \\
&= p! \frac{\sum_{d_1+d_2+\cdots+d_{p+1}=n-p} C_{d_1}C_{d_2}\cdots C_{d_{p+1}}}{C_n}
\end{aligned}$$

Notice that $\sum_{d_1+d_2+\cdots+d_{p+1}=n-p} C_{d_1}C_{d_2}\cdots C_{d_{p+1}} = [z^{n-p}]C(z)^{p+1}$. Apply Proposition 4.2.1, we obtain

$$\begin{aligned}
\lim_{n\to\infty} E([O_n]_p) &= p! \lim n \to \infty \frac{\sum_{d_1+d_2+\cdots+d_{p+1}=n-p} C_{d_1}C_{d_2}\cdots C_{d_{p+1}}}{C_n} \\
&= p! \lim_{n\to\infty} \frac{[z^{n-p}]C(z)^{p+1}}{[z^{n-p}C(z)} \cdot \frac{[z^{n-p}C(z)}{[z^n C(z)} \\
&= p! \cdot (p+1)2^p \cdot \frac{1}{4^p} \\
&= \frac{(p+1)!}{2^p}
\end{aligned}$$

$\square$

If we can find a probability measure ν on $\mathbb{N}$ such that $\int_{\mathbb{N}} [x]_p d\nu = \frac{(p+1)!}{2^p}$ and ν is determined by its moments, one can immediately deduce from Corollary 2.5.14 that this sequence O_n converges in distribution to ν. The general technique to obtain the moment generating function from the factorial moments is through the probability generating function:

Definition 5.3.2. Let ν be a probability measure on $\mathbb{N}$ with mass p_k at each $k = 0, 1, \cdots$. Then the probability generating function of ν is defined as $G(z) = \sum_{k=0}^{\infty} p_k z^k$.

Remark 5.3.3. Recall from Definition 2.5.9 that the moment generating function $M(t)$ is defined by $M(t) = \sum_{k=0}^{\infty} p_k e^{kt}$. It is immediate that $M(t) = G(e^t)$.

Remark 5.3.4. It is immediate that $G(1) = \sum_{k=0}^{\infty} p_k = 1$. Assuming its convergence radius is great than 1 and thus we can consider its analytic expansion around $z = 1$: for any positive integer p, the p-th derivative evaluated at 1:

$$\frac{d^p}{dz^p} G(z) \mid_{z=1} = \sum_{k=0}^{\infty} p_k [k]_p$$

which is exactly the p-th factorial moment of ν. Hence if G_p are the p-th factorial moments of the measure ν, then

$$G(z) = \sum_{p=0}^{\infty} \frac{G_p}{p!} (z-1)^p$$

Combining these two observations, we can reach the following main result regarding the asymptotic distribution of O_n:

Theorem 5.3.5. *$O_n \xrightarrow{d} NB(1/3, 2)$, where $NB(1/3, 2)$ is a probability measure on $\mathbb{N}$ with mass $\frac{4}{9}(k+1)\frac{1}{3^k}$ at each $k = 0, 1, 2, \cdots$.*

Proof. Let $G(z) = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{(k+1)!}{2^k} (z-1)^k$. Then if O_n converges to a probability measure ν in distribution, we must have that the probability generating function of ν is $G(z)$. We first compute $G(z)$ explicitly:

$$\begin{aligned} G(z) &= \sum_{k=0}^{\infty} \frac{1}{k!} \frac{(k+1)!}{2^k} (z-1)^k \\ &= \sum_{k=0}^{\infty} \frac{(k+1)}{2^k} (z-1)^k \\ &= \frac{4}{(z-3)^2} \end{aligned}$$

The corresponding moment generating function is $M(t) = G(e^t) = \frac{4}{(3-e^t)^2}$. This matches the moment generating function of the negative binomial distribution $NB(2, \frac{1}{3})$, where it has probability mass $p_k = \frac{4}{9}(k+1)\frac{1}{3^k}$ at each $k = 0, 1, 2, \cdots$. The result follows from Theorem 2.5.8

$\square$

References

[1] P. Biane. Some properties of crossings and partitions. *Discrete Mathematics*, 175(1-3):41–53, 1997.

[2] B. Bollobas. *Random Graphs*. Cambridge University Press, 2001.

[3] M. Bożejko, M. Leinert, and R. Speicher. Convolution and limit theorems for conditionally free random variables. *Pacific J. Math*, 175:357–388, 1996.

[4] R.M. Dudley. *Real Analysis and Probability*. Cambridge University Press, 2002.

[5] P.H. Edelman. Chain enumeration and non-crossing partitions. *Discrete Mathematics*, 31(2):171–180, 1980.

[6] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[7] V. L. Goncharov. On the distribution of cycles in permutations. *Dokl. Akad. Nauk SSSR*, 35:299301, 1942.

[8] V. L. Goncharov. Some facts from combinatorics. *Izvestia Akad. Nauk. SSSR, Ser. Mat.*, 8:348, 1944.

[9] G. Kreweras. On the noncrossing partitions of a cycle. *Discrete Mathematics*, 1(4):333350, 1972.

[10] J. McCammond. Noncrossing partitions in surprising locations. *Amer. Math. Monthly*, 113:598–610, 2006.

[11] A. Nica. Multi-variable subordination distributions for free additive convolution. *Journal of Functional Analysis*, 257:428–463, 2009.

[12] A. Nica and R. Speicher. *Lectures on the Combinatorics of Free Probability*. Cambridge University Press, 2006.

[13] V. Reiner. Non-crossing partitions for classical reflection groups. *Discrete Mathematics*, 177(1-3):195–222, 1997.

[14] J.S. Rosenthal. *A First look at Rigorous Probability Theory*. World Scientific Publishing, 2006.

[15] G.C. Rota. On the foundations of combinatorial theory i. theory of mobius functions. *Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 2:340–368, 1964.

[16] L.A. Shepp and S.P. Lloyd. Ordered cycle lengths in a random permutation. *Transactions of American Mathematical Society*, 121(2):340–357, 1966.

[17] R. Simion. Noncrossing partitions. *Discrete Mathematics*, 217:367409, 2000.

[18] N. Sloane. The on-line encyclopedia of integer sequences. A005043. Motzkin sums: a(n) = (n-1)*(2*a(n-1)+3*a(n-2))/(n+1). Also called Riordan numbers or ring numbers.

[19] R.P. Stanley. *Enumerative Combinatorics, Volume 2*. Cambridge Studies in Advanced Mathematics, 2001.