Mikołaj Marciniak

DOCTORAL THESIS

Dynamic asymptotic combinatorics

Main supervisor prof. dr hab. Piotr Śniady
Foreign supervisor prof. Benoit Collins

NICOLAUS COPERNICUS
UNIVERSITY
IN TORUŃ

Interdisciplinary Doctoral School
“Academia Copernicana”
Faculty of Mathematics and Computer Science
Nicolaus Copernicus University in Toruń
Poland
11 January 2023

Preamble

Summary

The main branch of mathematics relevant to this dissertation is combinatorics. In addition, we also use elements of graph theory, maps (i.e., graphs on oriented surfaces) and probability. The results obtained by us are also of an asymptotic nature. The dissertation consists of three chapters, and each of them is a separate paper.

The first chapter is almost identical to the article entitled *"Hydrodynamic limit of the Robinson–Schensted–Knuth algorithm"* [Mar22a] published in the journal *Random Structures and Algorithms*. The starting point is the RSK algorithm, which in its best known form takes as input a sequence of numbers and returns as output a pair of Young tableaux with the same shape: the insertion tableau and the recording tableau. The main tool we use is the result of Romik and Śniady describing the asymptotic position of the last added box in the RSK algorithm applied to a random sequence of numbers with a fixed last number z. The problem we consider is the dynamics of the RSK algorithm from the viewpoint of a single box. More precisely, we consider the evolution over time of the position of the box with the number z in the insertion tableau when the RSK algorithm is applied to a sequence of random numbers containing the number z. We prove that when the length of the sequence tends to infinity, the typical trajectory after scaling converges uniformly in probability to a certain deterministic curve.

The second chapter is almost identical to the article entitled *"Quadratic Coefficients of Goulden–Rattan Character Polynomials"* [Mar22b] published in the journal *Annals of Combinatorics*. The starting point is the Kerov polynomials, which, for any Young diagram λ, present the normalized character of the symmetric group corresponding to the diagram λ, as some polynomial of the free cumulants of the diagram λ. The main problem considered is the Goulden–Rattan Conjecture of non-negativity of coefficients of Goulden–Rattan polynomials. By using the

combinatorial interpretation of the coefficients of Kerov polynomials given by Dołęga, Féray and Śniady [DFŚ10] and the relationship between these polynomials, the Goulden–Rattan Conjecture translates to certain inequalities concerning maps, i.e., graphs on oriented and connected surfaces. The main result is three bijections based on the newly defined operation "edge sliding", which prove the Goulden–Rattan Conjecture in a special case.

The third chapter is a shortened version of a preprint, currently in preparation, based on collaborative research with Piotr Śniady. This unfinished preprint is not yet publicly available, but when completed, we plan to publish it as a joint paper. The main problem we consider is the position of a new box when we apply one step of the RSK algorithm to a random tableau with a fixed shape and a random number. The main tool is the cumulative function of tableau F_T as the most natural random variable to study the position of the new box. In addition, we define several types of weighted graphs and identify with each graph a certain rational function. We also use the fact that certain algebraic identities correspond to some combinatorial operations on graphs. The main goal is to create a tool to study the random variable F_T when the size of the diagram tends to infinity. The main result is a closed formula for the cumulants of F_T. We believe that this formula will be necessary for proving more precise versions of the results proved by Romik and Śniady. We also hope that this formula, by its generality, will also be useful for studying the position of the new box for different shapes of diagrams and will allow formulating new asymptotic results.

Authors' contributions

Chapter 1, "Hydrodynamic limit of the Robinson–Schensted–Knuth algorithm". This chapter is essentially identical to an article *Hydrodynamic limit of the Robinson–Schensted–Knuth algorithm* with Mikołaj Marciniak as the sole author which was published in a peer-reviewed journal *Random Structures & Algorithms* and is available at `https://doi.org/10.1002/rsa.21016`.

Chapter 2, "Quadratic coefficients of Goulden-Rattan character polynomials". This chapter is essentially identical to an article *Quadratic coefficients of Goulden-Rattan character polynomials* with Mikołaj Marciniak as the sole author which was published in a peer-reviewed journal *Annals of Combinatorics* and is available as `https://doi.org/10.1007/s00026-022-00611-5`.

Chapter 3, "The cumulants of the cumulative function of a random Poissonized tableau with a fixed shape". This chapter is a truncated version of a preprint by Mikołaj Marciniak and Piotr Śniady. The authors' contribution to this chapter is as follows.

Sections 3.1 and 3.2 (*"Introduction"*, *"The general form of RSK insertion fluctuations"*) were written by Piotr Śniady; their aim is to place the result of the Candidate in a wider context.

Section 3.3 (*"The cumulative function of a tableau and its cumulants"* was written solely by Mikołaj Marciniak.

Section 3.4 (*"Anti-Pieri growth process"*) was written by Piotr Śniady. In particular, Lemma 3.4.1 and the idea to use the anti-Pieri growth process to study the moments and the cumulants of the random variable $F_T(u_0)$ comes from an unpublished preprint from 2013.

Section 3.5 (*"Decomposition into simple fractions"*) was written solely by Mikołaj Marciniak.

Section 3.6 (*"The moments of the cumulative function"*), which is devoted to the regularization trick (used in Sections 3.6.2 to 3.6.5), was written by Piotr Śniady.

Section 3.7 (*"Proof of Theorem 3.3.2"*) contains the proof of the main result; it was written by Mikołaj Marciniak.

Figures 3.1 to 3.11 were created by Piotr Śniady. Figures 3.12 to 3.22 were created by the Mikołaj Marciniak.

Funding

Mikołaj Marciniak was supported by Narodowe Centrum Nauki, grant number 2017/26/A/ST1/00189 and Narodowe Centrum Badań i Rozwoju, grant number POWR.03.05.00-00-Z302/17-00.

Acknowledgments

I would like to thank my supervisor, Professor Piotr Śniady, for his support, understanding, patience, positive outlook, guidance, valuable advice, and useful and constructive recommendations during writing this dissertation. I would like to thank also to Dr Jacinta Torres and Dr Stephen Moore for the help with the text editing.

Contents

Chapter 1

Hydrodynamic limit of the Robinson–Schensted–Knuth algorithm

1.1 Introduction

The final published version can be found in [Mar22a].

1.1.1 Notations

A *partition* of a natural number n is a break up of n into a sum $n = \lambda_1 + \lambda_2 + \cdots + \lambda_k$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$ are positive integer numbers. The vector $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ is usually used to denote a partition. Let $\lambda \vdash n$ denote that λ is a partition of a number n. A *Young diagram* $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ is a finite collection of boxes arranged in left-justified rows with the row length λ_j of the jth row. Thus the Young diagram λ is a graphical interpretation of the partition λ. A *Young tableau* is a Young diagram filled with numbers. If the entries strictly decrease along each column from top to bottom and weakly increase along each row from left to right, a tableau is called *semistandard*. A *standard Young tableau* is a semistandard Young tableau with n boxes which contains all numbers $1, 2, \ldots, n$. Figure 1.1 shows examples of a Young diagram and of a standard Young tableau.

The *Robinson–Schensted–Knuth (RSK) algorithm* is a bijective algorithm which takes a finite sequence of numbers of length n as the input and returns a pair of Young tableaux (P, Q) with the same shape $\lambda \vdash n$. The semistandard tableau P

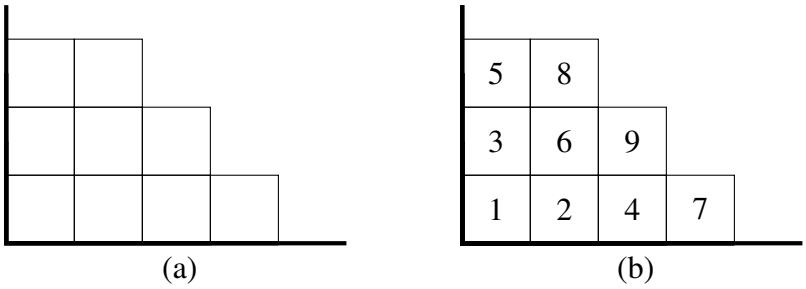


Figure 1.1: (a) The Young diagram of shape $(4, 3, 2) \vdash 9$ and (b) a standard Young tableau of shape $(4, 3, 2) \vdash 9$.

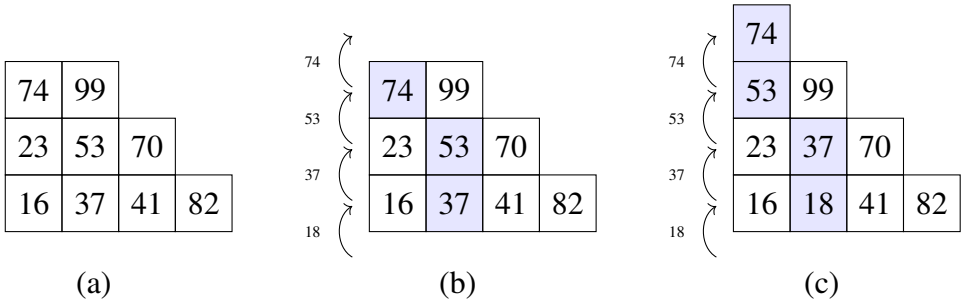


Figure 1.2: (a) The original tableau P. (b) The highlighted boxes form the bumping route which corresponds to an insertion of the number 18. The numbers next to the arrows indicate the bumped numbers.(c) The output of the RSK insertion step.

is called an *insertion tableau*, and the standard tableau Q is called a *recording tableau*. In particular, the RSK algorithm assigns to any permutation σ a pair (P, Q) of standard Young tableaux. A detailed description of the RSK algorithm can be found in [Rom15, Chapter 1.6].

The RSK algorithm is based on applying the *insertion step* to successive numbers from a given finite sequence $(X_j)_{j=1}^n$. The insertion step takes as input the previously obtained tableau $P\left(X_1, X_2, \ldots, X_{j-1}\right)$ and the next number X_j from the sequence. It produces as the output a new tableau $P\left(X_1, X_2, \ldots, X_j\right)$ with shape "increased" by one box; this tableau is obtained in the following way (see Figure 1.2). The RSK-insertion step starts in the first row with the number $x := X_j$. The insertion step consists of inserting the number x into the leftmost

box in this row containing a number y greater than x. We then move to the next row with the number y and repeat the action. At some row we are forced to insert the number at the end of the row, which will end the insertion step. The collection of rearranged boxes is called the *bumping route*.

We can say that the boxes with numbers are moved along the bumping route during an RSK insertion step.

1.1.2 Motivations

The RSK algorithm is an important tool in algebraic combinatorics, especially in the context of Littlewood–Richardson coefficients and the plactic monoid [FH91, Lecture 4].

For many years mathematicians have been studying the asymptotic behavior of the insertion tableau when we apply the Robinson–Schensted–Knuth algorithm to a random input. In the following paragraphs we will see several examples of such considerations.

The Ulam–Hammersley problem [Rom15, Chapter 1.1] concerns the typical length of the longest increasing subsequence in a random permutation. This problem corresponds to the problem of finding the typical length of the first row in the Young tableau obtained by the RSK algorithm from the sequence of independent random variables $(X_j)_{j=1}^n$ with the uniform distribution $U(0,1)$ on the unit interval $(0,1)$.

More generally, the RSK algorithm applied to the sequence of independent and identically distributed random variables with the uniform distribution $U(0,1)$ on the unit interval $(0,1)$ generates the *Plancherel measure* on Young diagrams [Rom15, Chapter 1.8]. The Plancherel measure is an important element of the representation theory of the symmetric groups because it describes how the left regular representation decomposes into irreducible components [FH91, Chapter 3.3].

Logan and Shepp [LS77] and Vershik and Kerov [KV86a] described the limit shape of the insertion tableau $P(X_1, X_2, \ldots, X_n)$ obtained when we apply the RSK algorithm to a random finite sequence.

Romik and Śniady [RŚ16] considered the limit shape of the bumping routes obtained when applying an RSK insertion step with a fixed number w to an existing insertion tableau $P(X_1, X_2, \ldots, X_n)$ obtained from a random finite sequence. In [RŚ15] they also considered the limit shape of *jeu de taquin* obtained from the recording tableau $Q(w, X_1, X_2, \ldots, X_n)$ made from a random finite sequence preceded by a fixed number w.

1.1.3 The main problem

This paper concerns the asymptotic behavior of the insertion tableau when we apply the RSK algorithm to a random input. *What can we say about the evolution over time of the insertion tableau from the viewpoint of box dynamics, when we apply the RSK algorithm to a sequence of independent random variables with the uniform distribution $U(0,1)$? How do the boxes move in the insertion tableau? If we investigate the scaled position of a box with a fixed number, will we get a deterministic limit, when the number of boxes tends to infinity?*

More specifically, we consider the insertion tableau $P\left(X_1, X_2, \ldots, X_n, w, X_{n+1}, \ldots, X_m\right)$ obtained by applying the RSK algorithm to a random finite sequence containing a fixed number w at some index. The box with this fixed number w is being bumped by the RSK insertion step along the bumping routes. We will describe the scaled limit position of the box with the number w depending on the ratio of the numbers m and n. This problem was also stated by Duzhin [Duz19].

Our main result for describing the scaled limit position is Theorem 1.3.1. A proof of Theorem 1.3.1 is given in Section 1.3.

1.2 Tools

1.2.1 The function $G(x)$

Following Romik and Śniady in [RŚ15, Chapter 5.1] we define the functions $F_{SC}, \Omega_\star, u, v$ needed to describe the macroscopic position of the new box in the RSK insertion step and additionally the function G

$$
\begin{aligned}
F_{SC}(y) &= \frac{1}{2} + \frac{1}{\pi}\left(\frac{y\sqrt{4-y^2}}{4} + \sin^{-1}\left(\frac{y}{2}\right)\right) && (-2 \le y \le 2)\,,\\
\Omega_\star(y) &= \frac{2}{\pi}\left(\sqrt{4-y^2} + y\sin^{-1}\left(\frac{y}{2}\right)\right) && (-2 \le y \le 2)\,,\\
u(x) &= F_{SC}^{-1}(x) && (0 \le x \le 1)\,,\\
v(x) &= \Omega_\star\left(u(x)\right) && (0 \le x \le 1)\,,\\
G(x) &= \left(\frac{v(x)+u(x)}{2}, \frac{v(x)-u(x)}{2}\right) && (0 \le x \le 1)\,.
\end{aligned}
$$

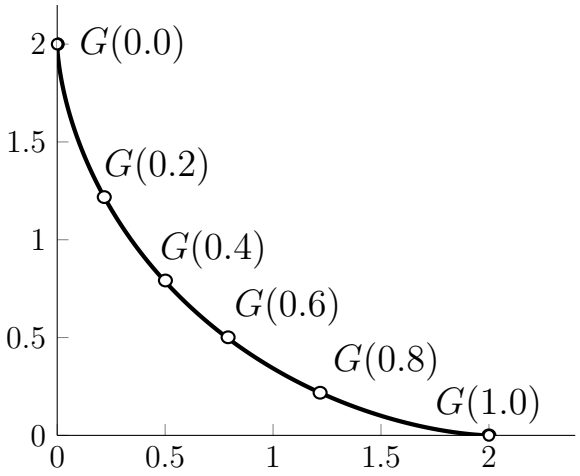


Figure 1.3: The graph of the function $G(x)$ with specified values for the argument $x = 0.0$, $x = 0.2$, $x = 0.4$, $x = 0.6$ $x = 0.8$ and $x = 1.0$.

The function F_{SC} is the cumulative distribution function of *Wigner's semicircle distribution*. The function $\Omega_\star$ is the limit shape of the Young tableau sampled from the *Plancherel measure*. This curve is called the Logan–Shepp–Vershik–Kerov curve [LS77; KV86a]. The function $x \mapsto \big(u(x), v(x)\big)$ is a special parameterization of the function $\Omega_\star$ and describes the limit position of the new box in the RSK insertion step applied with the number x to the random Young tableau [RŚ15, Chapter 5.1]. The continuous function $G\colon [0,1] \to \mathbb{R}^2_+$ is the function $\big(u(x), v(x)\big)$ rotated by 45 degrees (see [RŚ15, Chapter 5.1]).

Figure 1.3 shows the graph of the function $G\colon [0,1] \to \mathbb{R}^2_+$.

1.2.2 The result of Romik and Śniady

In the proof of Theorem 1.3.1 we will need the following result of Romik and Śniady [RŚ15, Theorem 5.1].

Let $(X_j)_{j=1}^{\infty}$ be a sequence of independent random variables with the uniform distribution $U(0,1)$. Let $\square_n(x) \in \mathbb{N}^2$ denote the position of the box with the maximal number in the recording tableau $Q\left(X_1, \ldots, X_n, x\right)$ when we apply the RSK insertion step for the number $x \in [0,1]$ to the previously obtained tableau $P\left(X_1, \ldots, X_n\right)$.

Theorem 1.2.1. *For each $x \in [0,1]$ the position $\square_n(x)$, after scaling by $\frac{1}{\sqrt{n}}$, converges in probability to a specific point $G(x) \in \mathbb{R}^2_+$, when n tends to infinity:*

$$\frac{\square_n(x)}{\sqrt{n}} \xrightarrow{p} G(x).$$

1.2.3 The partial order on the plane

We define *the partial order $\prec$ on the plane* as follows: $(x_1, y_1) \prec (x_2, y_2)$ if and only if $x_1 \leq x_2$ and $y_1 \geq y_2$.

For example the function G is *increasing with respect to the relation $\prec$*, that is. if $x_1 \leq x_2$ then $G(x_1) \prec G(x_2)$. Likewise, as each row of the insertion tableau is nondecreasing, for each natural number $n \in \mathbb{N}$ the function $\square_n$ is increasing with respect to the relation $\prec$, that is, if $x_1 \leq x_2$ then $\square_n(x_1) \prec \square_n(x_2)$.

1.2.4 The random increasing sequence

Let $w \in (0,1]$ be a fixed number and let $\left(X'_j\right)_{j=1}^{m'}$ be a finite sequence of independent random variables with the uniform distribution $U(0,w)$. For $j \in \{1,2,\ldots,m'\}$ we define $z(j)$ as the jth order statistic, that is, the jth smallest number among $X'_1, X'_2, \cdots, X'_{m'}$. The sequence $\left(z\left(j\right)\right)_{j=1}^{m'}$ contains all elements of the sequence $\left(X'_j\right)_{j=1}^{m'}$ in the ascending order. The sequence $\left(z\left(j\right)\right)_{j=1}^{m'}$ will be called a *random increasing sequence with the uniform distribution on the interval* $[0,w]$.

In addition, the sequence $\left(X'_j\right)_{j=1}^{m'}$ is some permutation $\Pi = (\Pi_1, \Pi_2, \ldots, \Pi_{m'})$ of the sequence $\left(z\left(j\right)\right)_{j=1}^{m'}$. Thus

$$\left(X'_j\right)_{j=1}^{m'} = \left(z(\Pi_j)\right)_{j=1}^{m'} = z \circ \Pi,$$

where z is the function that acts pointwise on every element of the permutation Π. The permutation Π is a random permutation with the uniform distribution. We see that a finite sequence of independent random variables with the uniform distribution $U(0,w)$ has the same distribution as a finite random increasing sequence with the uniform distribution on the interval $[0,w]$ permuted by a random permutation with the uniform distribution.

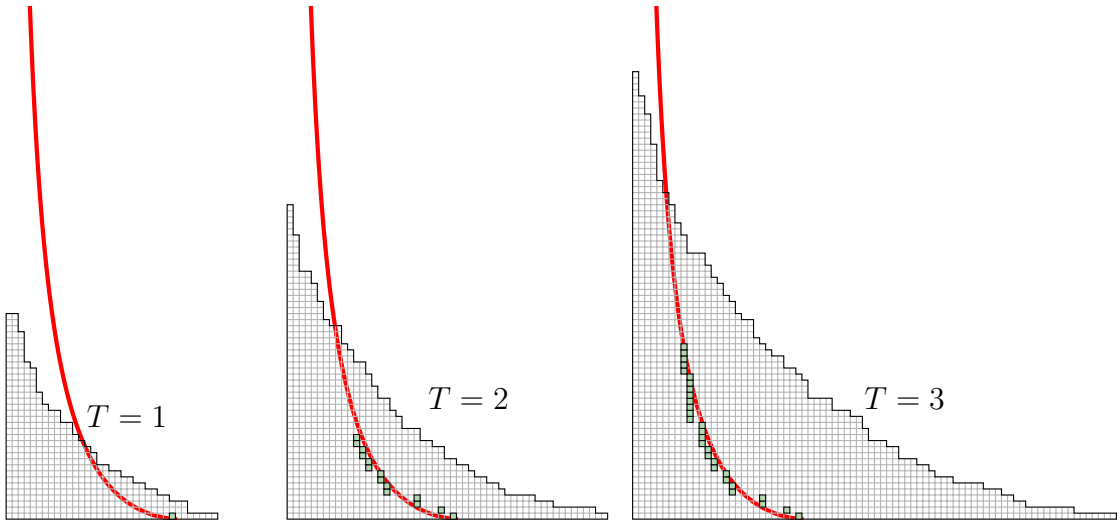


Figure 1.4: (a) The initial shape of the insertion tableau $P\left(X_1, \ldots, X_n, w\right)$ immediately after the new box with the number w was added (the higlighted box in the bottom row) for $n = 400$ and $w = 0.5$. (b) The shape of the insertion tableau $P\left(X_1, \ldots, X_n, w, X_{n+1}, \ldots, X_{\lfloor Tn \rfloor}\right)$ at the time parameter $T = 2$. The highlighted boxes indicate the trajectory of the box with the number w. The red smooth curve is the plot of H. (c) Analogous picture for $T = 3$.

1.3 The main result

1.3.1 Statement of main result

Our main result is Theorem 1.3.1 describing the asymptotic behavior of the box with a fixed number. It states that when the number of boxes tends to infinity then the (scaled down) trajectory of the box converges in probability to the curve $H\colon [1, \infty) \to \mathbb{R}^2_+$ given by

$$H(T) := \sqrt{T}\, G\left(\frac{1}{T}\right).$$

The same curve H also happens to be the limit shape of the bumping routes [RŚ16] in the RSK algorithm. Figure 1.4 shows the graph of the curve H and the experimentally determined trajectory of the box with the number $w = 0.5$.

More specifically, let $w \in (0,1]$ be a fixed number. Let $(X_j)_{j=1}^{\infty}$ be a sequence of independent random variables with the uniform distribution $U(0,1)$ on the unit interval $[0,1]$. For every $n \in \mathbb{N}$ we define the function $\mathrm{Pos}_n : \{n+1, n+2, \ldots\} \to \mathbb{N}^2$ by:

$$\mathrm{Pos}_n(j) = \mathrm{box}_w\left(P\left(X_1, \ldots, X_n, w, X_{n+1}, \ldots, X_j\right)\right)$$

for $j \in \{n+1, n+2, \ldots\}$, where for a tableau P we denote by $\mathrm{box}_w(P) \in \mathbb{N}^2$ the coordinates of the box with the number w.

Theorem 1.3.1. *Let $R \in (1,\infty)$ be a real number. For each number $T \in [1,R]$ the random variable $\mathrm{Pos}_n\left(\lfloor Tn \rfloor\right)$, after scaling by $\frac{1}{\sqrt{wn}}$, converges in probability to $H(T)$, when n tends to infinity. Moreover, the convergence is uniform, that is, for each $\epsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{T\in[1,R]} \left\| \frac{\mathrm{Pos}_n\left(\lfloor Tn \rfloor\right)}{\sqrt{wn}} - H(T) \right\| > \epsilon\right) = 0.$$

1.3.2 The strategy of the proof of Theorem 1.3.1

The proof will be presented in consecutive subsections of this section. Here we present a sketch of the proof

First, we prove pointwise convergence in the following way.

- We reduce the case to numbers less than w.
- We restrict our attention only to the permutation generated by the numbers.
- We use the property of the RSK algorithm that the insertion tableau is equal to the recording tableau of the inverse permutation.
- We return to the sequence of numbers from the interval $[0,1]$.
- We apply the Romik–Śniady result about bumping routes to deduce the pointwise convergence.

Second, using Dini's second theorem we prove the uniform convergence.

1.3.3 Pointwise convergence

First, we prove only pointwise convergence, that is, we prove that for each number $T \in [1, \infty)$ and for each $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\left\|\frac{\mathrm{Pos}_n\left(\lfloor Tn \rfloor\right)}{\sqrt{wn}} - H\left(T\right)\right\| > \epsilon\right) = 0.$$

Proof. We apply the RSK algorithm to a random sequence of real numbers containing the number w and investigate the position of the box with the number w in the insertion tableau. Any insertion step applied to a number greater that w does not change the position of the number w in the tableau, so it is enough to consider only the subsequence containing numbers not greater than w.

Now we will use this observation in the proof. Let $m = \lfloor Tn \rfloor$. The probability that the same number occurs twice in the sequence $(w, X_1, X_2, \ldots)$ is equal to 0, hence without losing generality we assume that the numbers $w, X_1, X_2, \ldots$ are all different. Let $\left(X'_j\right)_{j=1}^{\infty}$ be the subsequence of the sequence $\left(X_j\right)_{j=1}^{\infty}$ containing all elements of the sequence $\left(X_j\right)_{j=1}^{\infty}$ which are less than w. The sequence $\left(X'_j\right)_{j=1}^{\infty}$ is a sequence of independent random variables with the uniform distribution $U(0, w)$.

Let $n' = n'(n)$ and $m' = m'(n)$ denote the number of elements, respectively, of the sequences $\left(X_j\right)_{j=1}^{n}$ and $\left(X_j\right)_{j=1}^{m}$ which are smaller than w. Then there is an equality:

$$\begin{aligned}\mathrm{Pos}_n\left(\lfloor Tn \rfloor\right) &= \mathrm{box}_w\left(P\left(X_1, \ldots, X_n, w, X_{n+1}, \ldots, X_m\right)\right)\\ &= \mathrm{box}_w\left(P\left(X'_1, \ldots, X'_{n'}, w, X'_{n'+1}, \ldots, X'_{m'}\right)\right).\end{aligned}$$

The random variable $n' = \sum_{j=1}^{n}[X_j < w]$ counts how many numbers from the sequence $\left(X_j\right)_{j=1}^{n}$ are smaller than w, so n' is a random variable with the binomial distribution with parameters n and w. We denote it $n' \sim B(n, w)$. Likewise, the random variable $m' - n'$ counts how many numbers from the sequence $\left(X_j\right)_{j=n+1}^{m}$ are smaller than w, so $m' - n' \sim B(m - n, w)$. Moreover, the random variables n' and $m' - n'$ are independent, because the random variables $X_1, X_2, \ldots$ are independent.

From the strong law of large numbers [Dur19, Theorem 2.4.1] we know that if

n tends to infinity then the following limits exist almost surely:

$$\lim_{n\to\infty}\frac{n'}{n}=\mathbb{E}\left[X_j<w\right]=\mathbb{P}\left(X_j<w\right)=w,$$
$$\lim_{n\to\infty}\frac{m'-n'}{m-n}=\mathbb{E}\left[X_j<w\right]=\mathbb{P}\left(X_j<w\right)=w.$$

Therefore, also the following limits exist almost surely

$$\lim_{n\to\infty}\frac{m'}{n'}=1+\lim_{n\to\infty}\frac{m'-n'}{n'}=1+\lim_{n\to\infty}\frac{m-n}{n}=T$$

and

$$\lim_{n\to\infty}\frac{m'}{n}=\lim_{n\to\infty}\frac{n'}{n}\frac{m'}{n'}=wT.$$

We define the function $z:\{1,2,\dots m'\}\cup\left\{m'+1,n'+\frac{1}{2}\right\}\to[0,1]$ that assigns to a number $j\in\{1,2,\dots,m'\}$, the jth smallest number among $X'_1,X'_2,\cdots,X'_{m'}$ and additionally $z\left(m'+1\right)=w$ and $z\left(n'+\frac{1}{2}\right)=\frac{z\left(n'\right)+z\left(n'+1\right)}{2}$.

From the property of the random increasing sequence with the uniform distribution (Section 1.2.4) we have:

$$\left(X'_j\right)_{j=1}^{m'}=\left(z\left(\Pi_j\right)\right)_{j=1}^{m'}=z\circ\Pi,$$

where $\left(z\left(j\right)\right)_{j=1}^{m'}$ is a random increasing sequence with the uniform distribution on the interval $[0,w]$ and Π is a random permutation of range m' with the uniform distribution. The function z acts pointwise on every element of the permutation Π. Similarly, the function z acts on the insertion tableau by acting on each box individually. The function z by acting on the input of the RSK algorithm changes pointwise the numbers in the insertion tableau, since the function z is increasing. Then

$$\begin{aligned}\mathrm{Pos}_n\left(m\right)&=\mathrm{box}_w\left(P\left(X'_1,\dots,X'_{n'},w,X'_{n'+1},\dots,X'_{m'}\right)\right)\\&=\mathrm{box}_w\left(P\left(z\left(\Pi_1\right),\dots,z\left(\Pi_{n'}\right),w,z\left(\Pi_{n'+1}\right),\dots,z\left(\Pi_{m'}\right)\right)\right)\\&=\mathrm{box}_w\left(z\circ P\left(\Pi_1,\dots,\Pi_{n'},m'+1,\Pi_{n'+1},\dots,\Pi_{m'}\right)\right)\\&=\mathrm{box}_{m'+1}\left(P\left(\Pi_1,\dots,\Pi_{n'},m'+1,\Pi_{n'+1},\dots,\Pi_{m'}\right)\right).\end{aligned}$$

We denote the inverse permutation to Π by $\Pi^{-1} = \left(\Pi_1^{-1}, \ldots, \Pi_{m'}^{-1}\right)$ and we define the permutation

$$\Pi\uparrow = \left(\Pi_1, \ldots, \Pi_{n'}, m'+1, \Pi_{n'+1}, \ldots, \Pi_{m'}\right)$$

as a natural extension of the permutation Π. Then $\Pi\uparrow^{-1} = \left(\Pi\uparrow_1^{-1}, \ldots, \Pi\uparrow_{m'+1}^{-1}\right)$ where

$$\Pi\uparrow_j^{-1} = \begin{cases} \Pi_j^{-1} & \text{if} \quad \Pi_j^{-1} \leq n', \\ \Pi_j^{-1} + 1 & \text{if} \quad n' < \Pi_j^{-1} \leq m', \\ n'+1 & \text{if} \quad j = m'+1. \end{cases}$$

In addition, we will use the fact [Sch63] that for any permutation $\Pi\uparrow$ the insertion tableau of $\Pi\uparrow$ is equal to the recording tableau of the inverse permutation $\Pi\uparrow^{-1}$:

$$P\left(\Pi\uparrow\right) = Q\left(\Pi\uparrow^{-1}\right).$$

Therefore

$$\begin{aligned} \mathrm{Pos}_n\left(m\right) &= \mathrm{box}_{m'+1}\left(P\left(\Pi_1, \ldots, \Pi_{n'}, m'+1, \Pi_{n'+1}, \ldots, \Pi_{m'}\right)\right) \\ &= \mathrm{box}_{m'+1}\left(Q\left(\Pi\uparrow^{-1}\right)\right) \\ &= \mathrm{box}_{m'+1}\left(Q\left(\Pi\uparrow_1^{-1}, \ldots, \Pi\uparrow_{m'}^{-1}, n'+1\right)\right). \end{aligned}$$

We want to rewrite this in terms of $\square_n(x)$ so that we can apply Theorem 1.2.1. Now, by replacing the number $n'+1$ with the number $n'+\frac{1}{2}$ we get the inverse permutation Π^{-1}

$$\begin{aligned} \mathrm{Pos}_n\left(m\right) &= \mathrm{box}_{m'+1}\left(Q\left(\Pi\uparrow_1^{-1}, \ldots, \Pi\uparrow_{m'}^{-1}, n'+1\right)\right) \\ &= \mathrm{box}_{m'+1}\left(Q\left(\Pi\uparrow_1^{-1}, \ldots, \Pi\uparrow_{m'}^{-1}, n'+\frac{1}{2}\right)\right) \\ &= \mathrm{box}_{m'}\left(Q\left(\Pi_1^{-1}, \ldots, \Pi_{m'}^{-1}, n'+\frac{1}{2}\right)\right). \end{aligned}$$

Since the function z is increasing, when acting on the input of the RSK algorithm it will not change the content of the recording tableau. Using the function

z, we will get a sequence of independent random variables with the uniform distribution $U(0,1)$. Indeed

$$\begin{aligned}\mathrm{Pos}_n(m) &= \mathrm{box}_{m'}\left(Q\left(\Pi_1^{-1},\ldots,\Pi_{m'}^{-1},n'+\frac{1}{2}\right)\right)\\ &= \mathrm{box}_{m'}\left(z\circ Q\left(\Pi_1^{-1},\ldots,\Pi_{m'}^{-1},n'+\frac{1}{2}\right)\right)\\ &= \mathrm{box}_{m'}\left(Q\left(z\left(\Pi_1^{-1}\right),\ldots,z\left(\Pi_{m'}^{-1}\right),z\left(n'+\frac{1}{2}\right)\right)\right).\end{aligned}$$

The permutation Π is a random permutation with the uniform distribution on S_n, so Π^{-1} is also a random permutation with the uniform distribution on S_n. If we act with a random permutation on a random increasing sequence with the uniform distribution we will get a sequence of independent random variables with the uniform distribution, thus the sequence

$$z\circ\Pi^{-1} = \left(z\left(\Pi_1^{-1}\right),\ldots,z\left(\Pi_{m'}^{-1}\right)\right)$$

is a sequence of independent random variables with the uniform distribution $U(0,w)$.

We define the random variable A_n and the sequence $(Y_j)_{j=1}^{m'}$:

$$\begin{aligned}Y_j &= \frac{z\left(\Pi_j^{-1}\right)}{w} \qquad \text{for } j\in 1,2,\ldots,m',\\ A_n &= \frac{z\left(n'+\frac{1}{2}\right)}{w} = \frac{z(n')+z(n'+1)}{2w}.\end{aligned}$$

The sequence $(Y_j)_{j=1}^{m'}$ is a sequence of independent random variables with the uniform distribution $U(0,1)$, and the random variable A_n converges almost surely

to the limit $A = \frac{1}{T}$:

$$\begin{aligned}\lim_{n\to\infty} A_n &= \lim_{n\to\infty} \frac{z\left(n'\right) + z\left(n'+1\right)}{2w} \\ &= \lim_{n\to\infty} \frac{z\left(n'\right)}{w} + \lim_{n\to\infty} \frac{z\left(n'+1\right) - z\left(n'\right)}{2w} \\ &= \lim_{n\to\infty} \frac{n'}{m'} + 0 \\ &= \frac{1}{T}.\end{aligned}$$

Therefore

$$\begin{aligned}\mathrm{Pos}_n\left(m\right) &= \mathrm{box}_{m'+1}\left(Q\left(z\left(\Pi_1^{-1}\right), \ldots, z\left(\Pi_{m'}^{-1}\right), z\left(n' + \frac{1}{2}\right)\right)\right) \\ &= \mathrm{box}_{m'+1}\left(Q\left(\frac{z\left(\Pi_1^{-1}\right)}{w}, \ldots, \frac{z\left(\Pi_{m'}^{-1}\right)}{w}, \frac{z\left(n'+\frac{1}{2}\right)}{w}\right)\right) \\ &= \mathrm{box}_{m'+1}\left(Q\left(Y_1, \ldots, Y_{m'}, A_n\right)\right) \\ &= \square_{m'}\left(A_n\right).\end{aligned}$$

Let

$$G_n(x) := \frac{\square_{m'}(x)}{\sqrt{Tnw}}.$$

The function $\square_{m'}$ is increasing with respect to the relation $\prec$ (of Section 1.2.3). Hence G_n is also increasing with respect to the relation $\prec$, that is, if $x_1 \leq x_2$, then

$$G_n(x_1) \prec G_n(x_2). \tag{1.3.1}$$

From Theorem 1.2.1 for each $x \in [0,1]$ the random variable $G_n(x)$ converges in probability to the limit $G(x)$, when n tends to infinity. Indeed

$$G_n(x) = \sqrt{\frac{m'}{n}\frac{1}{Tw}}\ \frac{\square_{m'}(x)}{\sqrt{m'}} \xrightarrow{p} \sqrt{\frac{Tw}{Tw}}\ G(x) = G(x).$$

□

1.3.4 The uniform convergence on the interval

In this section we prove the uniform convergence on the interval $[1, R]$.

Proof. For each x, the sequence $G_n(x)$ converges in probability to the limit $G(x)$, when n tends to infinity. Both coordinates of each function $G_n(x)$ are monotonic and the limit function $G(x)$ is continuous. Then from Dini's second theorem we get that for each x, the sequence $G_n(x)$ converges uniformly in probability to the limit $G(x)$, when n tends to infinity.

Since

$$\left\|G_n(A_n) - G(A)\right\| \leq \left\|G_n(A_n) - G(A_n)\right\| + \left\|G(A_n) - G(A)\right\|,$$

the sequence $G_n(A_n)$ converges uniformly in probability to the limit $G(A)$, when n tends to infinity. Then for each $A \in [1, R]$ and each $\epsilon > 0$ we have:

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{A\in\left[\frac{1}{R},1\right]} \left\|G_n(A_n) - G(A)\right\| > \epsilon\right) = 0.$$

Using the inequality $T \leq R$ we obtain

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{A\in\left[\frac{1}{R},1\right]} \sqrt{T}\left\|G_n(A_n) - G(A)\right\| > \sqrt{R}\epsilon\right) = 0,$$

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{T\in[1,R]} \left\|\frac{\mathrm{Pos}_n\left(\lfloor Tn\rfloor\right)}{\sqrt{wn}} - H\left(T\right)\right\| > \sqrt{R}\epsilon\right) = 0.$$

Thus for each $\epsilon > 0$ we have

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{T\in[1,R]} \left\|\frac{\mathrm{Pos}_n\left(\lfloor Tn\rfloor\right)}{\sqrt{wn}} - H\left(T\right)\right\| > \epsilon\right) = 0.$$

□

1.3.5 Further questions

Instead of one box with the number w we can simultaneously observe the position of r boxes. If r is finite, then Theorem 1.3.1 describes the hydrodynamic behavior of all r boxes. Is this also true if r goes to infinity?

The function H behaves asymptotically for $T \to \infty$ like the function $\left(\frac{1}{\sqrt{T}}, 2\sqrt{T}\right)$. The function H describes the asymptotic position of the box with a fixed number w. The x-coordinate of this position after scaling tends to zero. If we did not scale it, would it also be similar? Will the box with the number w be moved to the first column? We expect that almost surely yes. How long will we have to wait for this? We expect that if the number w appears in the RSK algorithm after n steps, the waiting time for the number in the first column is $O(n^2)$. Is this really true? More formally, let $\left(X_j\right)_{j=1}^{\infty}$ be a sequence of independent random variables with the uniform distribution $U(0,1)$. What is the probability that the box with a fixed number w will be in the first column in the insertion tableau $P\left(X_1, \ldots, X_n, w, X_{n+1}, \ldots, X_{\lfloor Tn^2 \rfloor}\right)$ depending on parameter T?

Chapter 2

Quadratic Coefficients of Goulden–Rattan Character Polynomials

2.1 Introduction

The final published version can be found in [Mar22b].

2.1.1 Normalized characters

Characters are a basic tool of representation theory. After normalization, they are also useful in asymptotic problems.

If $k \leq n$ are natural numbers, then any permutation $\pi \in S_k$ can also be treated as an element of the larger symmetric group S_n by adding $n-k$ additional fixpoints. For any permutation $\pi \in S_k$ and any irreducible representation ρ^λ of the symmetric group S_n which corresponds to the Young diagram λ, we define *the normalized character*

$$\Sigma_\pi(\lambda) = \begin{cases} n(n-1)\cdots(n-k+1)\frac{\operatorname{Tr}\rho^\lambda(\pi)}{\text{dimension of } \rho^\lambda} & \text{for } k \leq n, \\ 0 & \text{otherwise.} \end{cases}$$

Of particular interest are the character values on the cycles, therefore we will use the shorthand notation

$$\Sigma_k(\lambda) = \Sigma_{(1,2,\ldots,k)}(\lambda).$$

2.1.2 Free cumulants

Free cumulants are an important tool of free probability theory [VDN92] and random matrix theory [Voi91]. In the context of the representation theory of the symmetric groups they can be defined as follows, see [Bia03]. For a Young diagram λ we define its free cumulants $R_2(\lambda), R_3(\lambda), \ldots$ as

$$R_k(\lambda) = \lim_{s\to\infty} \frac{1}{s^k}\Sigma_{k-1}(s\lambda),$$

where the diagram $s\lambda$ is created from the diagram λ by dividing each box of λ into an $s \times s$ square.

The free cumulants have been defined in such a way as to be very helpful for studying asymptotic behaviour of the characters on a cycle of length k when the size of the Young diagram tends to infinity [Bia98].

2.1.3 Kerov character polynomials

Kerov [Ker00] formulated the following result: for each permutation π and any Young diagram λ, the normalized character $\Sigma_\pi(\lambda)$ is equal to the value of some polynomial $K_\pi(R_2(\lambda), R_3(\lambda), \ldots)$ (now called *the Kerov character polynomial*) with integer coefficients. The first published proof of this fact was provided by Biane [Bia03]. The Kerov character polynomial is *universal* because it does not depend on the choice of λ. We are interested in the values of the characters on cycles; therefore, for $\pi = (1, 2, \ldots, k)$ we use the simplified notation

$$\Sigma_k = K_k(R_2, R_3, \ldots) \tag{2.1.1}$$

for such Kerov polynomials. The first few Kerov polynomials K_k are as follows:

$$\begin{aligned}
K_1 &= R_2,\\
K_2 &= R_3,\\
K_3 &= R_4 + R_2,\\
K_4 &= R_5 + 5R_3,\\
K_5 &= R_6 + 15R_4 + 5R_2^2 + 8R_2,\\
K_6 &= R_7 + 35R_5 + 35R_3R_2 + 84R_3,\\
K_7 &= R_8 + 180R_2 + 224R_2^2 + 14R_2^3 + 56R_3^2 + 469R_4 + 84R_2R_4 + 70R_6.
\end{aligned}$$

Kerov conjectured that the coefficients of the polynomial K_k are *non-negative* integers. Goulden and Rattan [GR07] found an explicit formula for the coefficients

of the Kerov polynomial K_k; unfortunately, their formula was complicated and did not give any combinatorial interpretation to the coefficients. Later, Féray proved positivity [Fér09] and together with Dołęga and Śniady found a combinatorial interpretation of the coefficients [DFŚ10]. In this paper, we will use the combinatorial interpretation given by them in the special case of linear and square coefficients.

2.1.4 Goulden–Rattan conjecture

Goulden and Rattan [GR07] introduced a family of functions $C_2, C_3, \dots$ on the set of Young diagrams given by $C_0 = 1$, $C_1 = 0$ and

$$C_k^\lambda = \frac{24}{k(k+1)(k+2)} \lim_{s\to\infty} \frac{1}{s^k}\Big(\Sigma_{k+1}(s\lambda) - R_{k+2}(s\lambda)\Big)$$

for $k \geq 2$.

Śniady [Śni06] proved the explicit form of C_k (conjectured by Biane [Bia03]) as a polynomial in the free cumulants $R_2, R_3, \dots$ given by

$$C_k = \sum_{\substack{j_2,j_3,\dots\geq 0 \\ 2j_2+3j_3+\dots=k}} (j_2+j_3+\dots)! \prod_{i\geq 2} \frac{\big((i-1)R_i\big)^{j_i}}{j_i!} \tag{2.1.2}$$

for $k \geq 2$. The aforementioned formula of Goulden and Rattan for the Kerov polynomials was naturally expressed in terms of these quantities $C_2, C_3, \dots$ [GR07]. More specifically, they constructed an explicit polynomial L_k with rational coefficients such that

$$K_k - R_{k+1} = L_k(C_2, C_3, \dots). \tag{2.1.3}$$

These polynomials are called *the Goulden–Rattan polynomials*. They formulated the following conjecture:

Goulden–Rattan conjecture. *The coefficients of the Goulden–Rattan polynomials are non-negative numbers with small denominators.*

The first few Goulden–Rattan polynomials are as follows [GR07]:

$$\begin{aligned}
K_1 - R_2 &= 0\\
K_2 - R_3 &= 0\\
K_3 - R_4 &= C_2,\\
K_4 - R_5 &= \frac{5}{2}C_3,\\
K_5 - R_6 &= 5C_4 + 8C_2,\\
K_6 - R_7 &= \frac{35}{4}C_5 + 42C_3,\\
K_7 - R_8 &= 14C_6 + \frac{469}{3}C_4 + \frac{203}{3}C_2^2 + 180C_2.
\end{aligned}$$

Linear coefficients of the Goulden–Rattan polynomials are non-negative, because they are equal to certain scaled coefficients of the Kerov polynomial:

$$[C_j]L_k = \frac{1}{j-1}[R_j]K_k.$$

In this paper we will prove that the coefficient of C_2^2 is non-negative. We hope that edge sliding we will define in this article will also be a useful tool in proving non-negativity of the square coefficients $[C_iC_j]L_k$. The next step towards the proof of Goulden–Rattan conjecture would be to understand the cubic coefficients $[C_iC_jC_u]L_k$; we hope that our methods will still be applicable there, nevertheless, there seem to be some difficulties related to the inclusion-exclusion principle.

2.1.5 Graphs on surfaces, maps and expanders

We will consider graphs drawn on an oriented surface. Each face of such a graph has some number of edges ordered cyclically by going along the boundary of the face and touching it with the right hand. We will call it *the clockwise boundary direction*. If we use the left hand and visit the edges in the opposite order, we will call it the counterclockwise boundary direction.

By *a map* we mean a bipartite graph drawn without intersections on an oriented and connected surface with minimal genus. The maps which we consider have a fixed choice of colouring of the vertices, i.e., each vertex is coloured black or white, with the edges connecting the vertices of the opposite colours. An example of a map is shown in Figure 2.1.

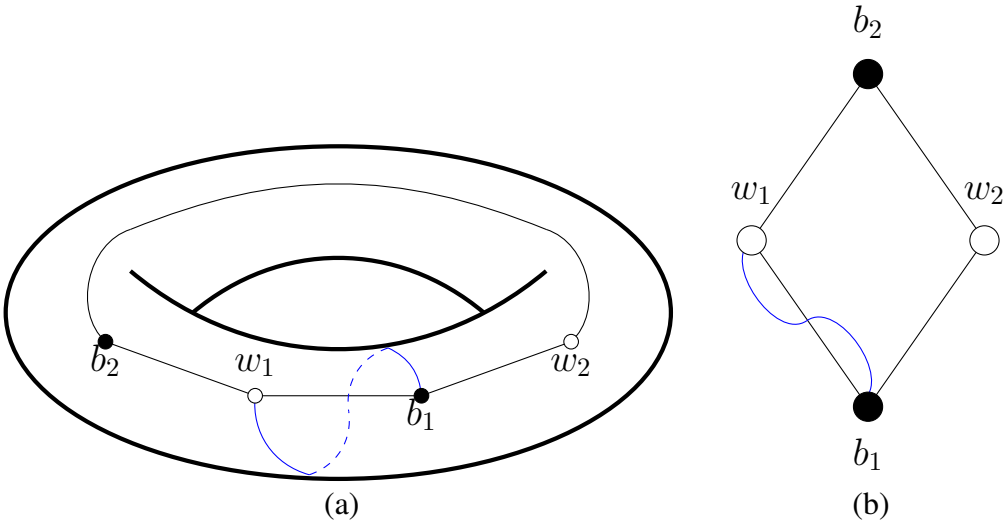


Figure 2.1: (a) An example of a map with 4 vertices and 5 edges drawn on a torus. (b) The same map drawn for simplicity on the plane.

An expander [Śni19, Appendix A.1] is a map with the following properties.

- It has a distinguished edge (known as the root) and one face.
- Each black vertex is assigned a natural number, known as a weight, such that each non-empty proper subset of the set of black vertices has more white neighbours than the sum of its weights.
- The sum of all weights is equal to the number of white vertices.

The map from Figure 2.1 is an expander if each black vertex has weight 1 (any choice of the root is valid).

Using the Euler characteristic we get

$$2 - 2g = \chi = V - k + 1 \tag{2.1.4}$$

where g denotes the genus of the surface, V denotes the number of vertices and k denotes the number of edges.

2.1.6 Combinatorial interpretation of the Kerov polynomial coefficients

The following two theorems Theorem 2.1.1 and Theorem 2.1.3 give a combinatorial interpretation to the linear and square coefficients of the Kerov character polynomials [DFŚ10, Theorem 1.2, Theorem 1.3]. The first is as follows.

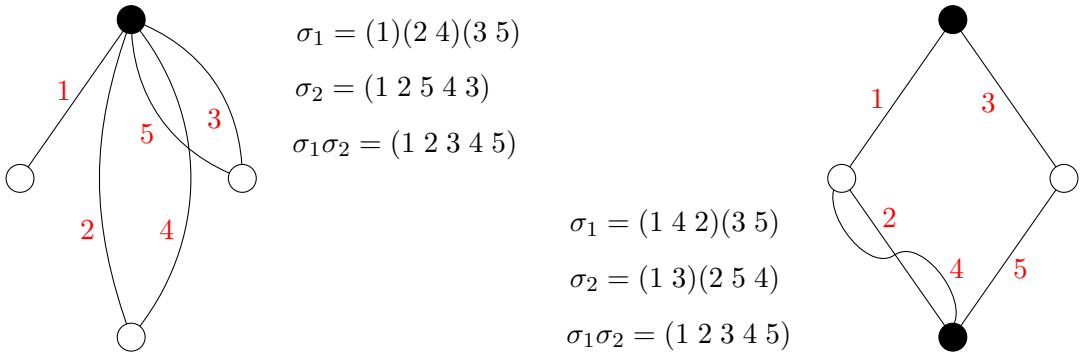
Theorem 2.1.1. *For all integers $l \geq 2$ and $k \geq 1$ the coefficient $[R_l]K_k$ is equal to the number of pairs (σ_1, σ_2) of permutations $\sigma_1, \sigma_2 \in S(k)$ such that $\sigma_1\sigma_2 = (1\ 2\ \cdots\ k)$ and such that σ_2 consists of one cycle and σ_1 consists of $l-1$ cycles. (We use the convention $\sigma_1\sigma_2 = \sigma_2(\sigma_1) = \sigma_2 \circ \sigma_1$.)*

The expanders are a graphical interpretation of these pairs of permutations. There is a natural bijection between such pairs of permutations (σ_1, σ_2) and the expanders with one face, one black vertex, $l-1$ white vertices and k edges. Additionally, one edge is selected as the root and the unique black vertex has a weight $l-1$. More precisely:

- The edges are numbered $1, 2, \ldots, k$. The edge with number 1 is selected as the root.
- *The counterclockwise angular cyclic order of the edges on a given vertex* (i.e., the order of edges ending at this vertex around it) corresponds to a cycle of a permutation depending on the colour of this vertex, i.e., σ_1 for white and σ_2 for black (in this case we have a unique cycle of the permutation σ_2).
- The unique face corresponds to the unique counterclockwise boundary cycle of the permutation $(1\ 2\ \cdots\ k)$.

Since there is only one face, the root determines the numbering of all edges. We can reformulate Theorem 2.1.1 as follows. (See Figure 2.2a for an example.)

Theorem 2.1.2. *For all integers $l \geq 2$ and $k \geq 1$ the coefficient $[R_l]K_k$ is equal to the number of expanders with k edges, $l-1$ white vertices and 1 black vertex with the weight $l-1$.*



(a) (b)

Figure 2.2: Examples of expanders with 5 edges and their corresponding pair of permutations σ_1, σ_2 such that $\sigma_1\sigma_2 = (1\ 2\ 3\ 4\ 5)$. The root is assigned the number 1. (a) The expander with one black vertex and three white vertices. (b) The expander with two black vertices and two white vertices.

Similarly we use the second theorem [DFŚ10, Theorem 1.3] for square coefficients.

Theorem 2.1.3. *For all integers $l_1, l_2 \geq 2$ and $k \geq 1$ the coefficient $[R_{l_1} R_{l_2}]K_k$ is equal to the number of triples (σ_1, σ_2, q) with the following properties.*

- *The permutations $\sigma_1, \sigma_2 \in S_k$ fulfill the equality $\sigma_1\sigma_2 = (1\ 2\ \cdots\ k)$.*
- *The permutation σ_1 consists of two cycles and the permutation σ_2 consists of $l_1 + l_2 - 2$ cycles.*
- *The function q associates the numbers l_1 and l_2 to the two cycles of σ_1. Furthermore, for each cycle c of σ_1 there exist at least $q(c)$ cycles of σ_2 which nontrivially intersect c.*

Analogously, we can also reformulate Theorem 2.1.3. (See Figure 2.2b for an example.)

Theorem 2.1.4. *For all integers $l_1, l_2 \geq 2$ and $k \geq 1$ the coefficient $[R_{l_1} R_{l_2}]K_k$ is equal to the number of expanders with k edges, $l_1 + l_2 - 2$ white vertices and 2 black vertices with weights $l_1 - 1, l_2 - 1$.*

2.1.7 Relationship between coefficients of Goulden–Rattan polynomials and coefficients of Kerov polynomials

The formula (2.1.2) allows us to express (C_i) in terms of free cumulants; we see that the coefficients of the terms R_iR_j, R_{i+j}, R_j^2 and R_{2j} in the expressions C_iC_j, C_{i+j}, C_j^2 and C_{2j} are given for $i \neq j$ by

$$\begin{aligned}
C_iC_j &= (i-1)(j-1)R_iR_j + 0R_{i+j} + (\text{sum of other terms}),\\
C_{i+j} &= 2(i-1)(j-1)R_iR_j + (i+j-1)R_{i+j} + (\text{sum of other terms}),\\
C_j^2 &= (j-1)^2R_j^2 + 0R_{2j} + (\text{sum of other terms}),\\
C_{2j} &= (j-1)^2R_j^2 + (2j-1)R_{2j} + (\text{sum of other terms}).
\end{aligned}$$

Moreover, any product $C_{i_1}C_{i_2}\cdots C_{i_t}$ of at least $t \geq 3$ factors does not contain any of the terms C_iC_j, C_{i+j}, C_j^2 and C_{2j}. It follows that the square coefficients of the Goulden–Rattan polynomial are related to the coefficients of the Kerov polynomial via

$$\begin{aligned}
\left.\frac{\partial^2 L_k}{\partial C_i \partial C_j}\right|_{0=C_1=C_2=\cdots} &= \frac{1}{(i-1)(j-1)}\left.\frac{\partial^2 K_k}{\partial R_i \partial R_j}\right|_{0=R_1=R_2=\cdots} - 2\left.\frac{\partial L_k}{\partial C_{i+j}}\right|_{0=C_1=C_2=\cdots}\\
&= \frac{1}{(i-1)(j-1)}\left.\frac{\partial^2 K_k}{\partial R_i \partial R_j}\right|_{0=R_1=R_2=\cdots} - \frac{2}{(i+j-1)}\left.\frac{\partial K_k}{\partial R_{i+j}}\right|_{0=R_1=R_2=\cdots}
\end{aligned}$$

for $i \neq j$. Whereas the quadratic coefficients are related via

$$\begin{aligned}
\left.\frac{\partial^2 L_k}{\partial C_j^2}\right|_{0=C_1=C_2=\cdots} &= \frac{1}{(j-1)^2}\left.\frac{\partial^2 K_k}{\partial R_j^2}\right|_{0=R_1=R_2=\cdots} - 2\left.\frac{\partial L_k}{\partial C_{2j}}\right|_{0=C_1=C_2=\cdots}\\
&= \frac{1}{(j-1)^2}\left.\frac{\partial K_k^2}{\partial R_j^2}\right|_{0=R_1=R_2=\cdots} - \frac{2}{(2j-1)}\left.\frac{\partial K_k}{\partial R_{2j}}\right|_{0=R_1=R_2=\cdots}
\end{aligned}$$

for any natural number j. Thus, we obtain the explicit formula for the square coefficients of the Goulden–Rattan polynomial:

$$[C_j^2]L_k = \frac{1}{(j-1)^2}[R_j^2]K_k - \frac{1}{2j-1}[R_{2j}]K_k \tag{2.1.5}$$

and

$$[C_iC_j]L_k = \frac{1}{(i-1)(j-1)}[R_iR_j]K_k - \frac{2}{i+j-1}[R_{i+j}]K_k \qquad \text{for } i \neq j. \tag{2.1.6}$$

2.2 The main result

Let $Y_k(u)$ denote the set of expanders with k edges, $u-1$ white vertices and one black vertex. Let $X_k(i,j)$ denote the set of expanders with k edges, $i+j-2$ white vertices and two black vertices with weights $i-1$ and $j-1$. Using Theorem 2.1.2 and Theorem 2.1.4 we can also reformulate the Goulden–Rattan conjecture for the square coefficients in terms of expanders, as follows.

Conjecture 2.2.1. *Let $i \neq j$ be natural numbers. Then*

$$(2j-1)\ \left\|X_k(j,j)\right\| \geq (j-1)^2\ \left\|Y_k(2j)\right\|$$

and

$$(i+j-1)\ \left\|X_k(i,j)\right\| \geq 2(i-1)(j-1)\ \left\|Y_k(i+j)\right\|$$

for any natural number k.

These inequalities are equivalent to the positivity of the coefficients $[C_j^2]L_k$ and $[C_iC_j]L_k$ respectively. In this text we prove only the first inequality in the special case $j=2$. We hope to present a proof of Conjecture 2.2.1 in its general form in a future paper.

Using (2.1.5) we can calculate several examples of the coefficient of C_2^2 of the Goulden–Rattan polynomials

$$\begin{aligned}
&[C_2^2]L_4 = 0-0=0,\\
&[C_2^2]L_5 = 5-\frac{1}{3}\cdot 15 = 0,\\
&[C_2^2]L_6 = 0-0=0,\\
&[C_2^2]L_7 = 224-\frac{1}{3}\cdot 469 = \frac{203}{3},\\
&[C_2^2]L_8 = 0-0=0.
\end{aligned}$$

Note that if k is even then $[C_2^2]L_k = 0$ because there does not exist an expander with 4 vertices and an even number of edges, since $2-2g = 2j-k+1$ by (2.1.4). Additionally, $[C_2^2]L_1 = 0$ and $[C_2^2]L_3 = 0$. Thus we can assume that the number of edges k is odd and $k \geq 5$.

Let

$$X_k = X_k(2,2), \tag{2.2.1}$$
$$Y_k = Y_k(4). \tag{2.2.2}$$

The set X_k consists of expanders with 2 black vertices and 2 white vertices such that each black vertex is connected with both white vertices; each black vertex necessarily has weight equal to 1. The set Y_k consists of expanders with one black vertex (which necessarily has weight 3) connected with all 3 white vertices. From now on we will omit the weights of the black vertices.

The main goal of this paper is to prove the following:

Theorem 2.2.2. *The inquality*

$$3\big\|X_k\big\| \geq \big\|Y_k\big\|$$

is true for any natural number k.

2.3 Edge sliding

In this section, we provide some necessary background details for the proof of Theorem 2.2.2. We will denote a transposition exchanging a and h by $(a\ h)$. For any set $H \subset \{1, 2, \dots\}$ such that $a, h \in H$, we can treat the transposition $(a\ h)$ as a permutation of the set H. (By adding fixed points to the transposition.) Therefore, for any permutation π of the set H, the products $\pi \circ (a\ h)$ and $(a\ h) \circ \pi$ are also permutations of the set H.

Let G be a graph with edges numbered $1, \dots, k$ drawn without intersections on an oriented and connected surface with minimal genus. There is a natural bijection between the graph G and a multiset of cycles M_G (cyclic permutations of a subset of $\{1, \dots, k\}$) such that each number j belongs to exactly two cycles. More precisely, the counterclockwise angular cyclic order of edges at a given vertex of the graph G (i.e., the order of edges ending at this vertex around it) corresponds to a single cycle of M_G. Two examples of a graph with its multiset of cycles are shown in Figure 2.3.

In this section, each end of every edge (or equivalently, each element of every cycle) will have one of three values assigned to it: clockwise direction, counterclockwise direction, or no direction. Note that the two ends of the same edge can be assigned different values. By default, we assume that a cycle element to which we have not assigned a direction has an assigned the value no direction.

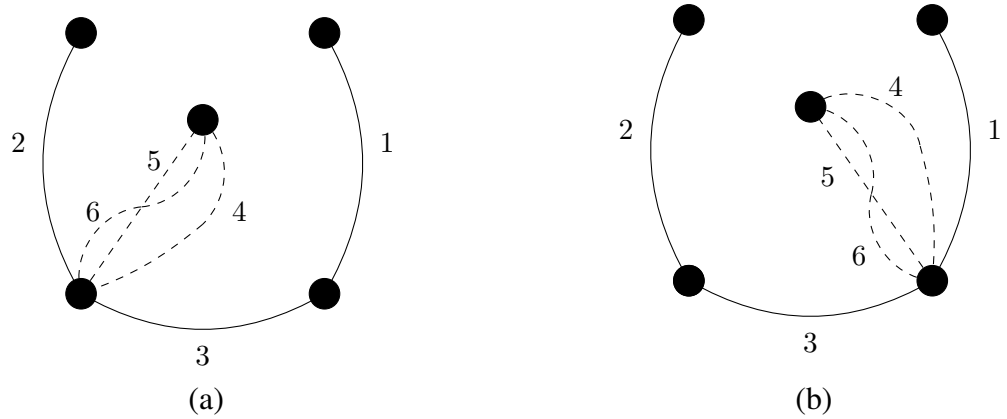


Figure 2.3: (a) The graph with the multiset of cycles $\big\{(2)(1)(5\ \ 6\ \ 4)(2\ \ 3\ \ 4\ \ 5\ \ 6)(3\ \ 1)\big\}$. (b) The graph with the multiset of cycles $\big\{(2)(1)(5\ 6\ 4)(2\ 3)(4\ 5\ 6\ 3\ 1)\big\}$.

2.3.1 Edge sliding for a single number

Let M_G be the multiset of cycles of a graph G. Let a belonging to a cycle $c \in M_G$ be a number with assigned the clockwise direction. (The number "a" from a cycle different from "c" and any number different from "a" from any cycle do not have a direction assigned to them.) Let $c' \neq c$ be the second cycle of the multiset M_G containing the number $h = c^{-1}(a)$. We assume that the number a does not belong to the cycle c'. We define *the single edge sliding for the number a in the clockwise boundary direction from the cycle c to the cycle c' along the number h* as the replacement of the cycles c, c' in the multiset M_G by two new cycles $s(c), s'(c')$ given by

$$s(c) = \big[c \circ (a\ h)\big] \setminus (a)\,,$$
$$s'(c') = (a\ h) \circ \big[(a)\ \ c'\big]\,.$$

Note that the product $c \circ (a\ h)$ consists of two cycles: the cycle (a) and the cycle formed from the cycle c by removing the number a. Therefore, the above transformation removes the number a from the cycle c and adds it to the cycle c' before the number h. Finally, we change the direction of the number a to the counterclockwise direction. We can naturally think of such operations as sliding of edges in a graph. An example of the single edge sliding is shown in Figure 2.4.

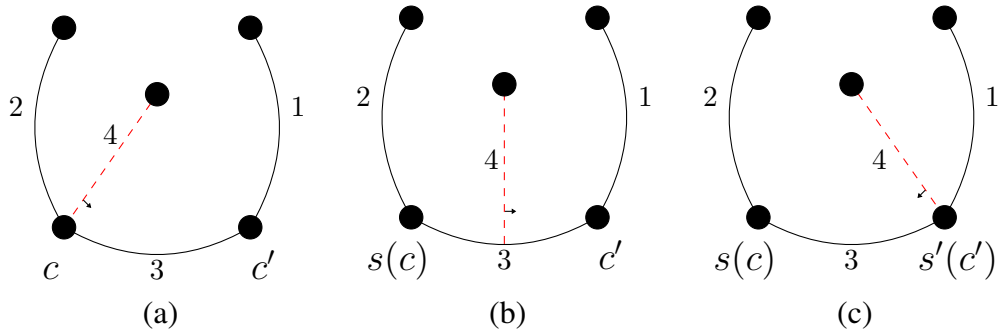


Figure 2.4: An example of the single edge sliding for the number $a = 4$ in the clockwise boundary direction from the cycle $c = (2\ 3\ 4)$ to the cycle $c' = (1\ 3)$ along the number $h = 3$. (a) The initial graph G. The edge labelled a is dashed and coloured red. (b) Visualization of the single edge sliding. (c) The resulting graph with two updated cycles: $s(c) = \big[(2\ 3\ 4) \circ (3\ 4)\big] \setminus (4) = (2\ 3)$ and $s'(c') = (3\ 4) \circ \big[(1\ 3)\ \ (4)\big] = (1\ 4\ 3)$.

Analogously, we define the single edge sliding for the number a in *the counterclockwise boundary direction* from the cycle c to the cycle c' along *the number* $h = c(a)$ as the replacement of the cycles c, c' in the multiset M_G by two new cycles $s(c), s'(c')$ given by

$$\begin{aligned} s(c) &= \big[(a\ h) \circ c\big] \setminus (a)\,, \\ s'(c') &= \big[(a)\ \ c'\big] \circ (a\ h). \end{aligned}$$

2.3.2 Edge sliding for a sequence of numbers

Let M_G be the multiset of cycles of a graph G. Let $a_1, \ldots, a_l$ be a sequence of successive but not all numbers belonging to a cycle $c \in M_G$, i.e. such that $a_j = c^{j-1}(a_1)$ for each index $j \leq l$. Moreover, to each of the numbers $a_1, \ldots, a_l$ is assigned the clockwise direction. Let $c' \neq c$ be the second cycle of the multiset of M_G containing the number $h = c^{-1}(a_1)$. We assume that the numbers $a_1, \ldots, a_l$ do not belong to the cycle c'. We define *the package edge sliding for the sequence of the numbers* $a_1, \ldots, a_l$ *in the clockwise boundary direction from the cycle* c *to the cycle* c' *along the number* h as the replacement of the cycles c, c' in the multiset M_G by two new cycles $s_l(c), s'_l(c')$ obtained by recursion with initial conditions

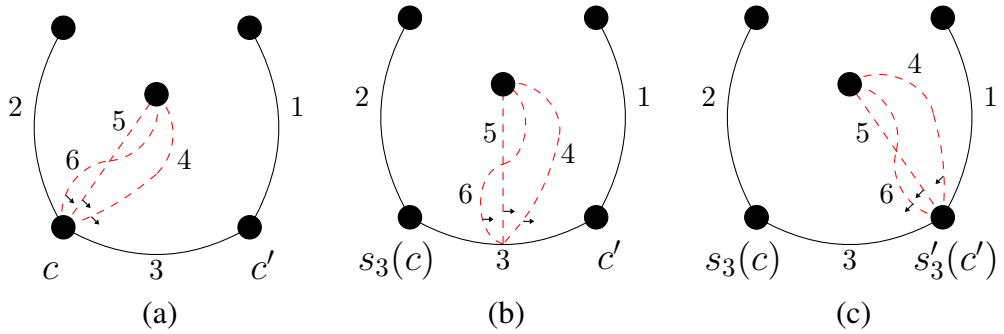


Figure 2.5: An example of the package edge sliding for the numbers $a_1 = 4$, $a_2 = 5$, $a_3 = 6$ in the clockwise boundary direction from the cycle $c = (2\ 3\ 4\ 5\ 6)$ to the cycle $c' = (1\ 3)$ along the number $h = 3$. (a) The initial graph G. Edges numbered a_1, a_2, a_3 are dashed and coloured red. (b) Visualization of the package edge sliding. (c) The resulting graph with two updated cycles: $s_3(c) = \big[(2\ 3\ 4\ 5\ 6) \circ (3\ 4) \circ (3\ 5) \circ (3\ 6)\big] \setminus \big[(4)\ (5)\ (6)\big] = (2\ 3)$ and $s'_3(c') = (3\ 6) \circ (3\ 5) \circ (3\ 4) \circ \big[(1\ 3)\ \ (4)\big] = (1\ 4\ 5\ 6\ 3)$.

$s_0(c) = c, s'_0(c') = c'$ and a recursive step given for each $1 < j \leq l$ by

$$s_j(c) = \Big[s_{j-1}(c) \circ \big(a_j\ h\big)\Big] \setminus \big(a_j\big)\,,$$
$$s'_j(c') = \big(a_j\ h\big) \circ \Big[\big(a_j\big)\ \ s'_{j-1}(c')\Big]\,.$$

Note that the package edge sliding for a sequence of numbers $a_1, \ldots, a_l$ is actually equivalent to the sequential single edge sliding for numbers $a_1, \ldots, a_l$. Therefore, the above transformation removes the numbers $a_1, \ldots a_l$ from the cycle c and adds them in the same order to the cycle c' before the number h. Finally, we change the directions of the numbers $a_1, \ldots, a_l$ to the counterclockwise boundary direction. Note that the single edge sliding is a special case of the package edge sliding. An example of the package edge sliding is shown in Figure 2.5.

Analogously, we define the package edge sliding for the numbers $a_1, \ldots, a_l$ in *the counterclockwise boundary direction* from the cycle c to the cycle c' along *the number* $h = c(a)$ as the replacement of the cycles c, c' in the multiset M_G by two new cycles $s_l(c), s_l(c')$ obtained by recursion with initial conditions $s_0(c) =$

$c, s'_0(c') = c'$ and a recursive step given for each $0 < j \leq l$ by

$$s_j(c) = \Big[\big(a_j\ h\big) \circ s_{j-1}(c)\Big] \setminus \big(a_j\big)\,,$$
$$s'_j(c') = \Big[\big(a_j\big)\ \ s'_{j-1}(c')\Big] \circ (a_j\ h).$$

In other words, if we would like to slide a number a in the clockwise boundary direction from a cycle c along a number h, then $c^{-1}(a) = h$ or before that, a number $c^{-1}(a)$ must be slid in the same direction. (Similarly for the counterclockwise boundary direction and number c(a).) For each number a in a cycle c with a fixed direction, the number h along which it will be slid is uniquely determined. Therefore, for simplicity, we will say that each of the numbers $a_1, \ldots, a_l$ from the cycle c is slid in a fixed direction. Of course, we assume that at least one number in the cycle c has no direction assigned.

2.3.3 Edge sliding in the general case

We define *the edge sliding* on a graph as follows. We start from the graph G in which some ends of certain edges are assigned a clockwise or counterclockwise boundary direction. We assume that:

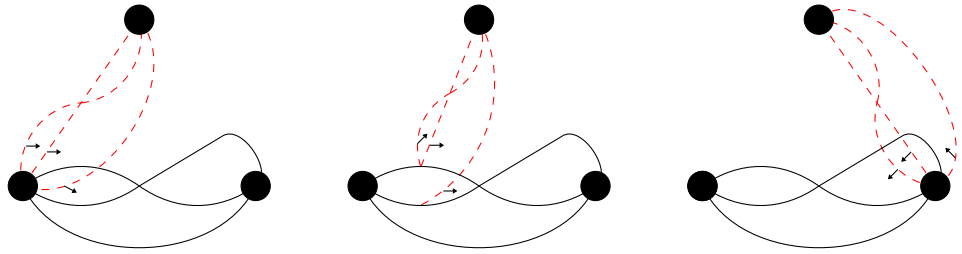
- At each vertex at least one edge has no direction assigned.
- Two consecutive edges do not have conflicting directions, i.e., there is no situation in which a number a from a cycle c has assigned the counterclockwise boundary direction and the number $c(a)$ has assigned the clockwise boundary direction.

Any such set of directions can be decomposed into an package edge sliding system for sequences of numbers:

$$\begin{cases} a_{1,1}, \ldots, a_{1,l_1} \text{ in a direction } d_1 \text{ from the cycle } c_1 \text{ to the cycle } c'_1 \text{ along the number } h_1, \\ \qquad\qquad\qquad\qquad\qquad\qquad \vdots \\ a_{t,1}, \ldots, a_{t,l_t} \text{ in a direction } d_t \text{ from the cycle } c_t \text{ to the cycle } c'_t \text{ along the number } h_t. \end{cases}$$

Furthermore, we require that:

- The numbers, along which other numbers are sliding, are not themselves sliding, i.e.,
$$\{a_{1,1}, \ldots, a_{t,l_t}\} \cap \{h_1, \ldots, h_t\} = \emptyset.$$



(a) (b) (c)

Figure 2.6: An example of edge sliding. (a) A graph with the slided edges dashed and coloured red. The directions of the edge ends are indicated by arrows. (b) The graph during the edge sliding in the clockwise boundary direction. (c) The graph after the edge sliding. Directions have already been reversed.

- Two ends of the same edge will not appear in the same vertex, i.e., if $a_{i_1,j_1} = a_{i_2,j_2}$, then
$$\{c_{i_1}, {c'}_{i_1}\} \cap \{c_{i_2}, {c'}_{i_2}\} = \emptyset.$$
(This condition can be weakened to ${c'}_{i_1} \neq {c'}_{i_2}$, but it is not necessary in this paper.)

- On one side of the edge along which we slide, the numbers do not slide in opposite directions, i.e., for each number h_j belonging to the cycles $c \neq c'$, there is no situation in which a number $c(h_j)$ has assigned the clockwise direction and the number ${c'}^{-1}(h_j)$ has assigned the counterclockwise direction.

We will call the selection of directions that satisfy the above conditions as *correct*.

We define the edge sliding as applying sequentially the package edge sliding for all sequences in any order. Such an action is well-defined, since permutation multiplication is associative and for any distinct numbers a_1, a_2, h_1, h_2 holds

$$(a_1\ h_1) \circ (a_2\ h_2) = (a_2\ h_2) \circ (a_1\ h_1).$$

An example of the edge sliding is shown in Figure 2.6.

The edge sliding is an involution on the set of graphs drawn on an oriented and connected surface with a correct selection of directions. Edge sliding is an invertible transformation, with the inverse also given by edge sliding.

In addition, it is easy to see that the edge sliding on a graph does not change the number of faces of this graph.

In the rest of this article, we will treat the edge sliding as transformation on a graph.

2.3.4 The set X_k of maps

We consider any map from the set X_k. This map has one face and an odd number of edges $k \geq 5$. We denote the black vertices by b_1, b_2 and the white vertices by w_1, w_2. There is at least one edge between each pair of the vertices of different colours. Of course, $\deg(b_1) + \deg(b_2) = k$ is an odd number. Without loss of generality we may assume that $\deg(b_1) > 0$ is an odd number and $\deg(b_2) > 0$ is an even number. Let $k_1, k_2 > 0$ denote the numbers of edges which connect the vertex b_1 with the vertices w_1, w_2, respectively. As $\deg(b_1) = k_1 + k_2$ is an odd number, without loss of generality we may assume that k_1 is even and k_2 is odd. For example, the unique (up to choice of the root) map from the set X_5 is shown in Figure 2.7a.

2.3.5 The set Y_k of maps

Let σ be the cycle that encodes the clockwise boundary cyclic order of the corners on the unique face of G. We will say that *the vertex w_j is a descendant of the vertex w_i* (we denote it by $w_i \to w_j$) if using the clockwise boundary order of the corners on the unique face of the map we can move (by walking along the edges and holding them with the right hand) in two steps from a certain corner c_i of the vertex w_i to a certain corner c_j of the vertex w_j, i.e., $\sigma^2(c_i) = c_j$.

We consider any map from the set Y_k. Any such map has one face and an odd number of edges $k \geq 5$. We denote the black vertex by b and the white vertices by w_1, w_2, w_3. We will write the set Y_k as a union of three sets which will be defined below.

Let $Y_k^{\mathrm{odd}} \subseteq Y_k$ be the set of maps for which there exists an *odd degree* white vertex (let us say it is w_3) which has the other two white vertices as descendants, i.e., $w_3 \to w_1$ and $w_3 \to w_2$. Let T_k^{odd} be the set of all maps from the set Y_k^{odd} with a distinguished vertex w_3 with this property. Moreover, each edge between the vertex w_3 and the vertex b, has a clockwise boundary direction assigned at the vertex b. The unique (up to choice of the root) map from the set T_5^{odd} is shown in Figure 2.7b. Clearly

$$|T_k^{\mathrm{odd}}| \geq |Y_k^{\mathrm{odd}}|. \tag{2.3.1}$$

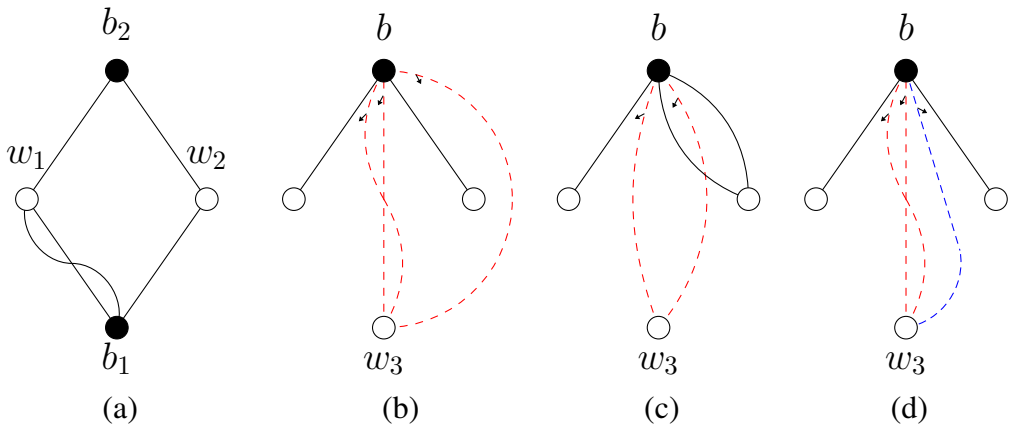


Figure 2.7: Examples of maps from the sets (a) X_5, (b) T_5^{odd}, (c) T_5^{even}, (d) T_5^{rest}. Slided edges are dashed and coloured. The directions are indicated by arrows, and the root is not marked.

Let $Y_k^{\text{even}} \subseteq Y_k$ be the set of maps such that there exists an *even degree* white vertex (let us say it is w_3) which has the other two white vertices as descendants, i.e., $w_3 \to w_1$ and $w_3 \to w_2$. Let T_k^{even} be the set of all the maps from the set Y_k^{even} with a distinguished vertex w_3 with this property. Moreover, each edge between the vertex w_3 and the vertex b, has a clockwise boundary direction assigned at the vertex b. The unique (up to choice of the root) map from the set T_5^{even} is shown in Figure 2.7c. Clearly

$$|T_k^{\text{even}}| \geq |Y_k^{\text{even}}|. \tag{2.3.2}$$

Let $Y_k^{\text{rest}} \subseteq Y_k$ be the set of maps not included in the sets Y_k^{odd} and Y_k^{even}, i.e.,

$$Y_k^{\text{rest}} = Y_k \setminus (Y_k^{\text{odd}} \cup Y_k^{\text{even}}). \tag{2.3.3}$$

Consider some map $m \in Y_k^{\text{rest}}$. Obviously $w_1 \to w_2 \to w_3 \to w_1$ or the other way around. Without loss of generality we may assume that $w_1 \to w_2 \to w_3 \to w_1$ and as a consequence $w_1 \nleftarrow w_2 \nleftarrow w_3 \nleftarrow w_1$.

Lemma. *The map m has a white vertex of odd degree, greater than* 1.

Proof. By contradiction, suppose this is not the case. The map m has at least one odd degree white vertex, because $\deg(w_1) + \deg(w_2) + \deg(w_3) = k$ is odd. Without loss of generality we may assume that $\deg(w_1)$ is odd. Since

$$\deg(w_1) + \deg(w_2) + \deg(w_3) = k > 3 = 1 + 1 + 1,$$

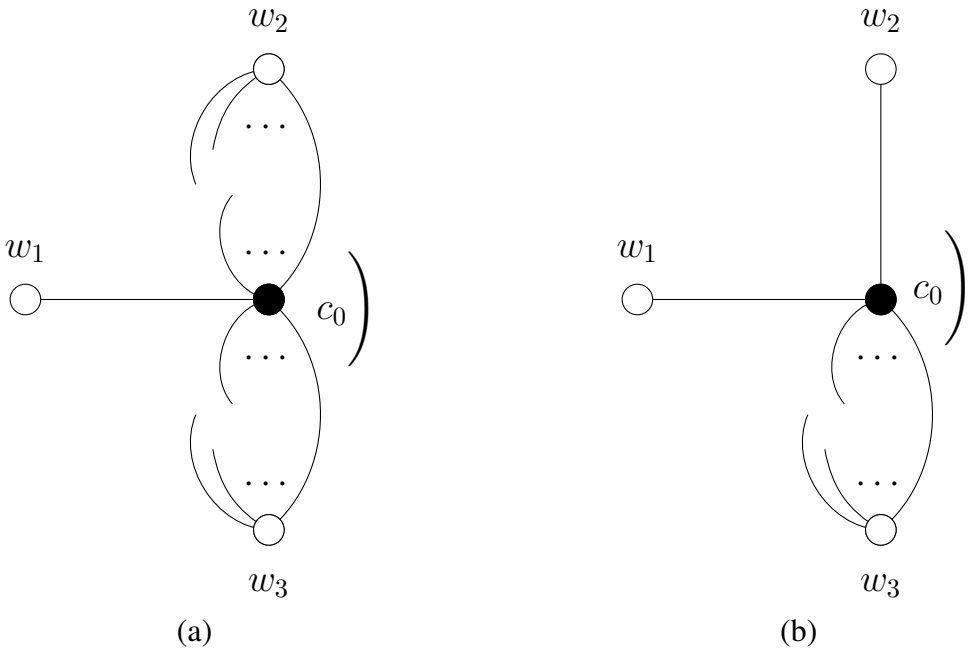


Figure 2.8: (a) The map m before replacement including the marked corner c_0. (b) A non-existent hypothetical map with four vertices and an even number of edges.

it follows that $\deg(w_1) = 1$ and $\deg(w_2), \deg(w_3)$ are even, because m does not have a white vertex with odd degree greater than 1. The vertex w_1 is a leaf and thus has a unique corner which we denote by c_1.

Naturally $\sigma^2(c_1)$ is a corner of the vertex w_2. Note that σ^2 is a permutation of the corners of the white vertices which has only one cycle, because the map m has only one face. The corners of the white vertices can be labelled 1, 2, 3 according to the names of the vertices they are in. If a corner c has the label a, its descendant $\sigma^2(c)$ has either the label a or $1 + a \bmod 3$. There is only one corner which has the label 1, so the corner labels of m (arranged in the cyclic order according to the unique cycle of σ^2) are $(1, 2, \ldots, 2, 3, \ldots, 3)$. Since there exists only one corner c_2 of the white vertex w_2 such that $\sigma^2(c_2)$ is a corner of the vertex w_3, then there exists a unique corner $c_0 = \sigma(c_2)$ of the vertex b such that $\sigma(c_0)$ is the corner of the vertex w_3 and $\sigma^{-1}(c_0)$ is the corner of the vertex w_2. Thus the clockwise angular cyclic order of the edges around the black vertex b is as follows: one edge connected to the vertex w_1, a certain number of edges connected to the vertex w_2,

a certain number of edges connected to the vertex w_3. Figure 2.8a visualizes this situation.

Let $k' = k - \big(\deg(w_2) - 1\big)$. The number k' is even because k is odd and $\deg(w_2)$ is even. We remove all edges except one of the white vertex w_2 obtaining a new graph m' with k' edges. Obviously, the clockwise angular cyclic order of the edges around the black vertex b is as follows: one edge connected to the vertex w_1, one edge connected to the vertex w_2, a certain number of edges connected to the vertex w_3. Whereas the corner labels of m' are $(1, 2, 3, \ldots, 3)$. Therefore, m' has one face. Figure 2.8b visualizes this situation. If m' does not have selected a root, we choose any edge of m' as the root. If the genus of the surface on which the map m' is drawn is not minimal, we draw the map m' on a surface with minimal genus. As a result, we obtain an expander from the set $Y_{k'}$ with 4 vertices, one face and an even number of edges k'. We get a contradiction because such a map does not exist (see (2.1.4)). Therefore, the map m has a white vertex with an odd degree greater than 1. □

Now, we fix the directions. To all ends in the vertex b of the edges between the vertex w_3 and b, we assign the direction in such way that among them there is an even number with the clockwise boundary direction and an odd number with the counterclockwise boundary direction. This is always possible, e.g. for a single edge with the counterclockwise boundary direction.

Let T_k^{rest} be the set of all the maps from the set Y_k^{rest} with a distinguished white vertex denoted by w_3 with a fixed choice of the set of special edges together with the directions of their ends satisfying the conditions just mentioned above. The unique (up to choice of the root) example of the map from the set T_5^{rest} is shown in Figure 2.7d. Clearly

$$|T_k^{\mathrm{rest}}| \geq |Y_k^{\mathrm{rest}}|. \tag{2.3.4}$$

2.4 Proof of main result

In this section we will construct three bijections which show the cardinalities of the corresponding sets are equal. Using these equalities and the definitions of these sets we will prove Theorem 2.2.2.

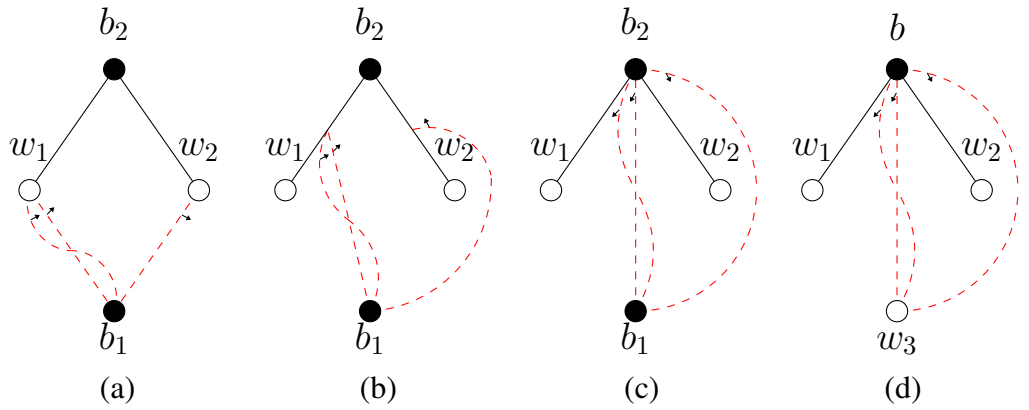


Figure 2.9: The example of the first bijection for the 5-edged map. (a) The map from the set X_5. (b) The map during the edge sliding. (c) The map after the edge sliding. (d) The map from the set Y_5^{odd}.

2.4.1 Three bijections

The first bijection between X_k and T_k^{odd}. We start from a map $m \in X_k$. Recall that we have assumed that $\deg(b_1)$ is odd. All edges connecting a vertex b_1 to white vertices, have the counterclockwise boundary direction assigned at the vertices w_1, w_2. The choice of directions is correct because m is a bipartite graph. We apply the edge sliding to the map m. Then we change the colour of the black vertex b_1 to white and its name to w_3, and the name of the vertex b_2 to b. Of course, the degree of the vertex w_3 does not change and is odd. In addition, $w_3 \to w_1$ and $w_3 \to w_2$, because any map from the set X_k has at least one edge between each pair of the vertices of different colours. We obtain a map from the set T_k^{odd}. (At all times one of the edges is selected as the root.) Moreover, each map from the set T_k^{odd} can be produced in this way. Such a transformation is a bijection between the set X_k and the set T_k^{odd}, since the edge sliding is reversible. Figure 2.9 shows an example of this bijection for $k = 5$. Thus

$$|X_k| = |T_k^{\text{odd}}|. \tag{2.4.1}$$

The second bijection between X_k and T_k^{even}. We start from a map $m \in X_k$. Recall that we have assumed that $\deg(b_2)$ is even. All edges connecting a vertex

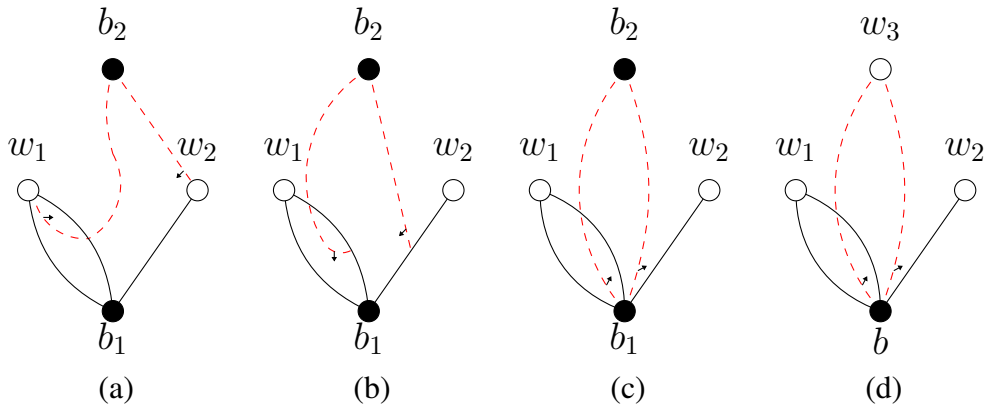


Figure 2.10: (a)—(d) The example of the second bijection for the 5-edged map.

b_1 to white vertices, have the counterclockwise boundary direction assigned at the vertices w_1, w_2. The choice of directions is correct because m is a bipartite graph. We apply the edge sliding to the map m. Then we change the colour of the black vertex b_2 to white and its name to w_3, and the name of the vertex b_1 to b. Of course, the degree of the vertex w_3 does not change and is even. In addition, $w_3 \to w_1$ and $w_3 \to w_2$, because any map from the set X_k has at least one edge between each pair of the vertices of different colours. We obtain a map from the set T_k^{even}. (At all times one of the edges is selected as the root.) Moreover, each map from the set T_k^{even} can be produced. Such a transformation is a bijection between the set X_k and the set T_k^{even}, since the edge sliding is reversible. Figure 2.10 shows an example of this bijection for $k = 5$. Thus

$$|X_k| = |T_k^{\text{even}}|. \tag{2.4.2}$$

The third bijection. We start from a map $m \in X_k$. Recall that we have assumed that $\deg(b_1)$ is odd. All edges connecting a vertex b_1 to white vertices, have the counterclockwise boundary direction assigned at the vertex w_1 and the clockwise boundary direction assigned at the vertex w_2. The choice of directions is correct because m is a bipartite graph. We apply the edge sliding to the map m. Then we change the colour of the black vertex b_1 to white and its name to w_3, and the name of the vertex b_2 to b. Of course, the degree of the vertex w_3 does not change and is odd. In addition, $w_3 \to w_1$ and $w_2 \to w_3$ (and $w_1 \to w_2$), because any map

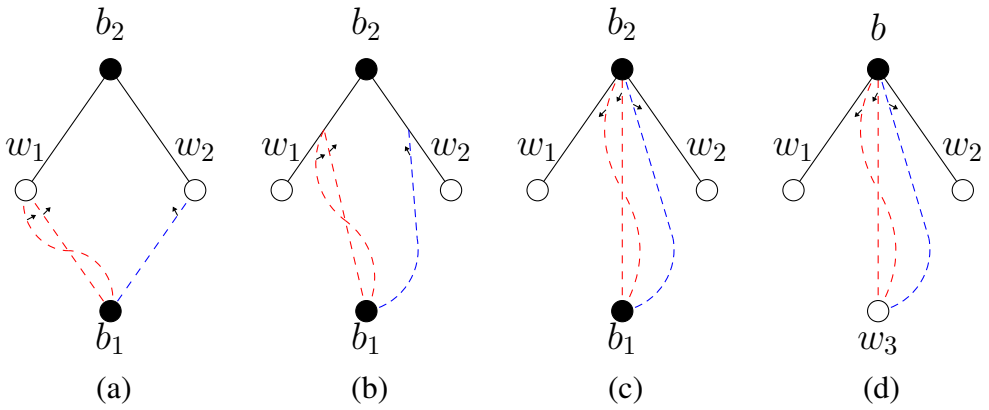


Figure 2.11: (a) - (d) The example of the third bijection for the 5-edged map.

from the set X_k has at least one edge between each pair of the vertices of different colours. We do not necessarily obtain a map from the set T_k^{rest} (it may be that we obtain a map from set T_k^{odd}), but it can be seen that each map from the set T_k^{rest} can be produced. (At all times one of the edges is selected as the root.) Such a transformation is a bijection between the set X_k and some superset of the set T_k^{odd}, since the edge sliding is reversible. Figure 2.11 shows an example of this bijection for $k = 5$. Thus

$$|X_k| \geq |T_k^{\text{rest}}|. \tag{2.4.3}$$

2.4.2 The conclusion of the proof

We can now proceed to the proof of Theorem 2.2.2, we have:

$$\begin{aligned}
3[C_2^2]L_k &= 3[R_2^2]K_k - [R_4]K_k && \text{by (2.1.5)}\\
&= 3|X_k| - |Y_k| && \text{by (2.2.1), (2.2.2)}\\
&\geq |T_k^{\text{odd}}| + |T_k^{\text{even}}| + |T_k^{\text{rest}}| - |Y_k| && \text{by (2.4.1), (2.4.2), (2.4.3)}\\
&\geq |Y_k^{\text{odd}}| + |Y_k^{\text{even}}| + |Y_k^{\text{rest}}| - |Y_k| && \text{by (2.3.1), (2.3.2), (2.3.4)}\\
&= |Y_k^{\text{odd}} \cap Y_k^{\text{even}}| && \text{by (2.3.3)}\\
&\geq 0.
\end{aligned}$$

□

Chapter 3

The cumulants of the cumulative function of a random Poissonized tableau with a fixed shape

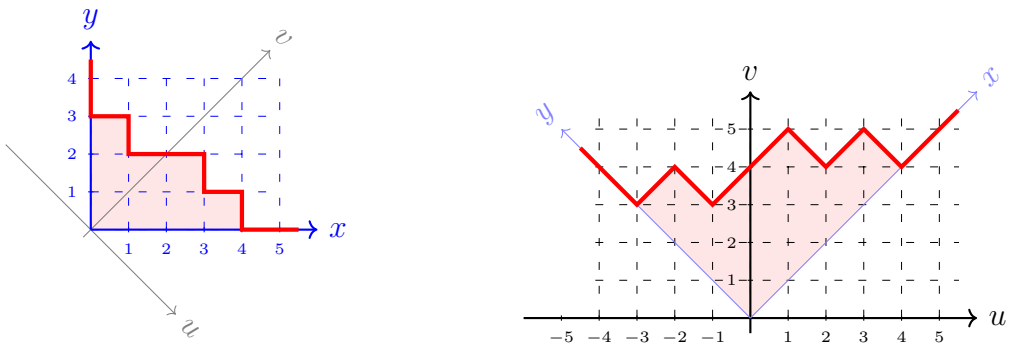
3.1 Introduction

3.1.1 Basic definitions

We start by recalling some basic combinatorial notions. For a more detailed treatment of the topic, we refer to [Ful97].

Young diagrams, tableaux

A Young diagram is a finite collection of boxes on the positive quarterplane, aligned to the left and to the bottom, see Figure 3.1a. This way of drawing Young diagrams is called *the French convention*. To a Young diagram with ℓ rows we associate the integer partition $\lambda = (\lambda_1, \ldots, \lambda_\ell)$, where λ_j denotes the number of the boxes in the j-th row (we count the rows from bottom to top). We identify a Young diagram with the corresponding partition λ and denote by $|\lambda| = \lambda_1 + \cdots + \lambda_\ell$ the number of its boxes.



(a) (b)

Figure 3.1: The Young diagram $(4, 3, 1)$ shown in (a) the French convention, and (b) the Russian convention. The solid red line represents the *profile* of the Young diagram. The coordinates system (u, v) corresponding to the Russian convention and the coordinate system (x, y) corresponding to the French convention are shown.

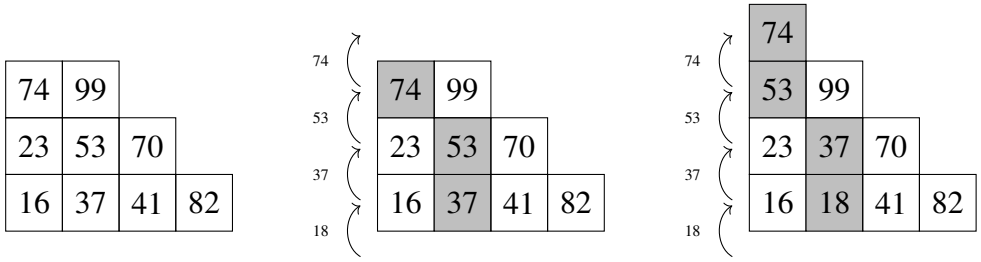
A *tableau* is a filling of the boxes of a Young diagram with numbers; we require that the entries should be weakly increasing in each row (from left to right) and strictly increasing in each column (from bottom to top). An example is given in Figure 3.2a. We say that a tableau T of shape λ is a *standard Young tableau* if it contains only entries from the set $\{1, 2, \ldots, |\lambda|\}$ and each element is used exactly once.

Schensted insertion

The *Schensted row insertion* is an algorithm which takes as an input a tableau T and some number x. The number x is inserted into the first row (i.e., the bottom row) of T in the leftmost box which contains an entry which is strictly bigger than x.

In the case when the row contains no entries which are bigger than x, the number x is inserted into the leftmost empty box in this row and the algorithm terminates.

If, however, the number x was inserted into a box which was not empty, the previous content x' of the box is *bumped* into the second row. This means that the algorithm is iterated but this time the number x' is inserted into the second row in the leftmost box which contains a number bigger than x'; if necessary this is repeated until some number is inserted into a previously empty box. This process is



(a) (b) (c)

Figure 3.2: (a) The original tableau T. (b) The highlighted boxes form the bumping route which corresponds to a Schensted insertion $T \leftarrow 18$. The numbers next to the arrows indicate the bumping entries. (c) The output $T \leftarrow 18$ of the Schensted insertion.

illustrated in Figures 3.2b and 3.2c. The outcome of Schensted insertion is defined as the new resulting tableau; it will be denoted by $T \leftarrow x$.

The *bumping route* consists of the boxes whose the content has changed by the action of Schensted insertion, see Figures 3.2b and 3.2c.

Robinson–Schensted–Knuth algorithm

For the purposes of this article we consider a simplified version of *Robinson–Schensted–Knuth algorithm*; for this reason we should rather call it *Robinson–Schensted algorithm*. Nevertheless, we use the first name because of its well-known acronym RSK. The RSK algorithm associates to a finite sequence $w = (w_1, \dots, w_n)$ a pair of tableaux: the *insertion tableau* $P(w)$ and the *recording tableau* $Q(w)$.

The insertion tableau

$$P(w) = \Big(\big((\emptyset \leftarrow w_1) \leftarrow w_2 \big) \leftarrow \cdots \Big) \leftarrow w_n \tag{3.1.1}$$

is defined as the result of iterative Schensted insertion applied to the entries of w, starting from the empty tableau $\emptyset$.

The recording tableau $Q(w)$ is defined as the standard Young tableau of the same shape as $P(w)$ in which each entry is equal to the number of the iteration of (3.1.1) in which the given box stopped being empty; in other words the entries of $Q(w)$ give the order in which the entries of the insertion tableau were filled.

Tableaux $P(w)$ and $Q(w)$ have the same shape; we will denote this common shape by $\mathrm{RSK}(w)$ and call it *the RSK shape associated to* w.

The RSK algorithm is an important tool of algebraic combinatorics and representation theory, especially in the context of the Littlewood–Richardson coefficients, see [Ful97; Sta99].

3.1.2 Context and motivations

A fruitful area of research is to investigate the RSK algorithm applied to a random input. We shall review below some selected highlights of this area.

Plancherel measure

The simplest example concerns the case when $w = (w_1, \ldots, w_n)$ is a uniformly random permutation in n letters. Since the recording tableau depends only on the relative order of the entries which form the input, the probability distribution of the recording tableau $Q(w)$ would not change if we replace the above probability distribution and take $w = (w_1, \ldots, w_n)$ to be a sequence of independent, identically distributed random variables with the uniform distribution $U(0, 1)$ on the unit interval $[0, 1]$. The corresponding probability distribution of $\mathrm{RSK}(w)$ is the celebrated *Plancherel measure* Pl_n *on the set of Young diagrams with* n *boxes* which appears naturally in the context of decomposition of *the left regular representation* of the symmetric group $\mathfrak{S}_n$ into irreducible components. This measure associates to a Young diagram λ (such that $|\lambda| = n$) the probability

$$\mathrm{Pl}_n(\lambda) = \frac{\left(f^\lambda\right)^2}{n!},$$

where f^λ denotes the number of standard Young tableaux of shape λ.

The most spectacular highlight related to the probability distribution of $\mathrm{RSK}(w)$ is the solution of *the Ulam–Hammersley problem* [BDJ99; Oko00] which shows a surprising link with *the Tracy–Widom distribution* which arises in random matrix theory. For a pedagogical introduction to this topic we recommend the book [Rom15].

Extremal characters of $\mathfrak{S}_\infty$

If $w_1, w_2, \dots$ is a sequence of independent, identically distributed random variables (possibly with a more complicated probability distribution which might have some atoms), and

$$\lambda^{(n)} = \mathrm{RSK}(w_1, \dots, w_n) \tag{3.1.2}$$

denotes the RSK shape corresponding to a prefix of size n, we may regard the growing sequence of Young diagrams

$$\emptyset = \lambda^{(0)} \nearrow \lambda^{(1)} \nearrow \cdots \tag{3.1.3}$$

as a Markov random walk in *the Young graph* which is a directed graph having the Young diagrams as the vertices and directed edges connecting pairs of diagrams which differ by exactly one box. This random walk (3.1.3) has some additional convenient properties which are out of scope of the current paper. Vershik and Kerov [VK81; KV86b] proved that a classification of the random walks with such additional properties (which is a problem on the boundary between probability theory, harmonic analysis on the Young graph, and ergodic theory) is equivalent to finding *the extremal characters of the infinite symmetric group* $\mathfrak{S}_\infty$. In this way Vershik and Kerov found a new, truly conceptual proof of Thoma's classification of such characters [Tho64]. As a byproduct, there is a convenient bijection between the extremal characters of $\mathfrak{S}_\infty$ and the probability laws of the random variables (w_n), and RSK provides a convenient, explicit way of generating the corresponding random walk (3.1.3).

The key motivation: is RSK an isomorphism of dynamical systems?

The recording tableau

$$Q_\infty(w_1, w_2, \dots) := \lim_{n\to\infty} Q(w_1, \dots, w_n)$$

corresponding to an *infinite* sequence $w_1, w_2, \dots$ is an *infinite standard tableau*, i.e., a filling of the boxes in (a subset of) the upper-right quarterplane such that each natural number appears exactly once. Just like its finite counterpart from Section 3.1.1, each entry of the infinite recording tableau $Q_\infty(w_1, w_2, \dots)$ is equal to the number of the iteration of an infinite sequence of row insertions

$$\big((\emptyset \leftarrow w_1) \leftarrow w_2\big) \leftarrow \cdots$$

in which the given box stopped being empty. This infinite recording tableau is another way of encoding the the infinite path (3.1.3) in the Young graph.

The aforementioned link between RSK and the Plancherel measure can be rephrased in a more abstract way as follows: Q_∞ is a homomorphism between the following two probability spaces:

- the infinite Cartesian power $[0,1]^\infty$ of the unit interval, equipped with the product of the Lebesgue measure (which clearly corresponds to a sequence $w_1, w_2, \dots$ of independent, identically distributed random variables with the uniform distribution on the unit interval),

and

- the set of infinite standard Young tableaux equipped with the *Plancherel measure* (which in light of the aforementioned results of Vershik and Kerov is fundamental for harmonic analysis on the infinite symmetric group $\mathfrak{S}_\infty$).

In fact, each of these two probability spaces can be equipped with a natural measure-preserving transformation (respectively, the *one-sided shift* and the *jeu de taquin transformation*) in such a way that Q_∞ becomes a homomorphism between two *measure-preserving dynamical systems*. It is natural to ask the following question.

Problem 3.1.1 (The key motivation)**.** *Is it true that Q_∞ is, in fact, an* isomorphism *of dynamical systems? If yes, how to construct the inverse map?*

The answer is not immediately obvious because in the finite case, when RSK is applied to a *finite* sequence w, in order to recover w we need information about the recording tableau $Q(w)$, *as well as about the insertion tableau* $P(w)$; the latter is *not* available in the infinite case. An affirmative answer to Problem 3.1.1 would shed light on some questions related to the harmonic analysis on $\mathfrak{S}_\infty$, for example whether jeu de taquin is ergodic.

These motivations were the starting point for the work of Romik and Śniady [RŚ15]; we present their findings in the following. In fact, Problem 3.1.1 has an affirmative answer.

3.1.3 The main problem: position of the new box

For asymptotic problems it is convenient to draw the Young diagrams in the *Russian convention*, see Figure 3.1b, which corresponds to the coordinate system (u, v) which is related to the usual (French) Cartesian coordinates by

$$u = x - y, \qquad v = x + y.$$

For a finite sequence $w = (w_1, \dots, w_{n+1})$ we denote by

$$\mathrm{Ins}(w_1, \dots, w_n;\ w_{n+1}) = (x_n, y_n)$$

the coordinates of the last box which was inserted to the Young diagram by the RSK algorithm applied to the sequence w. In other words, it is the box containing the biggest number in the recording tableau $Q(w)$. Above, (x_n, y_n) refer to the Cartesian coordinates of this box in the French convention, i.e., x_n is the number of the column and y_n is the number of the row. By

$$\text{u-}\mathrm{Ins}(w_1, \dots, w_n;\ w_{n+1}) = x_n - y_n$$

we denote the u-coordinate of the aforementioned box.

For later use, given a tableau T and a real number z we denote by $\mathrm{Ins}(T; z)$ the coordinates of the new box which was created by the Schensted row insertion $T \leftarrow z$; in other words it is the unique box of the skew diagram

$$\mathrm{shape}\left(T \leftarrow z\right) \setminus \mathrm{shape}\, T;$$

the quantity $\text{u-}\mathrm{Ins}(T; z)$ is defined in an analogous way as the u-coordinate of $\mathrm{Ins}(T; z)$.

In the current paper we concentrate on the aforementioned fundamental case when $w_1, w_2, \dots$ is a sequence of independent, identically distributed random variables with the uniform distribution $U(0, 1)$ on the unit interval $[0, 1]$. Romik and Śniady [RŚ15] noticed that the construction of the inverse map to Q_∞ (and, in consequence, a positive answer to Problem 3.1.1) requires a solution to the following somewhat vague question about the usual (finite) version of RSK applied to such a random input.

Problem 3.1.2 (The main problem)**.** *Let $w = (w_1, w_2, \dots)$ be a sequence of independent, identically distributed random variables with the uniform distribution $U(0, 1)$ on the unit interval.*

What is the relationship between

- *the value of the new entry w_{n+1}, and*
- *the position of the corresponding newly created box*

$$(x_n, y_n) = \mathrm{Ins}(w_1, \dots, w_n;\ w_{n+1})? \tag{3.1.4}$$

We are interested in this probabilistic question in the limit as $n \to \infty$.

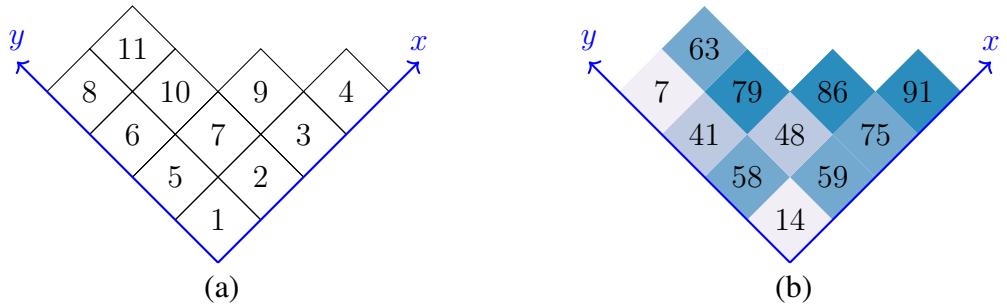


Figure 3.3: (a) The recording tableau $Q_{xy} = Q(w)$ (drawn in the Russian convention) which corresponds to the sequence $w = (14, 59, 75, 91, 58, 41, 48, 7, 86, 79, 63)$ which was selected from the interval $J = [0, 100]$. (b) The responsibility matrix $\left(w_{Q_{xy}}\right)$ obtained by replacing each entry of the recording tableau by the corresponding entry of the sequence w. The following layer tinting was used: the four colors of the background correspond to the values of the responsibility matrix in the four quarters of the interval J, namely, $[0, 25]$ (almost white), $(25, 50]$ (beige), $(50, 75]$ (blue), and $(75, 100]$ (dark blue).

This problem can be visualized by replacing the number in any box $\square = (x, y)$ of the recording tableau $Q_{xy} = Q(w)$ by the entry of the sequence $w = (w_1, \ldots, w_n)$ which was responsible for the creation of $\square$. The resulting matrix $\left(w_{Q_{xy}}\right)$ will be called *the responsibility matrix*, see Figure 3.3b for an example. In order to improve legibility and avoid writing real numbers, the entries of the sequence w in the example in Figure 3.3 are integers sampled from the interval $J = [0, 100]$.

Following the ideas of Pittel and Romik [PR07, Section 1.1], the responsibility matrix can be depicted geometrically as a three-dimensional stack of cuboids over the plane $\mathbb{R}^2 \times \{0\}$, where $w_{Q_{xy}}$ is the height of the cuboid which has the unit square $[x-1, x] \times [y-1, y] \times \{0\}$ as the base. Alternatively, the function $(x, y) \mapsto w_{Q_{xy}}$ can be thought of as the graph of the (non-continuous) surface of the upper envelope of this stack. By rescaling the unit squares on the plane $\mathbb{R}^2 \times \{0\}$ to be squares of side length $\frac{1}{\sqrt{n}}$ (recall that n is the length of the sequence w), the total base area of the cuboids becomes equal to 1, see Figure 3.4 for an example. In Figures 3.3b and 3.4 the elevation (i.e., the values of the responsibility matrix) was indicated by *layer tinting*.

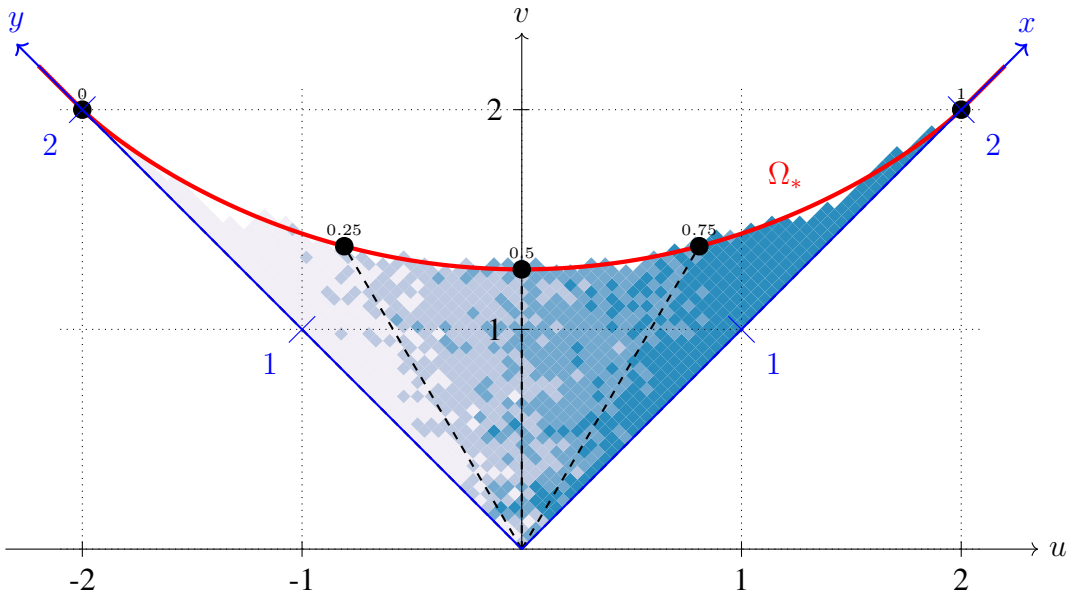


Figure 3.4: An analogue of Figure 3.3b for a sequence $w = (w_1, \ldots, w_{1000})$ of $n = 1000$ independent, identically distributed random variables with the uniform distribution $U(0, 1)$ on the unit interval $I = [0, 1]$. The layer tinting indicates the values of the responsibility matrix $\left(w_{Q_{xy}}\right)$: the four colors correspond to the values in the four intervals $\left[0, 1/4\right]$ (almost white), $\left(1/4, 1/2\right]$, $\left(1/2, 3/4\right]$, $\left(3/4, 1\right]$ (dark blue). The red solid line is the Logan–Shepp–Vershik–Kerov limit curve Ω_*. The five black dots indicate its natural parametrization; they divide the area between the curve and the Oxy axes into four curvilinear triangles of equal areas.

Monte Carlo simulations such as the one from Figure 3.4 suggest that the value of the new entry w_{n+1} in Problem 3.1.2 determines the ray (a halfline starting in the origin of the coordinate system) on which the coordinates $\mathrm{Ins}(w_1, \ldots, w_n;\ w_{n+1})$ will approximately appear (with high probability, asymptotically, as $n \to \infty$). Small values of the new entry (i.e., $w_{n+1} \approx 0$) seem to correspond to the rays closer to the Oy axis (the almost white area in Figure 3.4) while large values ($w_{n+1} \approx 1$) seem to correspond to the rays closer to the Ox axis (the dark blue area in Figure 3.4).

3.1.4 The limit shape and its parametrization

The first step towards understanding Problem 3.1.2 is the simple observation that the newly created box $\mathrm{Ins}(w_1, \ldots, w_n;\ w_{n+1})$ must be located in one of the concave corners of the Young diagram $\lambda^{(n)} = \mathrm{RSK}(w_1, \ldots, w_n)$. For this reason, for our purposes it is beneficial to understand the asymptotic behavior of the random Young diagram $\lambda^{(n)}$. Fortunately, the probability distribution of such a RSK shape is simply the Plancherel measure on Young diagrams with n boxes for which the limit shape is well-known; we will review it in the following. For a pedagogical introduction to this topic we refer to [Rom15, Chapter 1].

Scaling of Young diagrams

The boundary of a Young diagram λ is called its *profile*, see Figure 3.1a. In the Russian coordinate system the profile can be seen as the plot of the function $\omega_\lambda \colon \mathbb{R} \to \mathbb{R}_+$, see Figure 3.1b.

If $c > 0$ is a positive number, the output $c\lambda \subset \mathbb{R}_+^2$ of a homogeneous dilation with scale c applied to the Young diagram λ might no longer be a Young diagram. Nevertheless, its profile is still well defined as

$$\omega_{c\lambda}(u) = c\,\omega_\lambda\left(\frac{u}{c}\right).$$

The Logan–Shepp–Vershik–Kerov limit shape

Independently Logan and Shepp [LS77] as well as Vershik and Kerov [VK77] proved that a law of large numbers for the shapes holds true: as the number of boxes $n \to \infty$ tends to infinity, the scaled down profile of the random Young diagram converges in probability to some explicit limit curve, see Figure 3.4 for an illustration.

Define the function $\Omega_* : \mathbb{R} \to [0, \infty)$ by

$$\Omega_*(u) = \begin{cases} \frac{2}{\pi}\left[u \sin^{-1}\left(\frac{u}{2}\right) + \sqrt{4-u^2}\right] & \text{if } -2 \le u \le 2, \\ |u| & \text{otherwise.} \end{cases} \tag{3.1.5}$$

Theorem 3.1.3 (The limit shape of Plancherel-distributed Young diagrams [LS77; VK77])**.** *Let $\lambda^{(n)}$ be a random Young diagram distributed according to the Plancherel measure on Young diagrams with n boxes. Then we have a convergence in the supremum norm in probability*

$$\sup_{u \in \mathbb{R}} \left| \omega_{\frac{1}{\sqrt{n}}\lambda^{(n)}}(u) - \Omega_*(u) \right| \xrightarrow[n\to\infty]{P} 0.$$

In fact, the rate of convergence is quite fast and there are some quite precise results about the magnitude of the local fluctuations [BS07].

Location of a box with a specified entry

In the context of Problem 3.1.2, suppose for a moment that we *do not* know the value of the new entry w_{n+1}. Then the complete answer to Problem 3.1.2 would be the probability distribution of the new box (3.1.4). This probability distribution coincides with the distribution of the box containing the entry $n+1$ in a large Plancherel-distributed random standard Young tableau. An analogous problem was studied by Pittel and Romik [PR07, Section 1.2] for a different class of random tableaux.

Not very surprisingly, the probability distribution of this vector (after rescaling, and in the Russian coordinates)

$$(u_n, v_n) = \left(\frac{x_n - y_n}{\sqrt{n}}, \frac{x_n + y_n}{\sqrt{n}} \right)$$

converges, as $n \to \infty$, to a certain probability measure μ_* which is supported on the limit curve Ω_*. In order to specify this measure uniquely it is enough to find the limit distribution for the u-coordinates, i.e., for the random variables (u_n). It turns out that the sequence (u_n) converges in distribution to the *semicircle measure* on the interval $[-2, 2]$, see [RŚ15, Theorem 3.2]. The density of this measure is given by

$$f_{\mathrm{SC}}(u) = \frac{1}{2\pi}\sqrt{4 - z^2} \qquad \text{for } u \in [-2, 2].$$

We denote by $F_{\mathrm{SC}}\colon [-2,2] \to [0,1]$ the cumulative distribution function of this semicircle law, given by

$$F_{\mathrm{SC}}(u) = \frac{1}{2\pi} \int_{-2}^{u} \sqrt{4 - z^2}\, \mathrm{d}z \qquad \text{for } u \in [-2,2].$$

The natural parametrization of the limit curve

The following two *RSK-trigonometric functions*

$$\begin{aligned} \mathrm{RSKcos}\, z &= F_{\mathrm{SC}}^{-1}(z), \\ \mathrm{RSKsin}\, z &= \Omega_* \left(F_{\mathrm{SC}}^{-1}(z) \right) \end{aligned}$$

are defined for $z \in [0,1]$. The map

$$[0,1] \ni z \mapsto (\mathrm{RSKcos}\, z, \mathrm{RSKsin}\, z) \in \mathbb{R}^2 \tag{3.1.6}$$

is a a two-dimensional analogue of the *quantile function* (the inverse of the cumulative distribution function) in the context of the limit measure μ_*. This map provides a convenient parametrization of the limit curve Ω_*, see the five black dots in Figure 3.4. It is a natural analogue of the parametrization of the unit circle by the angle

$$z \mapsto (\cos z, \sin z)$$

as well as the parametrization of the hyperbole by the hyperbolic angle

$$z \mapsto (\cosh z, \sinh z)$$

since in all three cases the area of the curvilinear triangle between the the ray, the curve, and the y-axis (respectively, x-axis) is proportional to the value of the parameter z.

3.1.5 The main problem: asymptotic determinism of RSK insertion and its fluctuations

The most convenient way to state an answer to Problem 3.1.2 is by conditioning, i.e., assuming that the random variable w_{n+1} takes some fixed value $z \in [0,1]$. The following result of Romik and Śniady provides a first order asymptotics.

Theorem 3.1.4 ([RŚ15, Theorem 5.1]). *Let $z \in [0,1]$ be fixed; we denote by*

$$(u_0, v_0) = (\mathrm{RSKcos}\, z, \mathrm{RSKsin}\, z)$$

the corresponding point on the limit curve Ω_* via (3.1.6)*.*

Let $w_1, w_2, \ldots$ be a sequence of independent, identically distributed random variables with the uniform distribution $U(0,1)$ on the unit interval; we denote

$$(x_n, y_n) = \mathrm{Ins}(w_1, \ldots, w_n;\ z).$$

Then we have the following convergence in probability:

$$\left(\frac{x_n - y_n}{\sqrt{n}}, \frac{x_n + y_n}{\sqrt{n}} \right) \xrightarrow[n\to\infty]{P} (u_0, v_0). \tag{3.1.7}$$

This relationship between the value of the new entry being inserted and the location of the corresponding new box was visualized in Figure 3.4 by the layer tinting. Theorem 3.1.4 was illustrated by a Monte Carlo simulation in Figure 3.5.

Śniady stated the following conjecture which would give the ultimate answer to Problem 3.1.2.

Conjecture 3.1.5 ([Śni20]). *We keep the notation from Theorem 3.1.4.*

Then the scaled sequence of random points

$$\sqrt[4]{n} \left[\left(\frac{x_n - y_n}{\sqrt{n}}, \frac{x_n + y_n}{\sqrt{n}} \right) - (u_0, v_0) \right] = \left(\frac{x_n - y_n}{\sqrt[4]{n}} - \sqrt[4]{n}\, u_0,\ \ \frac{x_n + y_n}{\sqrt[4]{n}} - \sqrt[4]{n}\, v_0 \right) \in \mathbb{R}^2 \tag{3.1.8}$$

converges in distribution to a centered, degenerate Gaussian distribution which is supported on the line

$$\left\{ (u, v) : v = \Omega_*'(u_0)\, u \right\}$$

which is parallel to the tangent line to the limit curve Ω_* in u_0.

This Gaussian distribution is uniquely determined by the limit measure for the u-coordinates, which is as follows:

$$\sqrt[4]{n} \left[\frac{x_n - y_n}{\sqrt{n}} - u_0 \right] \xrightarrow[n\to\infty]{d} N\left(0, \sigma_z^2\right).$$

It converges, as $n \to \infty$, in distribution to the centered normal distribution with the variance

$$\sigma_z^2 = \frac{\pi}{3} \sqrt{4 - u_0^2}. \tag{3.1.9}$$

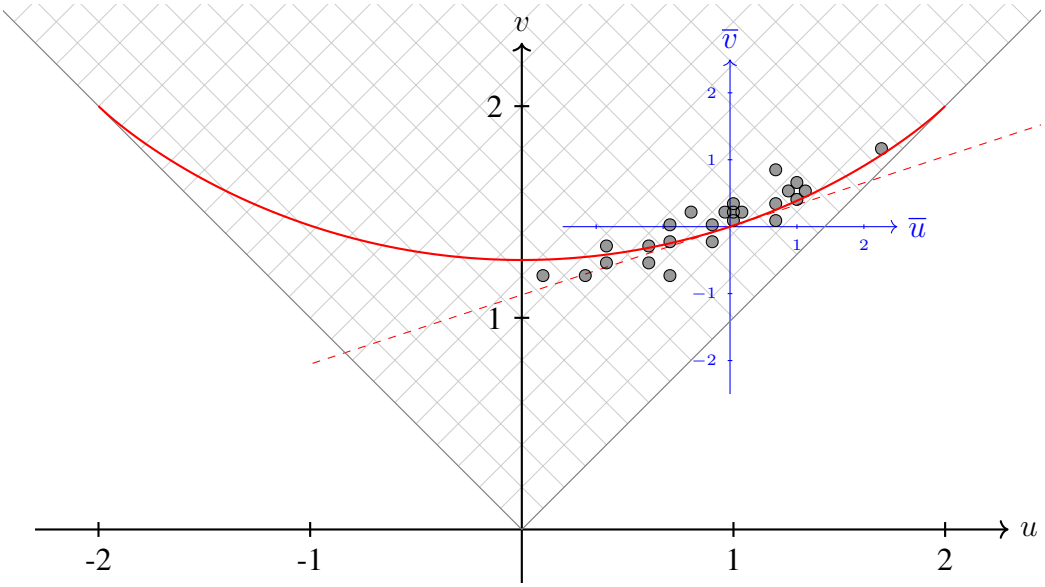


Figure 3.5: Grey circles indicate simulated (rescaled) positions of the new box $\mathrm{Ins}(w_1, \ldots, w_n; z)$ for the initial sequence $w_1, \ldots, w_n$ of i.i.d. $U(0,1)$ random variables of length $n = 100$ and the new entry $z = 0.8$. The uv coordinates of the circles are the quantities which appear in the law of large numbers (Theorem 3.1.4). The grid indicates the actual size of the boxes. The solid red curve is the Logan–Shepp–Vershik–Kerov limit shape; the red dashed line is tangent to this curve at the point which in the natural parametrization corresponds to z. The coordinates of the circles in the blue $(\overline{u}, \overline{v})$ coordinate system correspond to the quantities in Conjecture 3.1.5.

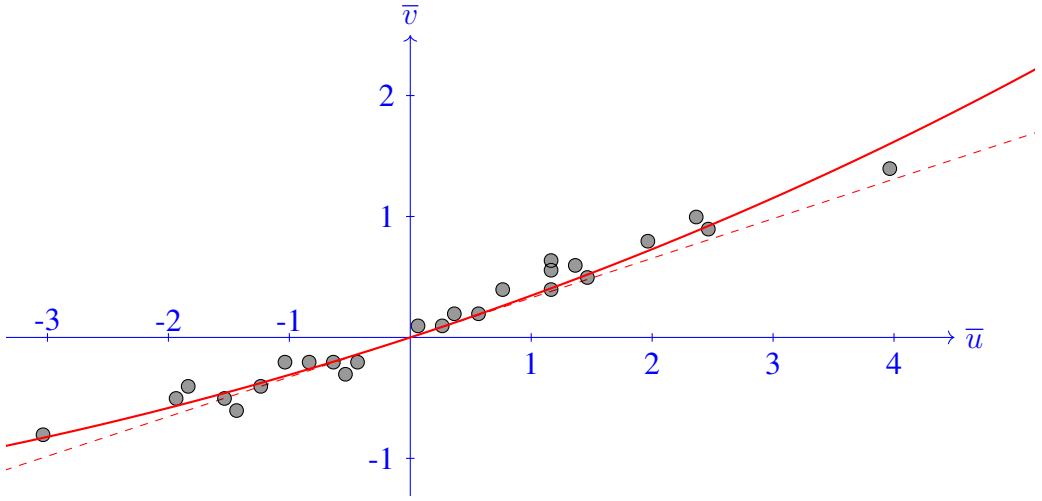


Figure 3.6: An analogue of Figure 3.5 for the length of the sequence $n = 10^4$ and $z = 0.8$. The picture was zoomed to focus on the $(\overline{u}, \overline{v})$ coordinate system. In order to improve visibility the grid with the actual size of the boxes was not shown.

This conjecture is visualized by Monte Carlo simulations in Figures 3.5 and 3.6. The blue coordinate system $(\overline{u}, \overline{v})$

$$(\overline{u}, \overline{v}) = \sqrt[4]{n} \left[\left(\frac{u}{\sqrt{n}}, \frac{v}{\sqrt{n}} \right) - (u_0, v_0) \right]$$

corresponds to the quantities which appear in the n-th random point (3.1.8).

A special case of Conjecture 3.1.5 for $z = \frac{1}{2}$ was conjectured by Wojtyniak [Woj19] (with a slightly different form of the variance (3.1.9)) based on extensive Monte Carlo simulations.

Conjecture 3.1.5 seems to be a special case of a more general result (Conjecture 3.2.2) which also provides a conceptual interpretation to the somewhat mysterious formula for the variance (3.1.9).

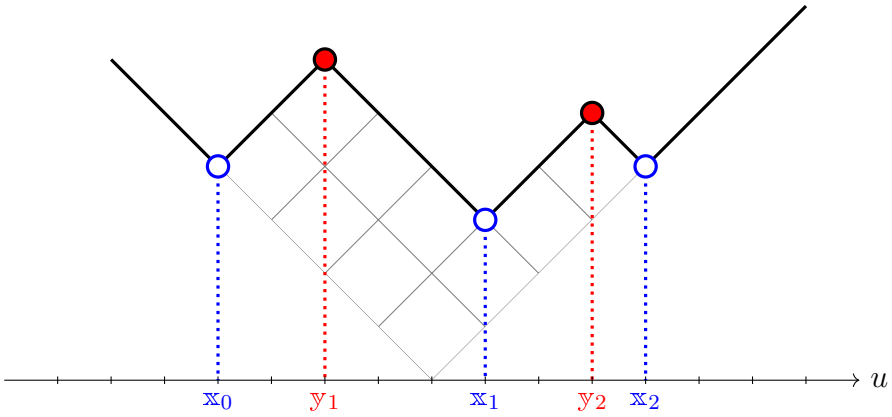


Figure 3.7: Concave corners (empty) and convex corners (filled) of a Young diagram $(4, 2, 2, 2)$ and their u-coordinates.

3.2 The general form of RSK insertion fluctuations

3.2.1 Plancherel growth process

Let $w_1, w_2, \dots$ be a sequence of i.i.d. random variables with the uniform distribution $U(0, 1)$. Let

$$\lambda^{(n)} := \mathrm{RSK}(w_1, \dots, w_n);$$

we say that the random sequence of Young diagrams

$$\emptyset = \lambda^{(0)} \nearrow \lambda^{(1)} \nearrow \cdots \tag{3.2.1}$$

is the *Plancherel growth process* [Rom15, Chapter 1.19]. It turns out that (3.2.1) is a Markov chain; below we will describe its transition probabilities.

Note that the above construction is a specific case of a more general setup which we already discussed in Section 3.1.2.

3.2.2 Transition measure of a Young diagram

For a given Young diagram λ with n boxes we denote by $\mathbb{x}_0 < \cdots < \mathbb{x}_{\mathbb{L}}$ the u-coordinates of its concave corners and by $\mathbb{y}_1 < \cdots < \mathbb{y}_{\mathbb{L}}$ the u-coordinates of its *convex corners*, see Figure 3.7. The *Cauchy transform* of λ is defined as the

rational function [Ker93]

$$\mathbf{G}_\lambda(z) = \frac{(z-\mathbb{y}_1)\cdots(z-\mathbb{y}_\mathbb{L})}{(z-\mathbb{x}_0)\cdots(z-\mathbb{x}_\mathbb{L})}. \tag{3.2.2}$$

The Cauchy transform can be written as a sum of simple fractions:

$$\mathbf{G}_\lambda(z) = \sum_{0\le i\le \mathbb{L}} \frac{p_i}{z-\mathbb{x}_i}$$

with the coefficients $p_0, \dots, p_\mathbb{L} > 0$ such that $p_0 + \cdots + p_\mathbb{L} = 1$. We define *the transition measure* of λ as the discrete measure

$$\mu_\lambda = p_0\delta_{\mathbb{x}_0} + \cdots + p_\mathbb{L}\delta_{\mathbb{x}_\mathbb{L}};$$

in this way

$$\mathbf{G}_\lambda(z) = \int_{\mathbb{R}} \frac{1}{z-x}\,\mathrm{d}\mu_\lambda(x)$$

is indeed the Cauchy transform of μ_λ.

Kerov [Ker93] proved that the transition probabilities of the Markov chain (3.2.1) are encoded by the transition measure. More specifically, the conditional probability that the new box will have the u-coordinate equal to $\mathbb{x}_i$ is given by

$$\mathbb{P}\left[u\left(\lambda^{(n+1)} \setminus \lambda^{(n)}\right) = \mathbb{x}_i \;\middle|\; \lambda^{(n)} = \lambda\right] = p_i = \operatorname{Res}_{\mathbb{x}_i} \mathbf{G}_\lambda, \tag{3.2.3}$$

i.e., by the corresponding atom of the transition measure as well as by the residue of the Cauchy transform.

We denote by

$$K_\lambda(z) = \mu_\lambda\big((-\infty, z]\big)$$

the cumulative distribution function of μ_λ.

3.2.3 Random Poissonized tableau of a given shape

By a *Poissonized tableau* [GR19] we mean any tableau which has the real numbers from the unit interval $[0, 1]$ as the entries. The set of Poissonized tableaux with shape λ will be denoted by $\mathcal{T}^\lambda$. There are two natural ways to equip $\mathcal{T}^\lambda$ with a probability measure, we will discuss them below.

Firstly, let us number the boxes of λ in an arbitrary way; it follows that each element of $\mathcal{T}^\lambda$ can be identified with an element of the unit cube $[0, 1]^n$, where n

is the number of boxes of λ. The requirement that the rows and the columns are increasing corresponds to a collection of inequalities between the coordinates of the points in the cube; it follows that $\mathcal{T}^\lambda$ can be identified with a convex polytope contained in $[0,1]^n$. This polytope has a strictly positive volume therefore it is possible to equip it with *the uniform probability measure*.

On the other hand, the canonical way of generating a random Poissonized tableau T with a prescribed number of boxes n is to consider the insertion tableau $P(w)$, where $w = (w_1, \dots, w_n)$ is a sequence of i.i.d. random numbers with the $U(0,1)$ distribution. If we condition the random tableau T so that its shape is equal to λ, it becomes a natural candidate for the notion of *a uniformly random Poissonized tableau of shape* λ.

The following lemma shows that the above two approaches give rise to the same probability distribution.

Lemma 3.2.1. *Let $w = (w_1, \dots, w_n)$ be a sequence of i.i.d. random variables with the $U(0,1)$ distribution and let λ be a Young diagram with n boxes.*

The conditional *probability distribution of the insertion tableau $P(w)$* under the condition that $\mathrm{RSK}(w) = \lambda$ *coincides with the uniform probability distribution on $\mathcal{T}^\lambda$.*

The proof is postponed to Section 3.4.1.

3.2.4 The interaction energy

If μ is a probability measure on the real line and $u_0 \in \mathbb{R}$ we define

$$\mathcal{E}_\mu(u_0) = \iint\limits_{\{(z_1,z_2):\ z_1<u_0<z_2\}} \frac{1}{z_2 - z_1}\,\mathrm{d}\mu(z_1)\,\mathrm{d}\mu(z_2) \tag{3.2.4}$$

whenever this double integral is finite. This quantity will play an important role in the statement of the conjectures of Śniady.

Let us interpret μ as the distribution of the electrostatic charge along a one-dimensional rod. If we split this rod at the point u_0, the two parts will act on one another with the electrostatic force. The integral (3.2.4) can be interpreted as the *interaction energy* between these two parts.

We now consider an alternative universe in which the space is two-dimensional so that the electrostatic force decays as the inverse of the distance between the charges. It follows that the double integral $\mathcal{E}_\mu(u_0)$ is equal to the *electrostatic force* between the two parts of the rod.

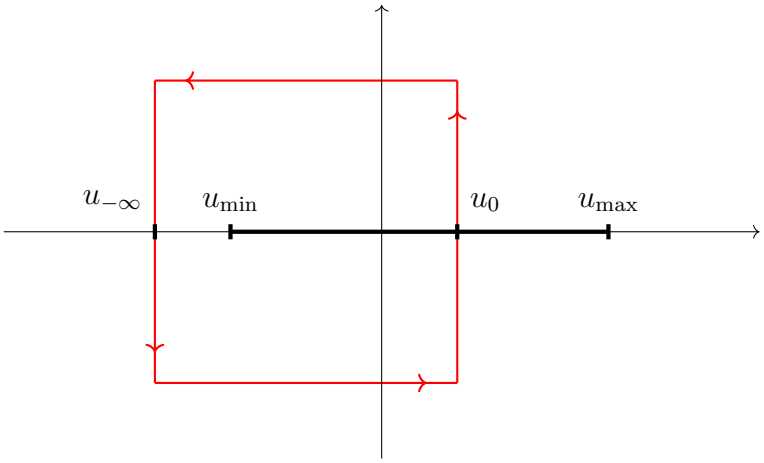


Figure 3.8: The contour C on the complex plane.

If we separate the two parts of the rod by an additional distance l, the aforementioned electrostatic force is the derivative of the total energy of the system with respect to the variable l. It is worth pointing out that in our context *the total energy* should be understood as the logarithmic energy which is ubiquitous in random matrix theory, Voiculescu's free entropy [Voi94], as well as in asymptotic representation theory. For this reason we suspect that the interaction energy $\mathcal{E}_\mu(u_0)$ can be found also in the context of random matrix theory.

Suppose that the support of the measure μ is contained in some interval $[u_{\min}, u_{\max}]$. We choose an arbitrary real number $u_{-\infty}$ such that $u_{-\infty} < u_{\min}$ and consider a contour C on the complex plane shown in Figure 3.8. This contour was chosen in such a way that it crosses the real line in two points, namely, $u_{-\infty}$ and u_0.

In the special case when the support of μ is a finite set and u_0 does not belong to the support of μ we may apply Cauchy's residue theorem and the interaction energy is equal to the contour integral

$$\mathcal{E}_\mu(u_0) = -\frac{1}{4\pi i} \oint_C \big[\mathbf{G}_\mu(z)\big]^2 \,\mathrm{d}z, \tag{3.2.5}$$

where

$$\mathbf{G}_\mu(z) = \int_{\mathbb{R}} \frac{1}{z-x} \,\mathrm{d}\mu(x)$$

is the Cauchy transform of μ. The formula (3.2.5) remains true for general measures μ as long as the Cauchy transform is sufficiently regular in the neighborhood of u_0, however the justification of its validity is more technically involved. In practical applications the formula (3.2.5) is more convenient than (3.2.4).

3.2.5 General form of the fluctuations

The following conjecture was stated by Śniady [Śni20].

Conjecture 3.2.2. *Let μ be a probability measure on the real line, and let $z \in [0,1]$ be a fixed number. Denote by $u_0 = F_\mu^{-1}(z)$ the corresponding value of the quantile function; we assume that the derivative $f = F_\mu'(u_0) > 0$ of the cumulative distribution function exists and is positive (it can be interpreted as the density of the measure μ in u_0).*

For each integer $n \geq 1$ let $\lambda^{(n)}$ be a random Young diagram and let a_n be a random variable. We assume that the following conditions hold true:

(a)

$$K_{\lambda^{(n)}}\left(\sqrt{n}\, u\right) \xrightarrow[n\to\infty]{P} F_\mu(u)$$

holds true for each $u \in \mathbb{R}$ in which F_μ is continuous;

(b) for each $c \in \mathbb{R}$

$$\sqrt[4]{n}\left[K_{\lambda^{(n)}}\left(\sqrt{n}\, a_n + \sqrt[4]{n}\, c\right) - z\right] \xrightarrow[n\to\infty]{P} f\, c.$$

Let $T^{(n)}$ be a uniformly random Poissonized tableau of shape $\lambda^{(n)}$ and let

$$(x_n, y_n) = \mathrm{Ins}\left(T^{(n)}; z\right).$$

Then

$$\sqrt[4]{n}\left[\frac{x_n - y_n}{\sqrt{n}} - a_n\right] \xrightarrow[n\to\infty]{dist} N\left(0, \frac{\mathcal{E}_\mu(u_0)}{f^2}\right).$$

The requirement (a) says that the (dilated by factor $\frac{1}{\sqrt{n}}$) Kerov's transition measure of $\lambda^{(n)}$ converges in the weak topology of probability measures to μ, in probability. In the simplest case when $a_n = u_0$ are all deterministic, the requirement (b) says that in a direct neighborhood of the number u_0 this convergence holds true very fast, with a specific rate.

Śniady [Śni20] also conjectured that the assumptions of Conjecture 3.2.2 are fulfilled if $\lambda^{(n)}$ is the Plancherel-distributed random Young diagram; in this way Conjecture 3.1.5 would be a consequence of Conjecture 3.2.2.

3.2.6 The content of the paper

The main goal of the current paper is to provide some computational tools (Theorem 3.3.2) which seem to be necessary to complete the proof of Conjecture 3.2.2. The proof of Conjecture 3.2.2 is out of scope of the current paper and we intend to address it in the future research.

3.3 The cumulative function of a tableau and its cumulants

3.3.1 The cumulative function of a tableau

Let T be a Poissonized tableau. For $u_0 \in \mathbb{R}$ we define

$$F_T(u_0) = \inf\big\{z \in [0,1] : \text{u-Ins}(T; z) \geq u_0\big\};$$

we recall that the quantity $\text{u-Ins}(T; z)$ was defined in Section 3.1.3 as the u-coordinate of the box $\text{Ins}(T; z)$. In the case when the infimum runs over the empty set, we declare that $F_T(u_0) = 1$, see Figure 3.9 for an example.

The function $F_T \colon \mathbb{R} \to [0,1]$ will be called *the cumulative function of* T. It has analogous properties as any cumulative distribution function of a probability measure with a finite support. It can be regarded as an inverse of the insertion function $z \mapsto \text{u-Ins}(T; z)$; more precisely the relationship between F_T and $\text{u-Ins}(T; \cdot)$ is the same as between the cumulative distribution function of a measure and the quantile function.

This relationship is the key motivation for studying the cumulative function F_T. Indeed, since Conjecture 3.1.5 and Conjecture 3.2.2 can be seen as limit statements about the probability distribution of the (random) function

$$z \mapsto \text{u-Ins}(T; z)$$

for some special choices for the random tableau T, a natural direction for proving these conjectures is to prove suitable limit results for its inverse function F_T.

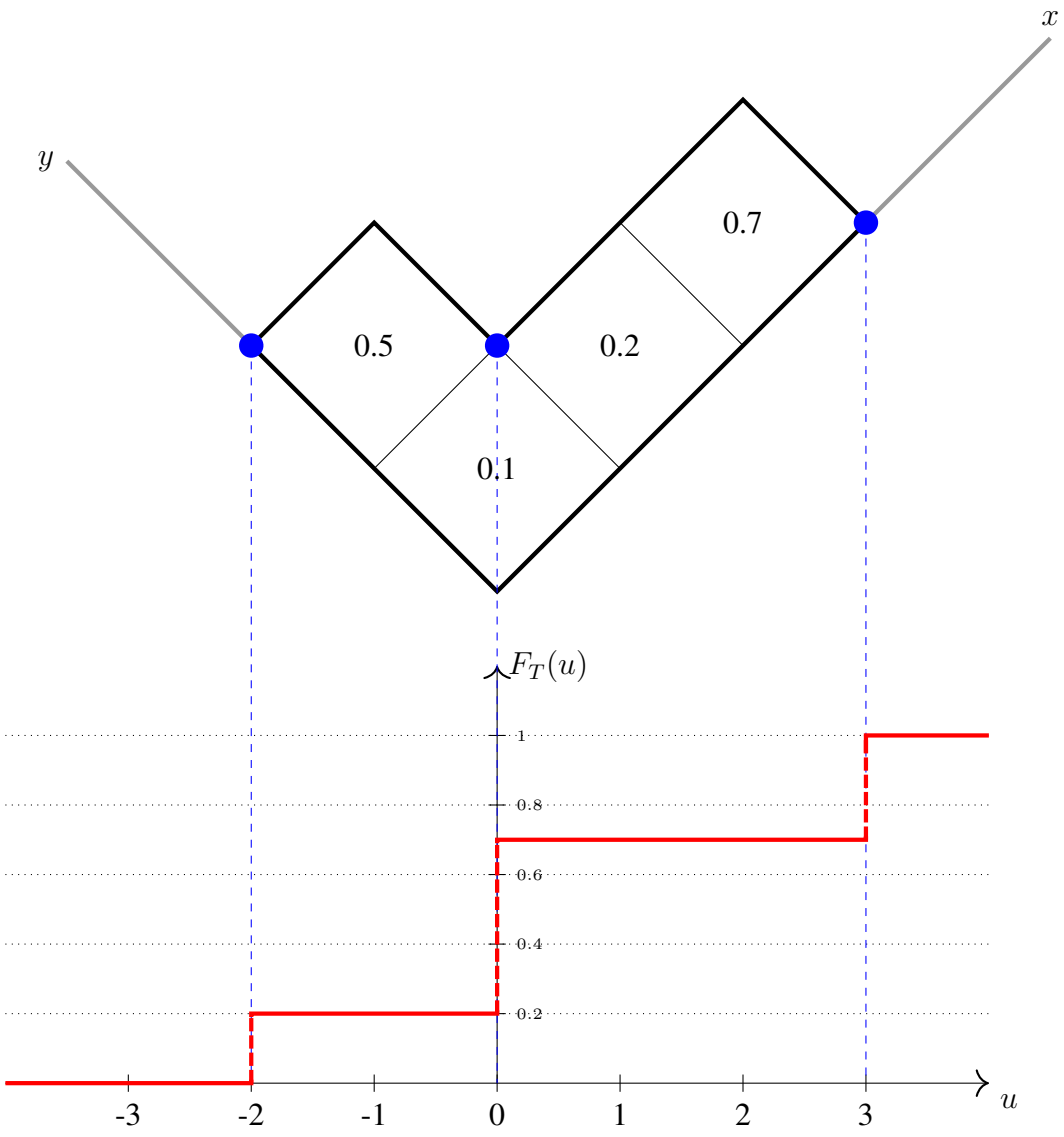


Figure 3.9: A Poissonized tableau T shown in the Russian coordinates. The red line depicts its cumulative function F_T.

3.3.2 Asymptotics of the cumulative function of random tableaux

Let us choose somehow (randomly or deterministically) a Young diagram λ, and let T be a random Poissonized tableau of shape λ. We will study the asymptotics of the random function $u_0 \mapsto F_T(u_0)$, in the limit as the number of boxes of λ tends to infinity.

A concrete version of this general problem is to take λ to be a Plancherel-distributed random Young diagram with n boxes. The resulting random tableau T can be then alternatively generated as the insertion tableau $T = P(w_1, \dots, w_n)$ corresponding to a sequence of i.i.d. $U(0,1)$ random variables, see Lemma 3.2.1. The simulations (Figures 3.10 and 3.11) indicate that in the asymptotic regime the cumulative distribution function F_T and the cumulative distribution function of the transition measure μ_λ are typically very close to each other.

The main result of the current paper (Theorem 3.3.2) is an answer to the following more modest problem related to the *pointwise* behavior of the cumulative function.

Problem 3.3.1. *Let λ be Young diagram and let T be a uniformly random normalized tableau with shape λ. For a given $u_0 \in \mathbb{R}$ find the probability distribution of the random variable $F_T(u_0)$.*

3.3.3 Cumulants and moments

Let X be a random variable with the sequence of moments $(m_k)_{k=1}^\infty$, where $m_k = \mathbb{E}X^k$. The formal power series

$$\mathbb{E}[e^{tX}] = \sum_{k=0}^{\infty} \frac{m_k}{k!} t^k$$

is its exponential moment generating function or *formal Fourier–Laplace transform*. The coefficients $(\kappa_k)_{k=1}^\infty$ of its formal logarithm

$$\log \mathbb{E}[e^{tX}] = \sum_{k=1}^{\infty} \kappa_k \frac{t^k}{k!}$$

are called the *cumulants* [LH02] of the random variable X. The first cumulant is the expected value and the second cumulant is the variance:

$$\kappa_1 = \mathbb{E}X,$$
$$\kappa_2 = \operatorname{Var} X.$$

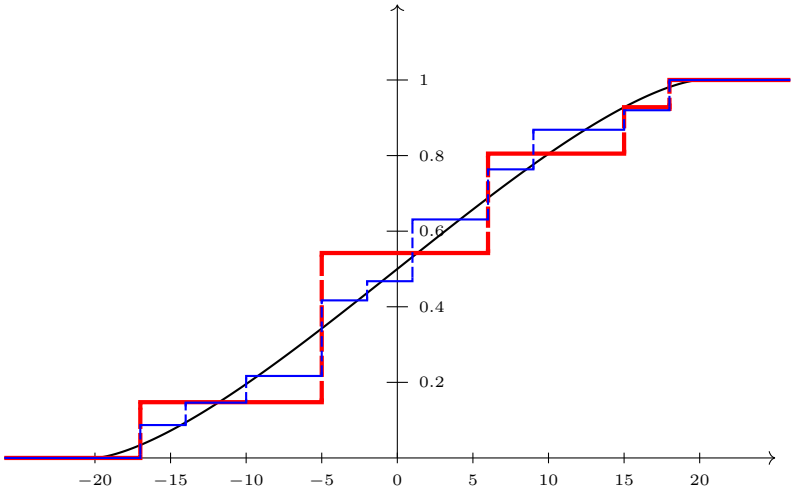


Figure 3.10: The insertion tableau $T = P(w_1, \ldots, w_n)$ with $n = 100$ boxes was sampled by applying the RSK algorithm to a sequence of i.i.d. $U(0, 1)$ random variables. The thin blue line shows the cumulative distribution function K_λ of the transition measure of the shape λ of T. The thick red line shows the cumulative function F_T. The smooth black line is the cumulative distribution function of the semicircle distribution supported on the interval $\left[-2\sqrt{n}, 2\sqrt{n}\right]$.

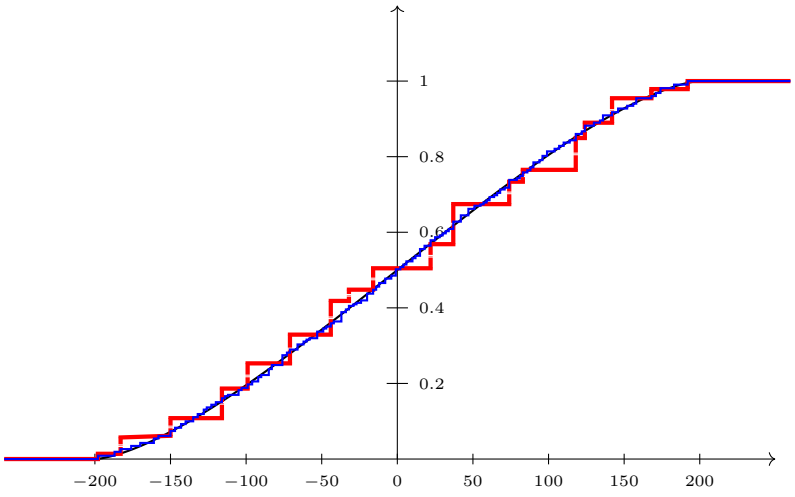


Figure 3.11: An analogue of Figure 3.10 for $n = 10^4$.

Cumulants are related with the moments via *the moment-cumulant formula*

$$m_k = \sum_{\pi \in \Pi_k} \prod_{b \in \pi} \kappa_{|b|}, \tag{3.3.1}$$

where the sum runs over all set-partitions π of the set $\{1, \ldots, k\}$, and the product runs over all blocks of the partition π. For example, for $k = 3$ there are 5 set-partitions of the set $\{1, 2, 3\}$, namely,

$$\big\{\{1\}, \{2\}, \{3\}\big\}, \big\{\{1, 2\}, \{3\}\big\}, \big\{\{1, 3\}, \{2\}\big\}, \big\{\{2, 3\}, \{1\}\big\}, \big\{\{1, 2, 3\}\big\}.$$

Therefore

$$m_3 = \kappa_1^3 + \kappa_2\kappa_1 + \kappa_2\kappa_1 + \kappa_2\kappa_1 + \kappa_3 = \kappa_1^3 + 3\kappa_2\kappa_1 + \kappa_3.$$

3.3.4 Notation

By *a directed graph* we mean any graph in which every edge has been directed. The edge outgoing from the vertex a and incoming to the vertex b will be denoted by (a, b). The directed graphs we consider do not have multiple edges, but may have *loops*, i.e., edges of the form (a, a). We will always assume that if $a \neq b$ and (a, b) is an edge of a graph then the opposite edge (b, a) is *not* an edge.

The graphs we consider will have their vertices colored black, red, or white. For the convenience of the readers of the non-colored printed version of this paper, the red vertices will be drawn with an additional ornament as crossed-out circles, see Figure 3.18. For a given graph H, we denote the set of its vertices by V_H, the set of its black vertices by B_H, the set of its red vertices by R_H, the set of its white vertices by W_H, and the set of its edges by E_H.

We say that a graph is a *weighted graph* if each of its edges is assigned a number called a *weight*. For a given edge e, we denote its weight by $w(e) \in \mathbb{R}$.

3.3.5 Decorations

Let $\mathbb{X} \subset \mathbb{R}$ be a fixed discrete set. Also, let $u_0 \in \mathbb{R}$ be a fixed real number. The elements of the interval $(-\infty, u_0]$ will be called *small* while the elements of the interval (u_0, ∞) will be called *big*.

For a given graph H we say that a function $\mathbf{x}\colon V_H \to \mathbb{X}$ is *a u_0-decoration of the graph H* if the following two conditions hold true:

$$\mathbf{x}(b) \text{ is small for each } b \in B_H,$$
$$\mathbf{x}(w) \text{ is big for each } w \in W_H.$$

Figure 3.12: (a) The unique non-crossing alternating tree with 1 vertex. (b) The unique non-crossing alternating tree with 2 vertices.

For simplicity we denote $x_v = \mathbf{x}(v)$ for $v \in V_H$. The set of all u_0-decorations of the graph H will be denoted by $D_H(u_0)$. When the value of u_0 is clear from the context, we will simply speak about *decorations* and write $D_H = D_H(u_0)$.

3.3.6 Non-crossing alternating trees

Let $k \geq 1$ be a natural number. We say that a tree with k vertices numbered $1, \ldots, k$ is *a non-crossing alternating tree* if the following conditions hold true:

(a) each vertex is colored either black or white;

(b) if an edge connects the vertices b and w for $b < w$, then the vertex b is black and the vertex w is white;

(c) there do not exist four vertices $v_1 < v_2 < v_3 < v_4$ such that v_1 is connected with v_3, and v_2 is connected with v_4,

see [Sta99, Exercise 6.19(p)] and solution to this exercise, as well as [GGP97, Section 6]. The condition (c) has a natural graphical interpretation: after drawing the vertices on the real line and the edges as arcs above the real line, we require that the edges do not cross.

In the exceptional case $k = 1$ we declare that there is only one non-crossing alternating tree with 1 vertex: it consists of a single black vertex (see Figure 3.12a). With this convention, the condition (b) implies that the coloring of the vertices can be uniquely recovered purely from the information about the edges.

We denote by $\mathbb{T}_k$ the set of all non-crossing alternating trees with k vertices. For example, $|\mathbb{T}_3| = 2$ (see Figure 3.13), and $|\mathbb{T}_4| = 5$ (see Figure 3.14).

In the following we will treat any non-crossing alternating tree as a directed graph in which any edge (b, w) is oriented from a black vertex and towards a white vertex; in other words from the left vertex to the right vertex.

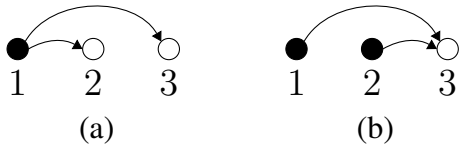


Figure 3.13: All non-crossing alternating trees with 3 vertices.

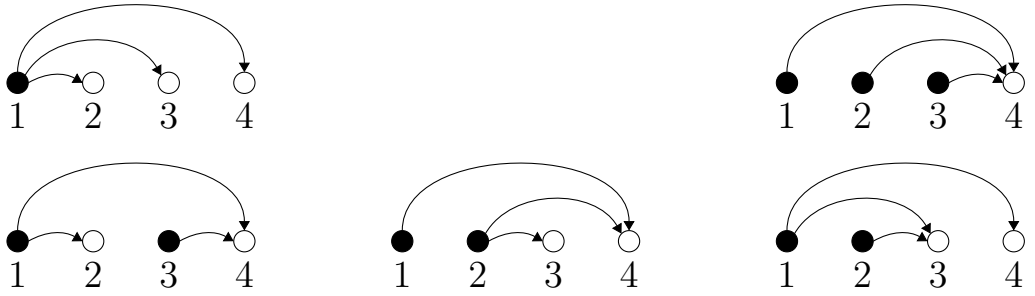


Figure 3.14: All non-crossing alternating trees with 4 vertices.

3.3.7 Closed formula for cumulants

Let λ be a fixed Young diagram. Let T be a uniformly random Poissonized tableau of shape λ, and $F_T(u_0)$ be the cumulative function of T. We will now present a closed formula for the cumulants of the random variable $F_T(u_0)$. This formula is convenient for proving that in a suitable asymptotic setting the random variable $F_T(u_0)$ (after suitable shift and rescaling) converges in distribution towards a Gaussian measure.

Assume that the discrete set $\mathbb{X}$ of the decoration values contains the set of u-coordinates of all concave corners of λ. In other words, we assume that the support of the transition measure of λ is contained in $\mathbb{X}$. For example, we may take $\mathbb{X} = \mathbb{Z}$ to be the set of integers. Note that the set $D_H(u_0)$ which is used below depends implicitly on this choice of $\mathbb{X}$.

Theorem 3.3.2 (The main result)**.** *With the above notations, for each $u_0 \in \mathbb{R}$ the k-th cumulant of the random variable $F_T(u_0)$ is given by*

$$\kappa_k\left(F_T\left(u_0\right)\right) = (k-1)! \sum_{H \in \mathbb{T}_k} \sum_{\mathbf{x} \in D_H(u_0)} \frac{(-1)^{|B_H|-1} \prod_{j=1}^{k} \mu_\lambda(x_j)}{\prod_{(b,w)\in E_H} (x_w - x_b + w - b)}, \tag{3.3.2}$$

where $\mu_\lambda(x_j)$ denotes the probability corresponding to the atom x_j of the transition measure μ_λ.

The proof is postponed to Section 3.7.12 .

Example 3.3.3. The expected value and the variance of $F_T(u_0)$ are given by

$$\mathbb{E}F_T\left(u_0\right) = \sum_{x_1 \le u_0} \mu_\lambda(x_1), \tag{3.3.3}$$

$$\operatorname{Var} F_T\left(u_0\right) = \sum_{\substack{x_1 \le u_0 \\ x_2 > u_0}} \frac{1}{x_2 - x_1 + 1}\, \mu_\lambda(x_1)\, \mu_\lambda(x_2). \tag{3.3.4}$$

The unique summand on the right-hand side of (3.3.3) corresponds to the tree in Figure 3.12a, and the unique summand on the right-hand side of (3.3.4) corresponds to the tree in Figure 3.12b. The third cumulant of $F_T(u_0)$ is given by

$$\begin{aligned} k_3\left(F_T\left(u_0\right)\right) = & \\ & \sum_{\substack{x_1 \le u_0 \\ x_2, x_3 > u_0}} \frac{2}{(x_2 - x_1 + 1)(x_3 - x_1 + 2)} \mu_\lambda(x_1)\, \mu_\lambda(x_2)\, \mu_\lambda(x_3) \\ & \quad - \sum_{\substack{x_1, x_2 \le u_0 \\ x_3 > u_0}} \frac{2}{(x_3 - x_2 + 1)(x_3 - x_1 + 2)} \mu_\lambda(x_1)\, \mu_\lambda(x_2)\, \mu_\lambda(x_3), \end{aligned}$$

where the first summand on the right-hand side corresponds to the tree in Figure 3.13a, and the second summand corresponds to the tree in Figure 3.13b.

Remark 3.3.4. The right-hand side of (3.3.2) can be interpreted as the expected value of the random variable Z defined in the following way. Let $x_1, \ldots, x_k$ be a sequence of independent, identically distributed random variables, with the distribution given by the transition measure μ_λ. Let $\mathbb{T}_k^{\mathbf{x}}$ denote the set of all trees

$H \in \mathbb{T}_k$ such that $\mathbf{x} = (x_1, \ldots, x_k)$ is a u_0-decoration of the tree H, i.e., $\mathbf{x} \in D_H$. The aforementioned random variable is defined as

$$Z = (k-1)! \sum_{H \in \mathbb{T}_k^{\mathbf{x}}} \frac{(-1)^{|B_H|-1}}{\prod\limits_{(b,w)\in E_H} (x_w - x_b + w - b)}.$$

3.3.8 Towards the proof. Rational functions associated to a graph

Our proof of Theorem 3.3.2 is based on algebraic identities fulfilled by some rational multivariate functions. It turns out that the class of the rational functions which we consider is naturally indexed by oriented and weighted graphs, and the aforementioned algebraic identities have a natural combinatorial interpretation as removal of loops from the graphs. We present the details in the following.

For an oriented weighted graph H with the vertex set $V_H = \{v_1, \ldots, v_t\}$, we consider the rational function

$$f_H = f_H(x_{v_1}, \ldots, x_{v_t}) = \frac{1}{\prod\limits_{e=(i,j)\in E_H} [x_j - x_i + w(e)]} \in \mathbb{Q}(x_{v_1}, \ldots, x_{v_t})$$

in the variables corresponding to the vertices of H. In the following we will usually consider the special case when H has the vertex set $V_H = \{1, \ldots, k\}$ so that

$$f_H = f_H(x_1, \ldots, x_k) \in \mathbb{Q}(x_1, \ldots, x_k).$$

For each tree $H \in \mathbb{T}_k$, we define the weight of an edge $e = (i, j) \in E_H$

$$w(e) = w(i, j) = j - i \tag{3.3.5}$$

as the difference of the endpoints. With this convention (3.3.2) can be written as

$$\kappa_k\left(F_T\left(u_0\right)\right) = (k-1)! \sum_{H\in\mathbb{T}_k} \sum_{\mathbf{x}\in D_H(u_0)} (-1)^{|B_H|-1} f_H(x_1, \ldots, x_k) \prod_{j=1}^{k} \mu_\lambda(x_j). \tag{3.3.6}$$

3.4 Anti-Pieri growth process

The aforementioned results of Romik and Śniady were proved using the *anti-Pieri growth process* [RŚ15, Section 4.1]. It should come as no surprise that we also will use this growth process.

3.4.1 Proof of Lemma 3.2.1

Proof of Lemma 3.2.1. In the following we consider the unit cube

$$[0,1]^n = \{(w_1, \ldots, w_n) : w_1, \ldots, w_n \in [0,1]\}$$

(equipped with the Lebesgue measure) with all hyperplanes $w_i - w_j = 0$ (over $1 \leq i < j \leq n$) removed. Since the removed hyperplanes have Lebesgue measure zero, this removal is irrelevant from the viewpoint of the measure theory. The hyperplanes divide the cube into $n!$ isometric simplices, each with the volume $\frac{1}{n!}$. The simplices are in a bijective correspondence with permutations in $\mathfrak{S}_n$; each simplex $\mathcal{S}_\sigma$ consists of the vectors with a prescribed linear order between the coordinates.

For a given Young diagram λ we denote by $\mathcal{T}^\lambda$ the set of Poissonized tableaux of shape λ, equipped with the Lebesgue measure. For simplicity we remove from this set all tableaux which have repeated entries; again this removal is irrelevant from the viewpoint of the measure theory.

With these notations, the Robinson–Schensted correspondence is a bijection between the aforementioned cube $[0,1]^n$ and the disjoint sum

$$\bigsqcup_{\lambda \vdash n} \mathcal{T}^\lambda \times f^\lambda, \tag{3.4.1}$$

where f^λ denotes the set of standard Young tableaux of shape λ. The second component of this correspondence, the map Q, restricted to the simplex $\mathcal{S}_\sigma$ is constant, equal to the recording tableau $Q(\sigma)$. On the other hand, the first component of this correspondence, the map P, restricted to $\mathcal{S}_\sigma$ acts by arranging the entries of $(w_1, \ldots, w_n)$ to the boxes of the diagram $\mathrm{RSK}(\sigma)$. It follows that the Robinson–Schensted correspondence is a piecewise isometry, hence it is a measure-preserving map if we equip f^λ with the counting measure and each summand in (3.4.1) with the product measure.

Conditioning over the event $\mathrm{RSK}(w) = \lambda$ corresponds therefore to considering the uniform measure (or, equivalently, the product measure multiplied by the scalar factor $\frac{n!}{(f^\lambda)^2}$) on a specific summand of (3.4.1), namely

$$\mathcal{T}^\lambda \times f^\lambda$$

which completes the proof. □

3.4.2 Anti-Pieri growth process

We will use the following notation which is intended as an analogue of the falling factorial

$$\mathbf{G}_\lambda^{\underline{k}}(x) = \begin{cases} \underbrace{\mathbf{G}_\lambda(x)\,\mathbf{G}_\lambda(x-1)\,\cdots\,\mathbf{G}_\lambda(x-k+1)}_{k \text{ factors}} & \text{if } k \geq 1, \\ 1 & \text{if } k = 0 \end{cases}$$

for an integer $k \geq 0$.

The following result is due to Śniady [Śni20, Lemma C]. We provide its proof for completeness.

Lemma 3.4.1. *Let λ be a fixed Young diagram with n boxes and $k \geq 1$ be an integer.*

Let

$$\lambda = \xi_n \nearrow \xi_{n+1} \nearrow \cdots \nearrow \xi_{n+k}$$

be given by the Plancherel growth process starting at λ and let $\mathbf{U} = (U_1, \ldots, U_k)$ be the sequence of the u-coordinates of the boxes added in each step, i.e.,

$$U_i = u(\xi_{n+i} \setminus \xi_{n+i-1})$$

for $i \in \{1, \ldots, k\}$.

(a) Let T be a random Poissonized tableau of shape λ. Then for any $u_0 \in \mathbb{R}$ the moment of the random variable $F_T(u_0)$ fulfills

$$\begin{aligned} m_k\left(F_T(u_0)\right) &= \mathbb{E}\left[\left(F_T(u_0)\right)^k\right] \\ &= k!\,\mathbb{P}\big(u_0 \geq U_1 > \cdots > U_k\big). \end{aligned}$$

(b) if $U_1 > \cdots > U_k$ then the tuple $\mathbf{U} = (U_1, \ldots, U_k)$ can be uniquely *written in the form*

$$\mathbf{x}^{\mathbf{a}} = \big(\underbrace{x_1, x_1 - 1, \ldots, x_1 - a_1 + 1}_{a_1 \text{ times}}, \ldots, \underbrace{x_\ell, x_\ell - 1, \ldots, x_\ell - a_\ell + 1}_{a_\ell \text{ times}}\big)$$

with $\mathbf{x} = (x_1, \ldots, x_\ell)$ and $\mathbf{a} = (a_1, \ldots, a_\ell)$, where $x_1 > \cdots > x_\ell$ are u-coordinates of some concave corners of λ and $a_1, \ldots, a_\ell \geq 1$ are integers such that $a_1 \cdots + a_\ell = k$.

(c) Let $x_1, \dots, x_\ell$ be the u-coordinates of some concave corners of λ. Let $a_1, \dots, a_\ell \geq 1$ be integers such that $a_1 + \cdots + a_\ell = k$. We assume that the following condition holds true:

(X) for each $i \in \{1, \dots, \ell\}$ the set

$$\{x_i - 1, x_i - 2, \dots, x_i - a_i + 1\}$$

and the set of u-coordinates of the concave corners of λ are disjoint.

Then

$$\mathbb{P}\big[\mathbf{U} = \mathbf{x}^{\mathbf{a}}\big] = \Theta(x_1, \dots, x_\ell) \prod_{1 \leq i \leq \ell} \frac{(-1)^{a_i - 1}}{a_i} \mu_\lambda(x_i)\, \mathbf{G}_\lambda^{a_i - 1}(x_i - 1), \quad (3.4.2)$$

where

$$\Theta(x_1, \dots, x_\ell) = \prod_{1 \leq i < j \leq \ell} \frac{(x_i - x_j)(x_i - x_j - a_i + a_j)}{(x_i - x_j + a_j)(x_i - x_j - a_i)}. \quad (3.4.3)$$

Note that the assumption (X) guarantees that on the right-hand side of (3.4.2) we do not evaluate the Cauchy transform $\mathbf{G}_\lambda$ or the function Θ in a singularity.

Proof. Proof of part (a). Our general strategy is to create a coupling and to create on a single probability space both a uniformly random Poissonized tableau with shape λ, as well as a Plancherel growth process starting in λ. A small additional difficulty is that our model will require some conditioning.

Let $w_1, \dots, w_{n+k}$ be a sequence of independent random variables with the uniform distribution $U(0, 1)$ and let $\Xi_i = \mathrm{RSK}(w_1, \dots, w_i)$; then

$$\Xi_0 \nearrow \cdots \nearrow \Xi_{n+k}$$

is the Plancherel growth process. For $i \in \{1, \dots, k\}$ we denote by $\mathcal{U}_i = u(\Xi_{n+i} \setminus \Xi_{n+i-1})$ the u-coordinate of the place where the growth occurs.

Clearly, the probability distribution of the Plancherel growth process $(\xi_n, \dots, \xi_{n+k})$ starting at $\xi_n = \lambda$ coincides with the *conditional* probability distribution of its counterpart $\big(\Xi_n, \dots, \Xi_{n+k}\big)$, under the condition that $\Xi_n = \lambda$. By Lemma 3.2.1, the probability distribution of the random Poissonized tableau T from the statement of the lemma coincides with the *conditional* probability distribution of the insertion tableau $\mathcal{T} := P(w_1, \dots, w_n)$, under the condition $\Xi_n = \lambda$.

These observations imply that it is enough to prove equality between the conditional expectations

$$\mathbb{E}\left[\left(F_{\mathcal{T}}(u_0)\right)^k \;\middle|\; \sigma\big(\Xi_n\big)\right] \;=\; k!\;\mathbb{E}\left[\mathbb{1}\big(u_0 \geq \mathcal{U}_1 > \cdots > \mathcal{U}_k\big) \;\middle|\; \sigma\big(\Xi_n\big)\right], \quad (3.4.4)$$

where $\sigma\big(\Xi_n\big)$ denotes the σ-algebra generated by the random Young diagram Ξ_n, and $\mathbb{1}\{A\}$ denotes the indicator random variable which takes the value 1 if the condition A holds true, and 0 otherwise.

For a moment let us fix the values in the prefix $x_1, \ldots, x_n$; the conditional probability

$$\begin{aligned}
&\mathbb{P}\left[F_{\mathcal{T}}(u_0) > w_{n+1} > \cdots > w_{n+k} \;\middle|\; \sigma(x_1, \ldots, x_n)\right] \qquad (3.4.5)\\
&= \mathbb{E}\left[\mathbb{1}\left\{F_{\mathcal{T}}(u_0) > w_{n+1} > \cdots > w_{n+k}\right\} \;\middle|\; \sigma(x_1, \ldots, x_n)\right]\\
&= \operatorname{vol}\left\{(x_{n+1}, \ldots, x_{n+k}) \in [0,1]^k \colon F_{\mathcal{T}}(u_0) > w_{n+1} > \cdots > w_{n+k}\right\}\\
&= \frac{1}{k!}\left[F_{\mathcal{T}}(u_0)\right]^k
\end{aligned}$$

is then directly related to the value of the random variable $F_{\mathcal{T}}(u_0)$.

The event which appears on the left hand side of (3.4.5) can be alternatively reformulated in the language of the Young diagrams $\Xi_n \nearrow \cdots \nearrow \Xi_{n+k}$ as follows:

$$\left\{F_T(u_0) > w_{n+1} > \cdots > w_{n+k}\right\} = \left\{u_0 \geq \mathcal{U}_1 > \cdots > \mathcal{U}_k\right\}. \qquad (3.4.6)$$

Indeed, the equivalence

$$F_T(u_0) > w_{n+1} \iff u_0 \geq \mathcal{U}_1$$

is a consequence of the definition of $F_T(u_0)$ while each of the equivalences

$$w_{n+i} > w_{n+i+1} \iff \mathcal{U}_i > \mathcal{U}_{i+1}$$

is the content of the Row Bumping Lemma [Ful97, page 9]. Thus, by taking the appropriate conditional expectation of both sides of (3.4.5), the desired equality (3.4.4) follows immediately.

The part (b) is obvious.

The proof of part (c) is a straightforward but tedious calculation of the transition probabilities of the Markov chain (ξ_m) (by taking the residue of the appropriate Cauchy transform, see Section 3.2.2), and then the product of these transition probabilities. Each of these expressions is a complicated rational function involving many factors, the only difficulty is related to keeping track of the factors which cancel. $\square$

Figure 3.15: An example of a spine graph with 6 vertices.

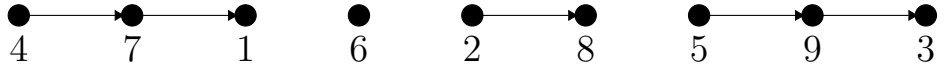


Figure 3.16: An example of a multi-spine graph with 9 vertices and 4 connected components.

3.5 Decomposition into simple fractions

In this section we will decompose the product Θ defined in (3.4.3) into a sum of simple fractions.

A *spine graph* with $\ell \geq 1$ vertices is defined as a connected directed graph F such that the set of its edges consists of $\ell - 1$ elements and is of the form

$$E_F = \{(v_1, v_2), \ldots, (v_{\ell-1}, v_\ell)\}.$$

Note that the vertices $v_1, \ldots, v_\ell$ are all different; otherwise, the graph would not be connected. We denote the set of all spine graphs with the vertex set $V = \{1, \ldots, \ell\}$ by S_ℓ; obviously $|S_\ell| = \ell!$. An example of a spine graph is shown in Figure 3.15.

A multi-spine graph is defined as any directed graph such that each component is a spine graph. In other words, a multi-spine graph is a forest of spine graphs. We denote the set of all multi-spine graphs with the vertex set $V = \{1, \ldots, \ell\}$ by MS_ℓ. An example of a multi-spine graph is shown in Figure 3.16.

Lemma 3.5.1. *Let $(a_1, \ldots a_\ell)$ be a sequence of numbers which has the property that the sum of the entries of any non-empty subsequence is non-zero (this condition holds, for example, if $a_1, \ldots, a_\ell > 0$ are all positive).*

Then the element $\Theta \in \mathbb{R}(x_1, \ldots, x_\ell)$ of the field of rational functions defined in (3.4.3) *can be written as the sum*

$$\Theta(x_1, \ldots, x_\ell) = \sum_{F \in \mathrm{MS}_\ell} \frac{\beta_F}{\prod\limits_{(i,j) \in E_F} (x_j - x_i + a_i)}. \tag{3.5.1}$$

Above, for any graph $F \in \mathrm{MS}_\ell$, the constant β_F is defined as

$$\beta_F = (-1)^{|V_F|} \frac{\prod\limits_{j=1}^{\ell} a_j}{\prod\limits_{F'} \left[(-1) \cdot \sum\limits_{i \in V_{F'}} a_i \right]}, \tag{3.5.2}$$

where the product over F' runs over all connected components of the graph F.

Proof. To simplify the notation, we put

$$z_j = x_j - a_j$$

for each index $j \in \{1, \dots, \ell\}$.

Let

$$A = \left[\frac{1}{x_i - z_j} \right]_{1 \le i,j \le \ell}$$

be the Cauchy matrix [Sch59]. Its determinant, called the *Cauchy determinant*, is given by the following product formula [Sch59]

$$\det A = \frac{\prod\limits_{1 \le i < j \le \ell} (x_i - x_j)(z_j - z_i)}{\prod\limits_{1 \le i,j \le \ell} (x_j - z_i)}.$$

The denominator of Θ differs from its counterpart in the Cauchy determinant only by the missing diagonal factors $x_j - z_i$ for $i = j$. Thus

$$\Theta = \left(\prod_{j=1}^{\ell} (x_j - z_j) \right) \det A = \left(\prod_{j=1}^{\ell} a_j \right) \det A.$$

Using the definition of the determinant we express Θ as a sum over permutations

$$\Theta = \left(\prod_{j=1}^{\ell} a_j \right) \sum_{\sigma \in \mathfrak{S}_k} \frac{(-1)^{\ell - c(\sigma)}}{\prod\limits_{i=1}^{\ell} (x_{\sigma(i)} - z_i)},$$

where $c(\sigma)$ denotes the number of cycles of the permutation σ. We can treat each permutation $\sigma \in \mathfrak{S}_n$ as a directed weighted graph with the vertex set $V_\sigma = \{1, \dots, \ell\}$ and with the edge set

$$E_\sigma = \left\{ \big(1, \sigma(1)\big), \dots, \big(\ell, \sigma(\ell)\big) \right\}.$$

We define the weight of an edge $e = \big(i, \sigma(i)\big)$ as $w(e) = a_i$. Therefore

$$\Theta = \left(\prod_{j=1}^{\ell} a_j \right) \sum_{\sigma \in \mathfrak{S}_k} (-1)^{\ell - c(\sigma)} f_\sigma = (-1)^\ell \left(\prod_{j=1}^{\ell} a_j \right) \sum_{\sigma \in \mathfrak{S}_k} \prod_{\sigma'} (-f_{\sigma'}), \quad (3.5.3)$$

where σ' runs over the connected components of the directed graph σ; note that each such connected component corresponds to a cycle of the permutation σ.

Let σ' be a connected component of the directed graph σ. Using the identity

$$\sum_{i \in V_{\sigma'}} a_i = \sum_{j \in V_{\sigma'}} x_j - \sum_{i \in V_{\sigma'}} z_i = \sum_{(i,j) \in E_{\sigma'}} \big(x_j - z_i\big)$$

we obtain

$$f_{\sigma'} \sum_{i \in V_{\sigma'}} a_i = f_{\sigma'} \sum_{(i,j) \in E_{\sigma'}} (x_j - z_i) = \sum_{F'} f_{F'}, \quad (3.5.4)$$

where F' runs over all spine graphs obtained from the cycle σ' by removing exactly one edge.

Equation (3.5.4) can be written as

$$f_{\sigma'} = \frac{1}{\sum\limits_{i \in V_{\sigma'}} a_i} \sum_{F'} f_{F'};$$

we apply this identity to each cycle σ' of the permutation $\sigma \in \mathfrak{S}_k$ on the right-hand side of (3.5.3). Note that the above equality holds true also in the special case when the cycle σ' is a fix-point; in this case we remove the loop from the directed graph σ', and the unique resulting graph F' has one isolated vertex and no edges.

If we remove one edge from each cycle of every permutation in all possible ways, we obtain each multi-spine graph on the vertex set $\{1, \dots, \ell\}$ exactly once. In this way we proved that

$$\Theta = (-1)^\ell \left(\prod_{j=1}^{\ell} a_j \right) \sum_{F \in \mathrm{MS}_\ell} \prod_{F'} \frac{-f_{F'}}{\sum\limits_{i \in V_{F'}} a_i}$$

where F' runs over the connected components of the graph F, as required. $\square$

3.6 The moments of the cumulative function

3.6.1 The first formula for the moments

A *composition* of a natural number k is an expression of k as an ordered sum of positive integers $k = a_1 + \cdots + a_\ell$. The set of all compositions of k will be denoted by Comp_k. For a given composition $\mathbf{a} = (a_1, \ldots, a_\ell) \in \mathrm{Comp}_k$ we denote the number of its parts by $\ell = \ell(\mathbf{a})$.

Using Lemma 3.4.1 we obtain

$$m_k\left(F_T\left(u_0\right)\right) = k! \sum_{\mathbf{a}\in\mathrm{Comp}_k} \sum_{\mathbf{x}} \mathbb{P}\big[\mathbf{U} = \mathbf{x}^{\underline{\mathbf{a}}}\big]$$

$$= k! \sum_{\mathbf{a}\in\mathrm{Comp}_k} \sum_{\mathbf{x}} \Theta(x_1, \ldots, x_\ell) \prod_{i=1}^{\ell} \frac{(-1)^{a_i-1}}{a_i}\, \mu_\lambda(x_i)\, \mathbf{G}_\lambda^{a_i-1}(x_i - 1), \quad (3.6.1)$$

where in each expression the second sum runs over $\mathbf{x} = (x_1, \ldots, x_\ell) \in \mathbb{X}$ such that

$$u_0 \geq x_1 > x_2 > \cdots > x_\ell \quad (3.6.2)$$

and such that the condition (X) from Lemma 3.4.1 is satisfied.

The aforementioned condition (X) turns out to be quite cumbersome in applications. For this reason our strategy is to obtain an analogue of the above formula (3.6.1) which would involve summation over *all* $x_1, \ldots, x_\ell \in \mathbb{X}$ which fulfill (3.6.2), i.e., to remove the requirement (X). Regretfully, without this additional condition it might happen that one of the factors in the falling product $\mathbf{G}_\lambda^{a_i-1}(x_i - 1)$ is evaluated in a singularity; thus the right-hand side of (3.6.1) might involve division by zero.

In order to avoid this difficulty instead of Young diagrams we will consider a more general class of objects, namely *interlacing sequences*, for which such a division by zero can be easily avoided. The formulas for the Young diagram λ can be then obtained by an appropriate limit.

3.6.2 Interlacing sequences

The following notations are based on the work of Kerov [Ker93]. We say that

$$\Lambda = (\mathbb{x}_0, \ldots, \mathbb{x}_\mathbb{L};\ \mathbb{y}_1, \ldots, \mathbb{y}_\mathbb{L}) \quad (3.6.3)$$

is an *interlacing sequence* if its entries are real numbers such that

$$\mathbb{x}_0 < \mathbb{y}_1 < \mathbb{x}_1 < \cdots < \mathbb{y}_{\mathbb{L}} < \mathbb{x}_{\mathbb{L}}.$$

Following Figure 3.7 and Section 3.2.2, each Young diagram can be regarded as a specific interlacing sequence. Conversely, each interlacing sequence can be visualized as the zig-zag curve analogous to the one from Figure 3.7; for this reason we will refer to the entries of the sequence $\mathbb{x}_0, \ldots, \mathbb{x}_{\mathbb{L}}$ as *concave corners* and to the entries of the sequence $\mathbb{y}_1, \ldots, \mathbb{y}_{\mathbb{L}}$ as *convex corners*.

The Cauchy transform $\mathbf{G}_\Lambda$ and the transition measure μ_Λ of an interlacing sequence Λ is defined in an analogous way as their counterparts for Young diagrams in Section 3.2.2.

3.6.3 Moments for interlacing sequences

Let an interlacing sequence Λ be fixed. We assume that the set of concave corners is *generic*, i.e., if $i \neq j$ then $\mathbb{x}_i - \mathbb{x}_j$ is *not* an integer. For the set of decoration values $\mathbb{X} := \{\mathbb{x}_0, \ldots, \mathbb{x}_{\mathbb{L}}\}$ we take the concave corners. Let u_0 be a fixed real number. We define the k-th *moment* for the interlacing sequence Λ as

$$M_k = M_k(\Lambda, u_0) = \\ k! \sum_{\mathbf{a} \in \mathcal{C}_n} \sum_{\mathbf{x}} \Theta(x_1, \ldots, x_\ell) \prod_{i=1}^{\ell} \frac{(-1)^{a_i - 1}}{a_i} \, \mu_\Lambda(x_i) \, \mathbf{G}_\Lambda^{a_i - 1}(x_i - 1), \quad (3.6.4)$$

where the sum over $\mathbf{x}$ runs over $x_1, \ldots, x_\ell \in \mathbb{X}$ such that (3.6.2) holds true, and $\ell = \ell(\mathbf{a})$ denotes the length of the composition $\mathbf{a}$ as before. The assumption that the set of concave corners is generic guarantees that the right-hand side is well-defined. One can ask if the quantity $M_k(\Lambda, u_0)$ has a probabilistic interpretation as a moment of some natural random variable associated to the interlacing sequence Λ; we expect that the answer for this question is negative. We will use M_k purely as an auxiliary tool for studying the moments of the random variable $F_T(u_0)$, see below.

The right-hand side of (3.6.4) is very similar to its counterpart (3.6.1); the only difference is that the second sum on the right-hand side of (3.6.1) runs over certain sequences $\mathbf{x}$ which *additionally* fulfill the condition (X) from Lemma 3.4.1(c).

Let us fix an integer $s \in \{0, \ldots, \mathbb{L} + 1\}$ and consider the set $W_{s,\mathbb{L}}$ of interlacing sequences Λ of the form (3.6.3) with the property that $\mathbb{x}_0, \mathbb{x}_1, \ldots, \mathbb{x}_{s-1} \leq u_0$ are all small and $\mathbb{x}_s, \ldots, \mathbb{x}_{\mathbb{L}} > u_0$ are all big; in other words

s is the cardinality of small entries of the set $\mathbb{X}$. Thanks to the aforementioned removal of the condition (X), the restriction of the function $\Lambda \mapsto M_k(\Lambda, u_0)$ to the set $W_{s,\mathbb{L}}$ is a rational function in the variables $\mathbb{x}_0, \dots, \mathbb{x}_{\mathbb{L}}, \mathbb{y}_1, \dots, \mathbb{y}_{\mathbb{L}}$. Our general strategy is to investigate this rational function M_k.

The price we have to pay for the aforementioned omission of the condition (X) is that the rational function M_k is singular in (some subset of) the set of non-generic interlacing sequences, in particular it is not clear how to evaluate $M_k(\Lambda, u_0)$ in the special case when the interlacing sequence Λ corresponds to a Young diagram λ (which is clearly non-generic). On the bright side, Lemma 3.6.1 below shows that there is a special way of taking the limit value of M_k at the singularity which provides a bridge with our main subject of investigations, the moment $m_k\left(F_T\left(u_0\right)\right)$. In fact, from the proof of Theorem 3.3.2 it will follow that the aforementioned singularity is removable and thus an analogue of Lemma 3.6.1 holds true for *any* way of taking the limit $\Lambda \to \lambda$.

3.6.4 Regularization

Let a Young diagram λ be fixed and let Λ be the corresponding interlacing sequence. For $\epsilon > 0$ we define the interlacing sequence

$$\Lambda^\epsilon = (\mathbb{x}_0^\epsilon, \dots, \mathbb{x}_{\mathbb{L}}^\epsilon;\ \mathbb{y}_1^\epsilon, \dots, \mathbb{y}_{\mathbb{L}}^\epsilon)$$

given by

$$\mathbb{x}_j^\epsilon = \mathbb{x}_j + j\epsilon, \qquad \mathbb{y}_j^\epsilon = \mathbb{y}_j + j\epsilon.$$

Note that if ϵ is small enough, the set of concave corners of Λ^ϵ is generic so that $M_k(\Lambda^\epsilon, u_0)$ is well-defined.

The distance

$$\mathbb{x}_j^\epsilon - \mathbb{y}_j^\epsilon = \mathbb{x}_j - \mathbb{y}_j \tag{3.6.5}$$

between any convex corner $\mathbb{y}_j^\epsilon$ and the next concave corner to the right $\mathbb{x}_j^\epsilon$ does not depend on the value of ϵ, and is a positive integer which has a natural interpretation for the original Young diagram λ, cf. Figure 3.7.

Lemma 3.6.1. *We suppose that u_0 is* not *an integer number. With the above notations, the moment m_k is equal to the limit of the moment M_k, when ϵ tends to zero:*

$$m_k\left(F_T\left(u_0\right)\right) = \lim_{\epsilon \to 0} M_k(\Lambda^\epsilon, u_0).$$

Proof. Let $s \in \{0, \dots, \mathbb{L}+1\}$ be the cardinality of the small concave corners of λ; with the notations of Section 3.6.3 this means that $\Lambda^\epsilon \in W_{s,\mathbb{L}}$ if $|\epsilon|$ is small enough. By writing $x_i = \mathbb{x}^\epsilon_{r_i}$ we may write (3.6.4) as

$$M_k(\Lambda^\epsilon, u_0) = \\ k! \sum_{\mathbf{a}\in\mathcal{C}_n} \sum_{s\geq r_1>\cdots>r_\ell\geq 1} \Theta(\mathbb{x}^\epsilon_{r_1}, \dots, \mathbb{x}^\epsilon_{r_\ell}) \prod_{i=1}^{\ell} \frac{(-1)^{a_i-1}}{a_i}\, \mu_\Lambda(\mathbb{x}^\epsilon_{r_i})\, \mathbf{G}^{a_i-1}_{\Lambda^\epsilon}(\mathbb{x}^\epsilon_{r_i} - 1). \tag{3.6.6}$$

Similarly (3.6.1) can be written as

$$m_k\left(F_T\left(u_0\right)\right) = \\ k! \sum_{\mathbf{a}\in\mathcal{C}_n} \sum \Theta(\mathbb{x}_{r_1}, \dots, \mathbb{x}_{r_\ell}) \prod_{i=1}^{\ell} \frac{(-1)^{a_i-1}}{a_i}\, \mu_\Lambda(\mathbb{x}_{r_i})\, \mathbf{G}^{a_i-1}_{\Lambda}(\mathbb{x}_{r_i} - 1); \tag{3.6.7}$$

the consequence of the condition (X) from Lemma 3.4.1 is that the second sum runs over $s \geq r_1 > \cdots > r_\ell \geq 1$ which additionally fulfill

$$\mathbb{x}_{r_i} - \mathbb{y}_{r_i} \geq a_i \qquad \text{for } i \in \{1, \dots, \ell\}; \tag{3.6.8}$$

in the special case when $r_i = 0$ and $\mathbb{y}_0$ is not defined the above condition is fulfilled by convention.

Let us consider a summand of (3.6.6) which corresponds to $\mathbf{a} \in \mathcal{C}_n$ and a tuple $(r_1, \dots, r_\ell)$ for which (3.6.8) is *not* satisfied thus $1 \leq \mathbb{x}_{r_i} - \mathbb{y}_{r_i} \leq a_i - 1$ for some choice of the index i. One of the factors in $\mathbf{G}^{a_i-1}_{\Lambda^\epsilon}(\mathbb{x}^\epsilon_{r_i} - 1)$ is equal to

$$\mathbf{G}_{\Lambda^\epsilon}\left(\mathbb{x}^\epsilon_{r_i} - (\mathbb{x}_{r_i} - \mathbb{y}_{r_i})\right) = \mathbf{G}_{\Lambda^\epsilon}\left(\mathbb{x}^\epsilon_{r_i} - (\mathbb{x}^\epsilon_{r_i} - \mathbb{y}^\epsilon_{r_i}\right) = \mathbf{G}_{\Lambda^\epsilon}\left(\mathbb{y}^\epsilon_{r_i}\right) = 0$$

by the very definition of the Cauchy transform; as a consequence the whole corresponding summand of (3.6.6) vanishes as well.

On the other hand, any summand in (3.6.6) for which (3.6.8) is satisfied is continuous at $\epsilon = 0$ and clearly converges as $\epsilon \to 0$ to its counterpart in (3.6.7) which completes the proof. □

3.6.5 Cumulants for interlacing sequences

For a given interlacing sequence Λ and u_0 we consider the corresponding sequence of moments $M_1, M_2, \dots$ with $M_k = M_k(\Lambda, u_0)$ given by (3.6.4). We

revisit Section 3.3.3 and consider the corresponding sequence of formal cumulants $K_1, K_2, \dots$ with $K_k = K_k(\Lambda, u_0)$ given by the expansion

$$\log \sum_{k=0}^{\infty} \frac{M_k}{k!} t^k = \sum_{k=1}^{\infty} K_k \frac{t^k}{k!}.$$

Since each cumulant K_k can be expressed as a polynomial in the moments $M_1, \dots, M_k$, Lemma 3.6.1 implies the following result.

Lemma 3.6.2. *Suppose that u_0 is* not *an integer number. With the above notations, the cumulants of the random variable $F_T(u_0)$ are given by*

$$\kappa_k \left(F_T(u_0) \right) = \lim_{\epsilon \to 0} K_k(\Lambda^{\epsilon}, u_0).$$

3.7 Proof of Theorem 3.3.2

3.7.1 The graph expansion for the moments

Using Lemma 3.5.1 and the fact that for any integer $r \geq 1$

$$\mathbf{G}_{\Lambda}(x_i - r) = -\sum_{x_{i,r}} \frac{\mu_{\Lambda}(x_{i,r})}{x_{i,r} - x_i + r}, \tag{3.7.1}$$

we may rewrite the formula (3.6.4) as follows.

Corollary 3.7.1. *If the interlacing sequence Λ is generic then the moment M_k is given by*

$$M_k(\Lambda, u_0) = k! \sum_{\mathbf{a} \in \mathrm{Comp}_k} \sum_{x_\ell < \dots < x_1 \leq u_0} \prod_{i=1}^{\ell} \frac{\mu_{\Lambda}(x_i)}{a_i} \times \prod_{r=1}^{a_i - 1} \sum_{x_{i,r}} \frac{\mu_{\Lambda}(x_{i,r})}{x_{i,r} - x_i + r} \times \sum_{F \in \mathrm{MS}_{\ell}} \frac{\beta_F}{\prod_{(i,j) \in E_F} \left(x_j - x_i + a_i \right)}, \tag{3.7.2}$$

where $\ell = \ell(\mathbf{a})$ is the length of the composition $\mathbf{a}$. Recall that the constant β_F was defined in (3.5.2)*. The above sums run over $x_i, x_{i,r} \in \{\mathbb{x}_1, \dots, \mathbb{x}_{\mathbb{L}}\}$.*

In the following we denote

$$x_{i,0} := x_i.$$

Now we will define multi-caterpillar graphs and with them we will simplify Corollary 3.7.1.

3.7.2 Multi-caterpillar graphs

By applying the distributive law to the right-hand side of (3.7.2) we obtain a sum of a lot of terms; to each of them we shall associate a certain directed weighted graph G. Each term is a product of

- the numerical factor

$$k!\,\beta_F \prod_{i=1}^{\ell} \frac{1}{a_i} \prod_{r\in\{0,\dots,a_i-1\}} \mu_\Lambda(x_{i,r})$$

  for some multi-spine graph F, and

- the reciprocal of the product of the polynomials of the form

$$(x_{i,r} - x_{i,0} + r) \quad \text{or} \quad (x_{j,0} - x_{i,0} + a_j).$$

The latter product of polynomials is in our focus.

We create *a multi-caterpillar graph* G with k vertices as follows. We create ℓ black vertices and $k-\ell$ red vertices. The black vertices correspond to the variables $x_{1,0},\dots,x_{\ell,0}$. For each $j\in\{1,\dots,\ell\}$ there are a_j-1 red vertices connected with the black vertex $x_{j,0}$; they correspond to the variables $x_{j,1},\dots,x_{j,a_j-1}$. We tag the vertices in such a way that the vertex which corresponds to the variable $x_{i,j}$ is tagged by the pair (i,j). Each factor $(x_{i_2,j_2} - x_{i_1,j_1} + a)$ corresponds to the oriented edge with the weight a from the vertex tagged (i_1,j_1) to the vertex tagged (i_2,j_2).

Example 3.7.2. The graph shown in Figure 3.17 was obtained from the term

$$\frac{1}{(x_{5,1}-x_5+1)(x_{5,2}-x_5+2)(x_{3,1}-x_3+1)(x_{2,1}-x_2+1)}\times \frac{1}{(x_5-x_1+1)(x_4-x_5+3)(x_3-x_4+1)}$$

which is one of the summands in Corollary 3.7.1 which corresponds to $\mathbf{a}=(1,2,2,1,3)$. Figure 3.18 shows the same graph without the Young diagram.

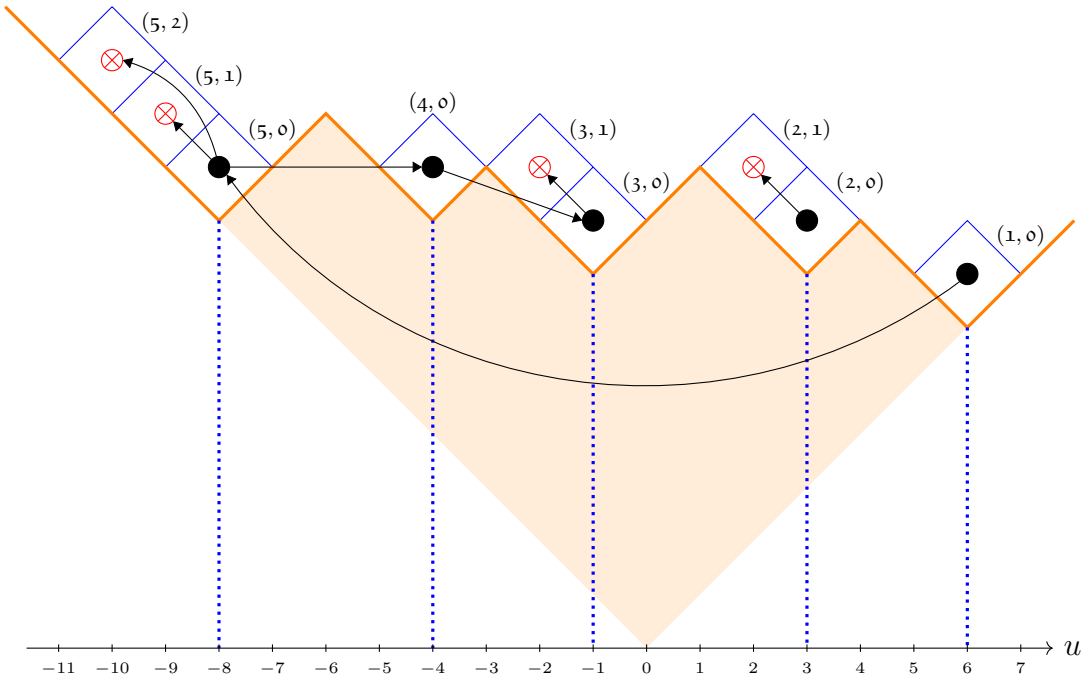


Figure 3.17: The multi-caterpillar graph considered in Example 3.7.2. The composition $\mathbf{a} = (1, 2, 2, 1, 3)$ was visualized as a configuration of white boxes which could occur in the anti-Pieri growth. The vertices of the multi-caterpillar graph (black vertices $\bullet$ and red vertices $\otimes$) correspond to the boxes of the Young diagram where the Plancherel growth occurred. In order to improve visibility the weights of the edges were not shown.

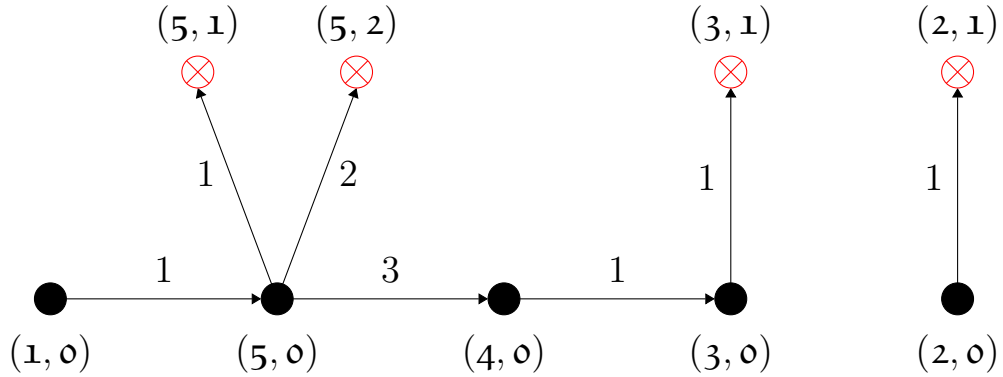


Figure 3.18: The multi-caterpillar graph from Figure 3.17 with the weights of the edges shown.

3.7.3 Multi-caterpillar graphs, the formal approach

More formally, a *multi-caterpillar graph G with tagged vertices* is a directed, weighted graph with black and red vertices, for which there exists a tuple of integers $a_1, \ldots, a_\ell \geq 1$ which fulfills the following properties.

Firstly, the subgraph composed of all black vertices and the edges between them forms a multi-spine graph with $\ell \geq 1$ vertices tagged $(1,0), \ldots, (\ell,0)$. For each $j \in \{1, \ldots, \ell\}$ the black vertex tagged $(j,0)$ has at most one outgoing edge to another black vertex; if such an edge exists, its weight is equal to a_j.

Secondly, if we remove the edges between the black vertices, each connected component of the resulting graph consists of a single black vertex $(j,0)$ for some value of $j \in \{1, \ldots, \ell\}$, and $a_j - 1$ red vertices tagged $(j,1), \ldots, (j, a_j - 1)$. There are no connections between the red vertices; for each $k \in \{1, \ldots, a_j - 1\}$ there is an oriented edge from the black vertex $(j,0)$ to the red vertex (j,k); this edge carries the weight k.

By $\mathrm{MC}_k^{\mathrm{tag}}$ we denote the set of all multi-caterpillar graphs with k tagged vertices.

3.7.4 Three systems of naming the vertices

In the following, we will have twice to use the technique of double counting. For this reason we will use the following three systems of naming the vertices in an oriented weighted graph with k vertices:

- the elements of the set $\mathbb{N} \times \mathbb{N}_0$ will be called *tags*; the naming of the vertices with tags (or, shortly, tagging) was used in Sections 3.7.2 and 3.7.3. An example of a graph with tagged vertices is shown in Figure 3.19a;
- the elements of the set $\{1, \ldots, k\}$ will be called *labels*; we will consider only labelings with the property that if a pair of vertices $v_1, v_2 \in \{1, \ldots, k\}$ is connected by an oriented edge $e = (v_1, v_2)$, its weight

$$w(e) = w(v_1, v_2) = v_2 - v_1 \tag{3.7.3}$$

  is equal to the difference of the vertex labels, see (3.3.5).
- by *marks* we understand the elements of an arbitrary fixed set which consists of k elements; in order to avoid confusion between marks and labels we may declare that the set of marks

$$\{\underline{1}, \underline{2}, \ldots, \underline{k}\} \tag{3.7.4}$$

consists of k underlined integers. An example of a graph with marked vertices is shown in Figure 3.19b;

3.7.5 Black-decreasing decorations

Let G be a multi-caterpillar graph with tagged vertices. The decoration $\mathbf{x} \in D_G$ is called *black-decreasing* if for any pair of black vertices $(p, 0)$ and $(q, 0)$ with $p < q$ the corresponding values of the decoration fulfill $x_{p,0} > x_{q,0}$. The set of all black-decreasing decorations of a multi-caterpillar graph G will be denoted by $D_G^{>}$.

Using Corollary 3.7.1 we can write the moment $M_k(\Lambda, u_0)$ as a sum over multi-caterpillar graphs, as follows. We replace the double sum in (3.7.2) over compositions and over multi-spine graphs by the sum over multi-caterpillar graphs $G \in \mathrm{MC}_k^{\mathrm{tag}}$. In addition, we replace the sum over the variables (x_i) and $(x_{i,r})$ by the sum over black-decreasing decorations. It follows that

$$M_k = k! \sum_{G \in \mathrm{MC}_k^{\mathrm{tag}}} \sum_{\mathbf{x} \in D_G^{>}} \alpha_G f_G, \tag{3.7.5}$$

where the constant α_G is defined as

$$\alpha_G = (-1)^{|B_G|} \left(\prod_{(i,j) \in V_G} \mu_\Lambda(x_{i,j}) \right) \left(\prod_{G'} \frac{-1}{|V_{G'}|} \right), \tag{3.7.6}$$

where G' runs over all connected components of the graph G. Note that α_G depends also on the choice of the decoration $\mathbf{x}$; in order to keep the notation lightweight we will make this dependence implicit.

3.7.6 Double counting

Let $\mathrm{MC}_k^{\mathrm{mark}}$ denote *the set of multi-caterpillar graphs with k marked vertices*, i.e., the set of weighted and oriented graphs G with the vertex set $\{\underline{1}, \ldots, \underline{k}\}$ such that there exists a way to tag the vertices in such a way that G becomes *a multi-caterpillar graph with k tagged vertices* in the sense considered in Section 3.7.3. Let $\mathrm{MC}_k^{\mathrm{tm}}$ denote the set of multi-caterpillar graphs with k vertices which are simultaneously tagged and marked. Examples of such graphs are shown in Figure 3.19.

For any graph $G \in \mathrm{MC}_k^{\mathrm{tag}}$ we have $k!$ ways to mark its k vertices by the elements of (3.7.4).

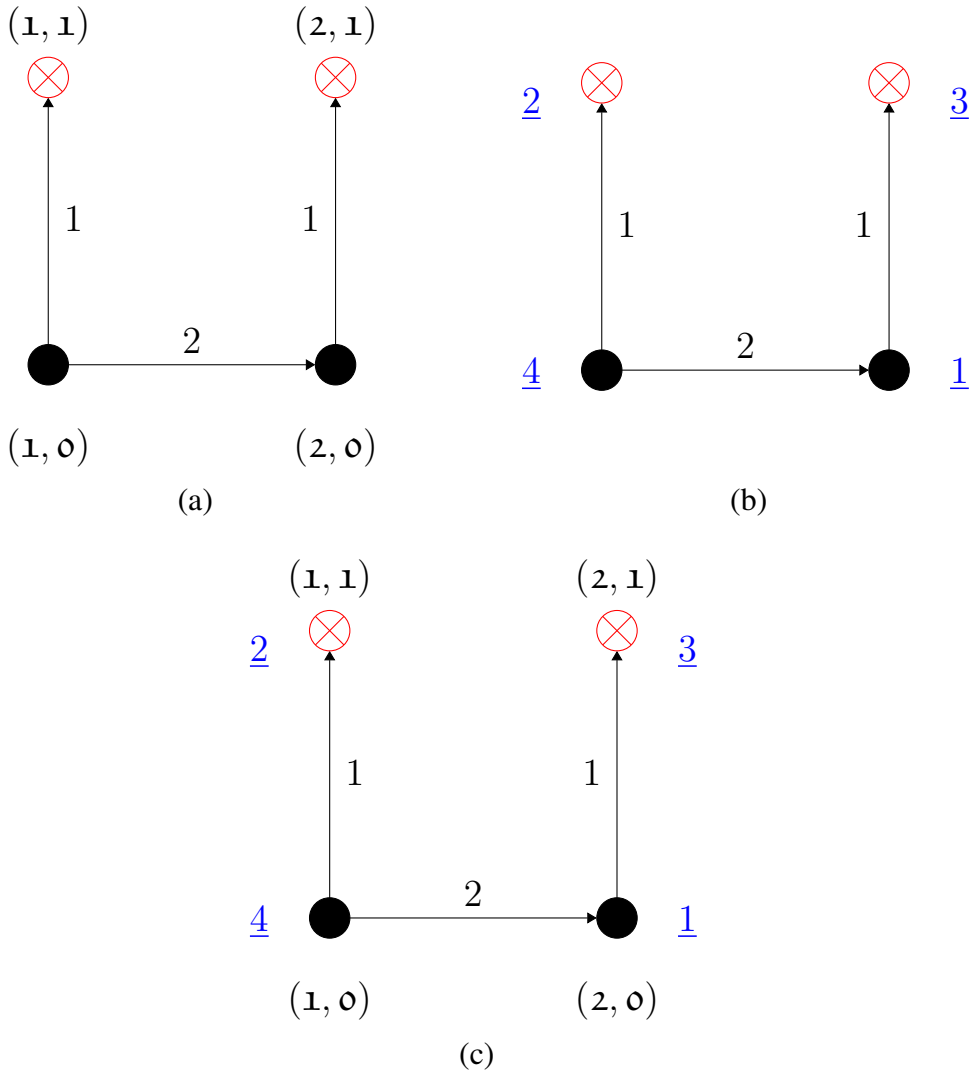


Figure 3.19: Examples of multi-caterpillar graphs: (a) with tagged vertices; the tags are printed black and belong to the set $\{(1,0),(1,1),(2,0),(2,1)\}$, (b) with marked vertices; the marks are printed blue and belong to the set $\{\underline{1},\underline{2},\underline{3},\underline{4}\}$, (c) with tagged and marked vertices.

Let G be a graph. Its decoration $\mathbf{x} \in D_G$ is called *black-injective* if $x_i \neq x_j$ for all pairs of black vertices $i, j \in B_G$ such that $i \neq j$. We denote the set of all black-injective decorations of G by $D_G^{\neq}$, and the set of non-black-injective decorations of G by $D_G^{=} = D_G \setminus D_G^{\neq}$.

Moreover, for each black-injective decoration $\mathbf{x}$ of $G \in \mathrm{MC}_k^{\mathrm{mark}}$ we can tag the vertices of G in a canonical way as follows. We tag the black vertices by $(1,0), (2,0), \dots$ in the opposite order to the one given by the decoration $\mathbf{x}$. Next, for each black vertex b with the tag $(j,0)$ we tag the white vertices connected with b by $(j,1), (j,2), \dots$ according to the increasing order of their corresponding weight of edges. In this way G becomes a caterpillar graph with tagged vertices, and $\mathbf{x}$ becomes a decreasing decoration.

Using these two observations and (3.7.5), we obtain

$$\begin{aligned} M_k &= k! \sum_{G \in \mathrm{MC}_k^{\mathrm{tag}}} \sum_{\mathbf{x} \in D_G^{>}} \alpha_G f_G \\ &= \sum_{G \in \mathrm{MC}_k^{\mathrm{tm}}} \sum_{\mathbf{x} \in D_G^{>}} \alpha_G f_G \\ &= \sum_{G \in \mathrm{MC}_k^{\mathrm{mark}}} \sum_{\mathbf{x} \in D_G^{\neq}} \alpha_G f_G. \end{aligned} \tag{3.7.7}$$

The constant α_G was defined in (3.7.6).

3.7.7 Sum over all decorations

Proposition 3.7.3. *The following double sum over multi-caterpillar graphs and their decorations remains the same when we restrict the sum to black-injective decorations, i.e., for each integer $k \geq 1$,*

$$\sum_{G \in \mathrm{MC}_k^{\mathrm{mark}}} \sum_{\mathbf{x} \in D_G} \alpha_G f_G = \sum_{G \in \mathrm{MC}_k^{\mathrm{mark}}} \sum_{\mathbf{x} \in D_G^{\neq}} \alpha_G f_G. \tag{3.7.8}$$

Proof. We consider the difference of the left-hand side and the right-hand side of (3.7.8)

$$\Delta = \sum_{G \in \mathrm{MC}_k^{\mathrm{mark}}} \sum_{\mathbf{x} \in D_G^{=}} \alpha_G f_G. \tag{3.7.9}$$

Our goal is to prove that $\Delta = 0$.

Let k be a fixed natural number, and let $B = \{b_1, \dots, b_l\} \subseteq \{\underline{1}, \dots, \underline{k}\}$ be a fixed set. Let $\mathrm{MS}(B)$ denote the set of all multi-spine graphs F with the vertex set $V_F = B$. In particular $\mathrm{MS}\left(\{\underline{1}, \dots, \underline{k}\}\right) = \mathrm{MS}_k$. Let $\mathcal{G}_k^{\mathrm{MC}}(B)$ denote the set of all multi-caterpillar graphs $G_\emptyset \in \mathrm{MC}_k^{\mathrm{mark}}$ with k marked vertices $\underline{1}, \dots, \underline{k}$ such that

- the set of black vertices of $G_\emptyset$ is given by $B_{G_\emptyset} = B$, and
- there is no edge in $G_\emptyset$ connecting two black vertices.

Let k be a fixed natural number. Every multi-caterpillar graph $G \in \mathrm{MC}_k^{\mathrm{mark}}$ can be decomposed in a unique way into the union of two graphs $G_\emptyset \in \mathcal{G}_k^{\mathrm{MC}}(B_G)$ and $F \in \mathrm{MS}(B_G)$. In other words, the graph F is the graph composed of all black vertices of the graph G and the edges between them, and the graph $G_\emptyset$ is the graph composed of all vertices of the graph G and the remaining edges. Furthermore, for each vertex $v \in V_F$ of the graph F we define the number a_v as the number of vertices in the connected component of the graph $G_\emptyset$ which contains the vertex v. With the notations from Lemma 3.5.1, the constant β_F given by (3.5.2) is equal to

$$\beta_F = (-1)^{|V_F|} \frac{\prod\limits_{v \in V_F} a_v}{\prod\limits_{G'} \left[(-1)\,|V_{G'}|\right]}, \tag{3.7.10}$$

where the product over G' runs over all connected components of the graph G. In addition, for each edge $e = (i, j) \in E_F$ we define its weight as $w(e) = a_i$.

Now we define the constant $\gamma_{G_\emptyset}$ so that

$$\alpha_G = \beta_F \, \gamma_{G_\emptyset}.$$

From (3.7.6) and (3.7.10) we obtain that the constant

$$\gamma_{G_\emptyset} = \frac{\alpha_G}{\beta_F} = \frac{\prod\limits_{v \in V_{G_\emptyset}} \mu_\Lambda(x_v)}{\prod\limits_{v \in B_{G_\emptyset}} a_v}$$

depends only on the graph $G_\emptyset$ and the decoration $\mathbf{x}$.

In addition, for any set $B \subseteq \{\underline{1}, \dots, \underline{k}\}$ the union of each pair of graphs $G_\emptyset \in \mathcal{G}_k^{\mathrm{MC}}(B)$ and $F \in \mathrm{MS}(B)$ as above is a multi-caterpillar graph with k marked vertices. Therefore, we can replace the sum in (3.7.9) over all multi-caterpillar

graphs with marked vertices by a triple sum over all possible sets of black vertices, over multi-caterpillar graphs, and over all multi-spine graphs. It follows that

$$\begin{aligned}\Delta &= \sum_{B\subseteq\{\underline{1},\dots,\underline{k}\}} \sum_{\substack{G\in \mathrm{MC}_k^{\mathrm{mark}} \\ B_G=B}} \sum_{\mathbf{x}\in D_G^{=}} \alpha_G f_G \\ &= \sum_{B\subseteq\{\underline{1},\dots,\underline{k}\}} \sum_{G_\emptyset\in \mathcal{G}_k^{\mathrm{MC}}(B)} \sum_{\mathbf{x}\in D_{G_\emptyset}^{=}} \gamma_{G_\emptyset} f_{G_\emptyset} \sum_{F\in \mathrm{MS}(B)} \beta_F f_F.\end{aligned}$$

Let $B = \{b_1, \dots, b_l\} \subseteq \{\underline{1}, \dots, \underline{k}\}$ be a fixed set and $\mathbf{x}_B = (x_{b_1}, \dots, x_{b_l})$ be a fixed non-black-injective decoration of B. Using the formula (3.5.1) (e.g., by temporarily renumbering the vertices $b_1, \dots, b_l$ into $1, \dots, l$), we obtain that the internal sum is equal to

$$\sum_{F\in \mathrm{MS}(B)} \beta_F f_F(x_{b_1}, \dots, x_{b_l}) = \Theta(x_{b_1}, \dots, x_{b_l}) = 0$$

since at least one of the factors in the numerator of Θ is equal to zero. (Recall that the definition of Θ was given in (3.4.3).) Thus $\Delta = 0$ as required. □

3.7.8 The first formula for the cumulants

We denote by $C_k^{\mathrm{mark}} \subset \mathrm{MC}_k^{\mathrm{mark}}$ the set of *connected multi-caterpillar graphs with k marked vertices*. Its elements will be called *caterpillar graphs with k marked vertices*.

Using Proposition 3.7.3 we transform the formula (3.7.7) to

$$M_k = \sum_{G\in \mathrm{MC}_k^{\mathrm{mark}}} \sum_{\mathbf{x}\in D_G} \alpha_G f_G.$$

We can look separately at each connected component G' of a multi-caterpillar graph G. The connected components correspond to the blocks of a set-partition.

Thus

$$\begin{aligned} M_k &= \sum_{G \in \mathrm{MC}_k^{\mathrm{mark}}} \sum_{\mathbf{x} \in D_G} \alpha_G f_G \\ &= \sum_{G \in \mathrm{MC}_k^{\mathrm{mark}}} \prod_{G'} \sum_{\mathbf{x} \in D_{G'}} \alpha_{G'} f_{G'} \\ &= \sum_{\pi} \prod_{b \in \pi} \sum_{G' \in C_b^{\mathrm{mark}}} \sum_{\mathbf{x} \in D_{G'}} \alpha_{G'} f_{G'} \\ &= \sum_{\pi} \prod_{b \in \pi} \tilde{K}_{|b|}, \end{aligned} \tag{3.7.11}$$

where π runs over all set-partitions of the set $\{\underline{1}, \ldots, \underline{k}\}$, and b runs over all blocks of π. Above $\tilde{K}_j$ is defined as

$$\begin{aligned} \tilde{K}_j :&= \sum_{G \in C_j^{\mathrm{mark}}} \sum_{\mathbf{x} \in D_G} \alpha_G f_G \\ &= \frac{1}{j} \sum_{G \in C_j^{\mathrm{mark}}} \sum_{\mathbf{x} \in D_G} \left(\prod_{v \in V_G} \mu_\Lambda(x_v) \right) (-1)^{|B_G|-1} f_G. \end{aligned} \tag{3.7.12}$$

In our setting the moment-cumulant formula (3.3.1) takes the form

$$M_k = \sum_{\pi \in \Pi_k} \prod_{b \in \pi} K_{|b|}.$$

(Here π still runs over set-partitions, and b runs over all blocks of π.) It can be viewed as a system of algebraic equations for the unknowns (K_k). This system of equations has an upper-triangular form in the sense that the k-th equation allows us to express the cumulant K_k as the sum of the moment M_k and some complicated polynomial in the variables $K_1, \ldots, K_{k-1}$. Such a system of equations can be solved recursively and clearly has a unique equation. Equation (3.7.11) shows that the sequence $(\tilde{K}_k)$ is a solution of this system of equations; since the solution is unique, it follows that the cumulant

$$K_k = \tilde{K}_k$$

is given by (3.7.12) after the substitution $j = k$.

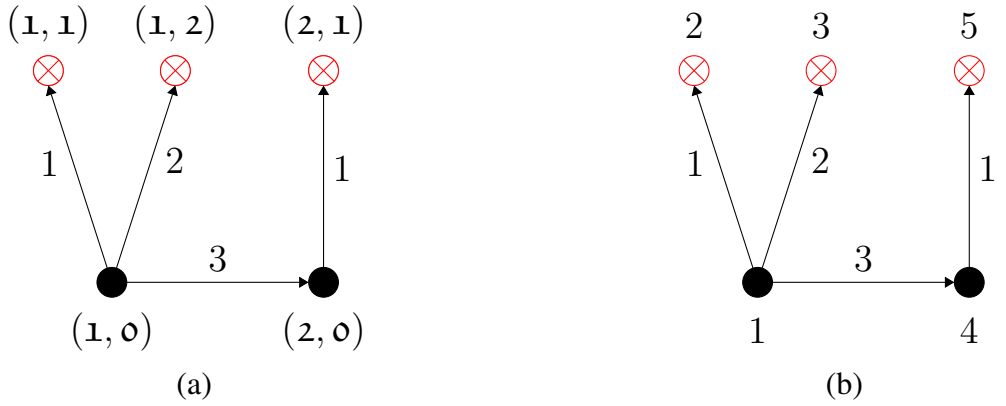


Figure 3.20: (a) A caterpillar graph with tagged vertices. (b) The same graph with labeled vertices. The labels belong to the set $\{1, 2, 3, 4, 5\}$.

3.7.9 Caterpillar graphs with labeled vertices

We say that a connected, weighted, oriented graph G is a *caterpillar graph with k labeled vertices* if its vertex set is equal to $\{1, \ldots, k\}$, the weights of the edges fulfill the convention (3.7.3), and there exists some way of tagging the vertices of G in such a way that G becomes an element of $\mathrm{MC}_k^{\mathrm{tag}}$, see Section 3.7.3. An example of a caterpillar graph with labeled vertices is given in Figure 3.20b. The set of caterpillar graphs with k labeled vertices will be denoted by C_k^{lab}. This definition may sound a bit abstract so we provide an alternative description below.

Note that for any connected of a graph $G \in \mathrm{MC}_k^{\mathrm{tag}}$ there is a unique way of labeling the vertices so that the requirement (3.7.3) is fulfilled, given as follows. We start by assigning the number 1 to the unique black vertex with no incoming edges. Then, in the order given by the weights of the edges, we number all endpoints of the edges outgoing from this vertex number 1 by successive natural numbers. We repeat the process at the unique black endpoint of an edge outgoing from 1, and continue until we visit all black vertices. In this way for any edge $e = (v_1, v_2)$ its weight is equal to the difference of the labels of the endpoints: $w(e) = v_2 - v_1$. The outcome is clearly an element of C_k^{lab}, and each element of C_k^{lab} can be obtained in this way.

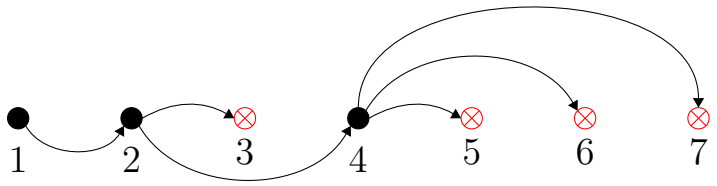


Figure 3.21: An example of a caterpillar graph with $k = 7$ labeled vertices. With the notations of (3.7.13) we have $\ell = 3$ black vertices and $b_1 = 1$, $b_2 = 2$, $b_3 = 4$; additionally we use the convention that $b_4 = 8$. The weights of the edges were not shown.

The above procedure shows that the elements of C_k^{lab} can be characterized as follows. For each $G \in C_k^{\mathrm{lab}}$ with the set of black vertices

$$B_G = \{b_1, \dots, b_\ell\} \subseteq \{1, \dots, k\}, \qquad b_1 < \cdots < b_\ell \tag{3.7.13}$$

we have that $\ell \geq 1$ and $b_1 = 1$. We will use the convention that $b_{\ell+1} = k + 1$. The black vertices are connected by a series of oriented edges

$$(b_1, b_2), \quad (b_2, b_3), \quad \dots, \quad (b_{\ell-1}, b_\ell).$$

Additionally, each black vertex b_i (with $i \in \{1, \dots, \ell\}$) is connected to the red vertices $b_i + 1, b_i + 2, \dots, b_{i+1} - 1$ which immediate follow it by a collection of oriented edges

$$(b_i, b_i + 1), \quad (b_i, b_i + 2), \quad \dots, \quad (b_i, b_{i+1} - 1),$$

see Figure 3.21 for an example.

In particular, since the structure of a caterpillar graph with labeled vertices is determined by its set of black vertices, it follows that $|C_k^{\mathrm{lab}}| = 2^{k-1}$.

3.7.10 The second formula for the cumulants

We continue the discussion from Section 3.7.8 and revisit the formula (3.7.12) for the cumulant K_k. As we already mentioned, the connected graph $G \in C_k^{\mathrm{mark}}$ can be labeled in a unique way so that it becomes an element of C_k^{lab}. On the other hand, for each graph $G \in C_k^{\mathrm{lab}}$, there exist $k!$ ways to mark the vertices so that the outcome is a caterpillar graph with k marked vertices. In this way we proved the following intermediate result.

Figure 3.22: (a) Caterpillar graph with one black and one red vertices. (b) Caterpillar graph with two black vertices.

Corollary 3.7.4. *Let Λ be an interlacing sequence with a generic set of concave corners. For each $u_0 \in \mathbb{R}$ the n-th formal cumulant considered in Section 3.6.5 is given by the following sum over caterpillar graphs with k labeled vertices*

$$K_k = (k-1)! \sum_{G \in C_k^{\mathrm{lab}}} \sum_{\mathbf{x} \in D_G} (-1)^{|B_G|-1} f_G \prod_{j \in \{1,\dots,k\}} \mu_\Lambda(x_j). \tag{3.7.14}$$

For example, for $k = 2$ we obtain

$$K_2 = \sum_{x_1 \le u_0} \sum_{x_2} \frac{\mu_\Lambda(x_1)\mu_\Lambda(x_2)}{x_2 - x_1 + 1} - \sum_{x_1 \le u_0} \sum_{x_2 \le u_0} \frac{\mu_\Lambda(x_1)\mu_\Lambda(x_2)}{x_2 - x_1 + 1}.$$

The first summand corresponds to the caterpillar graph shown of Figure 3.22a, and the second summand corresponds to the caterpillar graph shown of Figure 3.22b.

3.7.11 Sum over non-crossing alternating trees

Proposition 3.7.5. *Let Λ be an interlacing sequence with a generic set of concave corners. For each $u_0 \in \mathbb{R}$ the n-th formal cumulant considered in Section 3.6.5 is given by the following sum over noncrossing alternating trees*

$$K_k = (k-1)! \sum_{H \in \mathbb{T}_k} \sum_{\mathbf{x} \in D_H} (-1)^{|B_H|-1} f_H \prod_{j \in \{1,\dots,k\}} \mu_\Lambda(x_j). \tag{3.7.15}$$

Proof. In (3.7.14) we can reverse the order of the sums and write

$$K_k = -(k-1)! \sum_{x_1,\dots,x_k \in \mathbb{X}} \mathfrak{C}_k(x_1,\dots,x_k) \prod_{j \in \{1,\dots,k\}} \mu_\Lambda(x_j),$$

where

$$\mathfrak{C}_k(x_1,\dots,x_k) := \sum_{\substack{G\in C_k^{\mathrm{lab}}\\(x_1,\dots,x_k)\in D_G}} (-1)^{|B_G|} f_G(x_1,\dots,x_k).$$

Similarly, the right-hand side of (3.7.15) can be written as

$$-(k-1)! \sum_{x_1,\dots,x_k\in\mathbb{X}} \mathfrak{T}_k(x_1,\dots,x_k) \prod_{j\in\{1,\dots,k\}} \mu_\Lambda(x_j),$$

where

$$\mathfrak{T}_k(x_1,\dots,x_k) := \sum_{\substack{H\in\mathbb{T}_k\\(x_1,\dots,x_k)\in D_H}} (-1)^{|B_H|} f_H(x_1,\dots,x_k).$$

As a side remark note that $\mathfrak{T}_k(x_1,\dots,x_k)$ is a quantity which (up to a scaling factor) is closely related to the random variable Z from Remark 3.3.4. The result is a consequence of Lemma 3.7.6 below. □

Lemma 3.7.6. *With the above notations,*

$$\mathfrak{C}_k(x_1,\dots,x_k) = \mathfrak{T}_k(x_1,\dots,x_k)$$

holds true for any $k \geq 1$ and any $x_1,\dots,x_k \in \mathbb{R}$ for which the left-hand side of the equality does not involve division by zero.

Proof. In the special case $k = 1$ we have that the set of graphs $C_1^{\mathrm{lab}} = \mathbb{T}_1$ which contributes to $\mathfrak{C}_1(x_1)$, respectively to $\mathfrak{T}_1(x_1)$, consists of a single element depicted in Figure 3.12a. Thus

$$\mathfrak{C}_1(x_1) = \mathfrak{T}_1(x_1) = \begin{cases} -1 & \text{if } x_1 \leq u_0, \\ 0 & \text{if } x_1 > u_0. \end{cases}$$

Let $k \geq 2$. In the case when $x_k \leq u_0$ we obtain that

$$\mathfrak{T}_{x_1,\dots,x_k} = 0$$

because the rightmost vertex of any non-crossing alternating tree $H \in \mathbb{T}_k$ is white thus $(x_1,\dots,x_k)$ is not a decoration of H and the sum runs over the empty set.

Let $C^{\mathrm{lab}}(V)$ and $\mathbb{T}(V)$ denote, respectively, the set of caterpillar graphs and the set of non-crossing alternating trees such that their vertex set is equal to V.

In particular $C^{\mathrm{lab}}\big(\{1,\dots,k\}\big) = C_k^{\mathrm{lab}}$ and $\mathbb{T}\big(\{1,\dots,k\}\big) = \mathbb{T}_k$ for any natural number k.

Let $H \in \mathbb{T}_k$ be a non-crossing alternating tree with k vertices. Obviously H contains the edge $e = (1,k)$ connecting the leftmost and the rightmost vertex. After removing the edge e from the graph H, the resulting graph $H \setminus e$ has two connected components: $H_1 \in \mathbb{T}_{i-1}\big(\{1,\dots,i-1\}\big)$ and $H_2 \in \mathbb{T}_{k-i}\big(\{i,\dots,k\}\big)$ for some $i \in \{2,\dots,k\}$. In the special case when the graph H_2 consists of a single (white) vertex, we change its color to black. This way, each non-crossing alternating tree with the vertex set $\{1,\dots,k\}$ decomposes uniquely into a sum of the edge e and two non-crossing alternating trees H_1 with the vertex set $\{1,\dots,i-1\}$ and H_2 with the vertex set $\{i,\dots,k\}$.

If $i \neq k$ then $(x_1,\dots,x_k)$ is a decoration of H if and only if the prefix $(x_1,\dots,x_{i-1})$ is a decoration of H_1 and the suffix $(x_i,\dots,x_k)$ is a decoration of H_2; the special case when $i = k$ and the tree H_2 consists of a single vertex has to be considered separately. Therefore we obtain the following recurrence relation:

$$\mathfrak{T}_k(x_1,\dots,x_k) = \begin{cases} 0 & \text{if } x_k \le u_0, \\ \sum\limits_{i=2}^{k-1} \frac{\mathfrak{T}_{i-1}(x_1,\dots,x_{i-1})\mathfrak{T}_{k-i+1}(x_i,\dots,x_k)}{x_k - x_1 + k - 1} & \\ \qquad + \frac{\mathfrak{T}_{k-1}(x_1,\dots,x_{k-1})}{x_k - x_1 + k - 1} & \text{if } x_k > u_0. \end{cases}$$

We will prove that the sequence of functions $\mathfrak{C}_k$ satisfies the same recurrence relation.

If $k \ge 2$ and $x_k \le u_0$ then $\mathfrak{C}(x_1,\dots,x_k) = 0$ because we can pair caterpillar graphs from C_k^{lab} into pairs that differ only in the color of the far-right vertex, and the contribution of each pair to the sum is zero.

Let $k \ge 2$ be a natural number, let $x_k > u_0$ and let $G \in C_k^{\mathrm{lab}}$ be a caterpillar graph. There is a unique path $(i_1,\dots,i_t)$ with $t \ge 2$ from the vertex 1 to the vertex k which means that $i_1 = 1$ and $i_t = k$, and $(i_1,i_2),(i_2,i_3),\dots,(i_{t-1},i_t) \in E_G$. In the special case when $e = (1,k) \in E_G$, we have $t = 2$ and $e_1 = e$. Using the telescopic sum

$$x_k - x_1 + k - 1 = \sum_{j=1}^{t-1} \big(x_{i_{j+1}} - x_{i_j} + i_{j+1} - i_j\big)$$

we obtain

$$f_G(x_1,\dots,x_k) = \sum_{j=1}^{t-1} \frac{x_{i_{j+1}} - x_{i_j} + i_{j+1} - i_j}{x_k - x_1 + k - 1} f_G(x_1,\dots,x_k)$$
$$= \sum_{j=1}^{t-1} \frac{f_{G\setminus e_j}(x_1,\dots,x_k)}{x_k - x_1 + k - 1},$$

where $G \setminus e_j$ denotes the graph G with the edge e_j removed. Therefore,

$$\mathfrak{C}(x_1,\dots,x_k) = \sum_{\substack{G\in C_k^{\mathrm{lab}} \\ (x_1,\dots,x_k)\in D_G}} (-1)^{|B_G|} \sum_{j=1}^{t-1} \frac{f_{G\setminus e_j}(x_1,\dots,x_k)}{x_k - x_1 + k - 1}.$$

In addition, every graph $G \in C_k^{\mathrm{lab}}$ after removing any edge e_j splits in a unique way into the sum of two caterpillar graphs $G_1 \in C_{i-1}^{\mathrm{lab}}\left(\{1,\dots,i-1\}\right)$ and $G_2 \in C_{k-i+1}^{\mathrm{lab}}\left(\{i,\dots,k\}\right)$ for some $i \in \{2,\dots,k\}$. In the special case when $i = k$ and the graph G_2 consists a single (red) vertex, we change its color to black; this case will require separate analysis. In this way we can write $\mathfrak{C}(x_1,\dots,x_k)$ as a triple sum over all possible choices of the number i, over all graphs
$G_1 \in C_{i-1}^{\mathrm{lab}}\left(\{1,\dots,i-1\}\right)$ and over all graphs $G_2 \in C_{k-i+1}^{\mathrm{lab}}\left(\{i,\dots,k\}\right)$, i.e.,

$$\mathfrak{C}(x_1,\dots,x_k) =$$
$$\sum_{i=2}^{k-1} \sum_{\substack{G_1\in C_{i-1}^{\mathrm{lab}}\left(\{1,\dots,i-1\}\right) \\ (x_1,\dots,x_{i-1})\in D_{G_1}}} \sum_{\substack{G_2\in C_{k-i+1}^{\mathrm{lab}}\left(\{i,\dots,k\}\right) \\ (x_i,\dots,x_k)\in D_{G_2}}} \frac{(-1)^{|B_{G_1}|+|B_{G_2}|} f_{G_1} f_{G_2}}{x_k - x_1 + k - 1}$$
$$+ \sum_{\substack{G_1\in C_{i-1}^{\mathrm{lab}}\left(\{1,\dots,k-1\}\right) \\ B_{G_1}\subseteq B}} \frac{(-1)^{|B_{G_1}|} f_{G_1}}{x_k - x_1 + k - 1}$$
$$= \sum_{i=2}^{k-1} \frac{\mathfrak{C}_{i-1}(x_1,\dots,x_{i-1})\mathfrak{C}_{k-i+1}(x_i,\dots,x_k)}{x_k - x_1 + k - 1} + \frac{\mathfrak{C}_{k-1}(x_1,\dots,x_{k-1})}{x_k - x_1 + k - 1}$$

if $x_k > u_0$.

The sequences of rational functions $\mathfrak{C}_k$ and $\mathfrak{T}_k$ satisfy the same recurrence relation and have the same initial condition, which completes the proof. □

3.7.12 Proof of Theorem 3.3.2

Proof of Theorem 3.3.2. When the number u_0 is not an integer, we apply Lemma 3.6.2 and evaluate the cumulant $K_k(\Lambda^\epsilon, u_0)$ using Proposition 3.7.5.

However, when the number u_0 is an integer, the theorem is satisfied for any number $u \in (u_0, u_0 + 1)$, as shown above. Since $F_T(u_0)$ is a right-continuous function, then

$$F_T(u_0) = \lim_{u \to u_0} F_T(u),$$

and Theorem 3.3.2 also holds for the number u_0. □

Bibliography

[BDJ99] Jinho Baik, Percy Deift, and Kurt Johansson. "On the distribution of the length of the longest increasing subsequence of random permutations". In: *J. Amer. Math. Soc.* 12 (1999), pp. 1119–1178.

[Bia03] Philippe Biane. "Characters of symmetric groups and free cumulants". In: *Asymptotic combinatorics with applications to mathematical physics (St. Petersburg, 2001)*. Vol. 1815. Lecture Notes in Math. Springer, Berlin, 2003, pp. 185–200. DOI: 10.1007/3-540-44890-X_8.

[Bia98] Philippe Biane. "Representations of symmetric groups and free probability". In: *Adv. Math.* 138.1 (1998), pp. 126–181. DOI: 10.1006/aima.1998.1745.

[BS07] Leonid V. Bogachev and Zhonggen Su. "Gaussian fluctuations of Young diagrams under the Plancherel measure". In: *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 463.2080 (2007), pp. 1069–1080. DOI: 10.1098/rspa.2006.1808.

[DFŚ10] Maciej Dołęga, Valentin Féray, and Piotr Śniady. "Explicit combinatorial interpretation of Kerov character polynomials as numbers of permutation factorizations". In: *Adv. Math.* 225.1 (2010), pp. 81–120. DOI: 10.1016/j.aim.2010.02.011.

[Dur19] Rick Durrett. *Probability—theory and examples*. Vol. 49. Cambridge Series in Statistical and Probabilistic Mathematics. Fifth edition of [MR1068527]. Cambridge University Press, Cambridge, 2019, pp. xii+419. DOI: 10.1017/9781108591034.

[Duz19] Vasilii Duzhin. "Investigation of insertion tableau evolution in the Robinson-Schensted-Knuth correspondence". In: *Discrete and Continuous Models and Applied Computational Science* 27 (Dec. 2019),

pp. 316–324. DOI: 10.22363/2658-4670-2019-27-4-316-324.

[Fér09] Valentin Féray. "Combinatorial interpretation and positivity of Kerov's character polynomials". In: *J. Algebraic Combin.* 29.4 (2009), pp. 473–507. DOI: 10.1007/s10801-008-0147-y.

[FH91] William Fulton and Joe Harris. *Representation theory*. Vol. 129. Graduate Texts in Mathematics. A first course, Readings in Mathematics. Springer-Verlag, New York, 1991, pp. xvi+551. DOI: 10.1007/978-1-4612-0979-9.

[Ful97] William Fulton. *Young Tableaux: With Applications to Representation theory and Geometry*. Vol. 35. London Mathematical Society Student Texts. Cambridge: Cambridge University Press, 1997, pp. x+260.

[GGP97] Israel M. Gelfand, Mark I. Graev, and Alexander Postnikov. "Combinatorics of hypergeometric functions associated with positive roots". In: *The Arnold-Gelfand mathematical seminars*. Birkhäuser Boston, Boston, MA, 1997, pp. 205–221. DOI: 10.1007/978-1-4612-4122-5_10.

[GR07] Ian P. Goulden and Amarpreet Rattan. "An explicit form for Kerov's character polynomials". In: *Trans. Amer. Math. Soc.* 359.8 (2007), pp. 3669–3685. DOI: 10.1090/S0002-9947-07-04311-5.

[GR19] Vadim Gorin and Mustazee Rahman. "Random sorting networks: local statistics via random matrix laws". In: *Probab. Theory Related Fields* 175.1-2 (2019), pp. 45–96. DOI: 10.1007/s00440-018-0886-1.

[Ker00] Sergei Kerov. *Talk in Institute Henri Poincaré, Paris*. Jan. 2000.

[Ker93] Sergei Kerov. "Transition probabilities for continual Young diagrams and the Markov moment problem." In: *Funct. Anal. Appl.* 27 (1993), pp. 104–117.

[KV86a] Sergei V. Kerov and Anatoly M. Vershik. "The characters of the infinite symmetric group and probability properties of the Robinson-Schensted-Knuth algorithm". In: *SIAM J. Algebraic Discrete Methods* 7.1 (1986), pp. 116–124. DOI: 10.1137/0607014.

[KV86b] Sergei V. Kerov and Anatoly M. Vershik. “The characters of the infinite symmetric group and probability properties of the Robinson-Schensted-Knuth algorithm”. In: *SIAM J. on Algebraic and Discrete Methods* 7.1 (1986), pp. 116–124. DOI: 10.1137/0607014.

[LH02] Steffen L. Lauritzen and A. Hald. *Thiele: pioneer in statistics*. Thiele’s papers translated from the Danish by Steffen L. Lauritzen, With appreciations of Thiele’s work by Lauritzen and A. Hald. Oxford University Press, New York, 2002, pp. viii+264. DOI: 10.1093/acprof:oso/9780198509721.001.0001.

[LS77] Benjamin F. Logan and Lawrence A. Shepp. “A variational problem for random Young tableaux”. In: *Advances in Math.* 26.2 (1977), pp. 206–222.

[Mar22a] Mikołaj Marciniak. “Hydrodynamic limit of the Robinson–Schensted–Knuth algorithm”. In: *Random Structures & Algorithms* 60.1 (2022), pp. 106–116. DOI: https://doi.org/10.1002/rsa.21016. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.21016.

[Mar22b] Mikołaj Marciniak. “Quadratic Coefficients of Goulden–Rattan Character Polynomials”. In: *Annals of Combinatorics* (Dec. 2022). DOI: https://doi.org/10.1007/s00026-022-00611-5. eprint: https://link.springer.com/content/pdf/10.1007/s00026-022-00611-5.pdf.

[Oko00] Andrei Okounkov. “Random matrices and random permutations”. In: *Internat. Math. Res. Notices* 20 (2000), pp. 1043–1095. DOI: 10.1155/S1073792800000532.

[PR07] Boris Pittel and Dan Romik. “Limit shapes for random square Young tableaux”. In: *Adv. in Appl. Math.* 38.2 (2007), pp. 164–209. DOI: 10.1016/j.aam.2005.12.005.

[Rom15] Dan Romik. *The surprising mathematics of longest increasing subsequences*. Vol. 4. Institute of Mathematical Statistics Textbooks. Cambridge University Press, New York, 2015, pp. xi+353.

[RŚ15] Dan Romik and Piotr Śniady. “Jeu de taquin dynamics on infinite Young tableaux and second class particles”. In: *Ann. Probab.* 43.2 (2015), pp. 682–737. DOI: 10.1214/13-AOP873.

[RŚ16] Dan Romik and Piotr Śniady. “Limit shapes of bumping routes in the Robinson-Schensted correspondence”. In: *Random Structures Algorithms* 48.1 (2016), pp. 171–182. DOI: 10.1002/rsa.20570.

[Sch59] Samuel Schechter. “On the Inversion of Certain Matrices”. In: *Mathematics of Computation - Math. Comput.* 13 (Apr. 1959), pp. 73–73. DOI: 10.1090/S0025-5718-1959-0105798-2.

[Sch63] Marcel P. Schützenberger. “Quelques remarques sur une construction de Schensted”. In: *Math. Scand.* 12 (1963), pp. 117–128.

[Śni06] Piotr Śniady. “Asymptotics of characters of symmetric groups, genus expansion and free probability”. In: *Discrete Math.* 306.7 (2006), pp. 624–665. DOI: 10.1016/j.disc.2006.02.004.

[Śni19] Piotr Śniady. “Asymptotics of Jack characters”. In: *J. Combin. Theory Ser. A* 166 (2019), pp. 91–143. DOI: 10.1016/j.jcta.2019.02.020.

[Śni20] Piotr Śniady. “Conjectures related to Robinson–Schensted–Knuth algorithm applied to a random input”. Unpublished handwritten manuscript. 2020.

[Sta99] Richard P. Stanley. *Enumerative combinatorics. Vol. 2*. Vol. 62. Cambridge Studies in Advanced Mathematics. Cambridge: Cambridge University Press, 1999, pp. xii+581. DOI: 10.1017/CBO9780511609589.

[Tho64] Elmar Thoma. “Die unzerlegbaren, positiv-definiten Klassenfunktionen der abzählbar unendlichen, symmetrischen Gruppe”. In: *Math. Z.* 85 (1964), pp. 40–61. DOI: 10.1007/BF01114877.

[VDN92] Dan V. Voiculescu, Kenneth J. Dykema, and Alexandru Nica. *Free random variables*. Vol. 1. CRM Monograph Series. A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups. American Mathematical Society, Providence, RI, 1992, pp. vi+70. DOI: 10.1090/crmm/001.

[VK77] Anatoly M. Vershik and Sergei V. Kerov. “Asymptotics of the Plancherel measure of the symmetric group and the limit form of Young tableaux”. In: *Soviet Math. Dokl.* 18 (1977), pp. 527–531.

[VK81] Anatoly M. Vershik and Sergei V. Kerov. “Asymptotic theory of the characters of a symmetric group”. In: *Funktsional. Anal. i Prilozhen.* 15.4 (1981), pp. 15–27, 96.

[Voi91] Dan Voiculescu. “Limit laws for random matrices and free products”. In: *Invent. Math.* 104.1 (1991), pp. 201–220. DOI: 10.1007/BF01245072.

[Voi94] Dan Voiculescu. “The analogues of entropy and of Fisher’s information measure in free probability theory. II”. In: *Invent. Math.* 118.3 (1994), pp. 411–440. DOI: 10.1007/BF01231539.

[Woj19] Karolina Wojtyniak. “Asymptotyka losowych trajektorii jeu de taquin”. Bachelor’s Thesis. 2019.