

# Notes on Tracy-Widom Fluctuation Theory

Nathanaël Berestycki

May 7, 2007

In these notes we discuss a few probability models where the fluctuations of a certain random variable are governed not by the classical central limit theorem, but by a distribution called the Tracy-Widom distribution. This random variable is closely connected to determinantal processes: it turns out that it is distributed as the largest point of a determinantal point process on the real line where the kernel  $K$  is the so-called *Airy kernel*.

This kind of fluctuations arises (or is believed to arise) in a surprising variety of models: eigenvalues of random matrices, longest increasing subsequences of random permutations, shape fluctuations in first and last passage percolation, polynuclear growth models, frozen region of a random domino tiling of the aztec diamond, totally asymmetric exclusion process... And the list doesn't stop here. For this reason, physicists say that these models belong to the same *universality class*.

In this document we will try to give an outline of the proof in perhaps the simplest case: random permutations. However the ideas and methods carry over to other cases mentioned above.

**References.** The proof of Tracy-Widom fluctuations for random permutation was first given by Baik, Deift and Johansson [2]. Shortly after, Johansson [7] proved (using similar techniques as the ones below) that this was also the case for the shape fluctuations in last passage percolation with geometric and exponential weight distributions in 2 dimensions. I recommend the notes of Johansson [8], very accessible to probabilists, and the very readable article of Aldous and Diaconis [1], which has much material on the history and the mathematics of longest increasing subsequences for random permutations. Finally, the connection to Young tableaux plays an essential role here. For this part, the book of Sagan [13] is very clear and to the point.

**Acknowledgement.** Thanks to Sunil Chhita, who proofread the document, and together with whom I learnt this stuff week after week.

# 1 Longest Increasing Subsequences

Let  $n \geq 1$  and consider  $S_n$  the symmetric group on  $n$  elements. Let  $\sigma$  be a random permutation. An increasing subsequence of  $\sigma$  is a set  $x_1 < \dots < x_r$  such that  $\sigma(x_i) < \sigma(x_{i+1})$  for all  $1 \leq i \leq r - 1$ . Let  $L_n$  be the length of longest increasing subsequence (LIS). For instance, if

$$\sigma = \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 2 & 3 & 1 & 6 & 7 & 5 \end{array}$$

Then the longest increasing subsequence of  $\sigma$  is  $\{2, 3, 6, 7\}$  and has length  $L_n = 4$ .

**A geometric model.** Another geometric way to think of  $L_n$  is to drop  $n$  points  $U_1, \dots, U_n$  uniformly in the unit square. Consider piecewise linear paths from  $(0,0)$  to  $(1,1)$  (with angles only at some  $U_i$ ) such that the path goes always up and right. The maximal number of  $U_i$ 's on the path has the same distribution as  $L_n$ . See the following picture for an illustration.

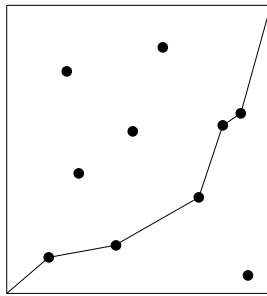


Figure 1: The geometrical model for  $n = 10$ . Here  $L_n = 5$ .

What is the asymptotic behavior of  $L_n$ ? There is a long history here. By a combination of efforts due to Hammersley [4] in the early 70's (subadditive arguments), Logan and Shepp [12] in the late 1970's, and independently Vershik and Kerov [10], it was first proved that

$$n^{-1/2}L_n \rightarrow 2 \tag{1}$$

in probability. This should be regarded as a law of large numbers for  $L_n$ .

**Why is this the right order of magnitude?** We present here an argument (I presume this is the same as the one used by Hammersley [4]) to show that  $\sqrt{n}$  is the order of magnitude of  $L_n$ . Consider the geometric representation. First,  $L_n$  should not change very much if we make  $n$  a Poisson random variable with mean

$n$ . I.e., we consider a Poisson process of points in the unit square with intensity  $n$  times Lebesgue measure. It is useful to blow up the picture and consider a Poisson process of points in the plane with unit Lebesgue intensity. Let  $c_L$  be the expected maximum number of points that can fit on an up-right path in a square of side-length  $L$ . Hence, we are interested in  $c_{\sqrt{n}}$ . Now, it is obvious that

$$c_{L+L'} \geq c_L + c_{L'} \tag{2}$$

because we can combine two optimal paths in a square of length  $L$  and a square of side length  $L'$  to get an admissible path in a square of side-length  $L + L'$  (attach the two squares so that their side are parallels and their only common point is the top-right corner of the bottom square). By (2),  $c_L$  is a super-additive function and we conclude that there exists  $\gamma > 0$  such that  $c_L/L \rightarrow \gamma$ . Specializing to  $L = \sqrt{n}$ , we conclude  $E(L_n) \approx \gamma\sqrt{n}$ . As with any subadditivity argument, obtaining the value of the constant  $\gamma$  is much harder than proving its existence. This is the subject of the work of Logan and Shepp [12] and of Vershik and Kerov [10].

**Fluctuations.** The next step is to study fluctuations. We are used to fluctuations being typically the square root of the expectations and the fluctuations to be asymptotically normal. The truth is far from it.

First, it was conjectured in the early 90s (by Kesten, inspired by analogous conjectures in first passage percolation) that

$$\text{var}(L_n) = \Theta(n^{1/3})$$

It was not until 1999, nearly 30 years after the first serious efforts on the problem, that the following astonishing result was proved by Baik, Deift, and Johansson [2].

**Theorem 1.** *For any  $\xi \in \mathbf{R}$ ,*

$$P\left(\frac{L_n - 2\sqrt{n}}{n^{1/6}} \leq \xi\right) \rightarrow F(\xi)$$

where  $F(\cdot)$  is the distribution function of a Tracy-Widom random variable.

It is time to define this famous Tracy-Widom distribution.

## 2 Airy process and Tracy-Widom distribution

As you will see, the definition of the distribution is not very explicit at first sight. Let  $u$  be the solution of the Painlevé II equation:

$$u''(x) = 2u^3 + xu$$

and  $u(x) \sim \text{Ai}(x)$  as  $x \rightarrow \infty$ . Here,  $\text{Ai}(x)$  denotes the Airy function:

$$\text{Ai}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i(t+is)^3/3+ix(t+is)} dt$$

where  $s > 0$  is arbitrary (it is easy to see that the choice of  $s$  doesn't matter in this definition). Then the analytic definition of the distribution function  $F_{TW}$  of a Tracy-Widom is: for any  $s \in \mathbf{R}$ ,

$$F_{TW}(s) = \exp\left(-\int_s^{\infty} (x-s)u(x)^2 dx\right). \quad (3)$$

There seems to be very little intuition to get from such a definition. I find this one a little more useful. Let

$$A(x, y) = \frac{\text{Ai}(x)\text{Ai}'(y) - \text{Ai}(y)\text{Ai}'(x)}{x - y}$$

be the *Airy kernel*. It can be checked that

$$A(x, y) = \int_0^{\infty} \text{Ai}(x+t)\text{Ai}(y+t) dt \quad (4)$$

The Airy kernel defines a determinantal point process  $\Xi$  on  $\mathbf{R}$  which is trace class, called the *Airy process*. Let  $X$  be the position of the rightmost point of  $\Xi$ . Then  $\Xi$  has the Tracy-Widom distribution (3). Thus a simple definition of the Tracy Widom distribution is:

**Lemma 1.** *We have the formula:*

$$F_{TW}(s) = \det(I - A)_{L^2(s, \infty)}. \quad (5)$$

*Proof.* The proof consists essentially in understanding the formula. What does it mean?  $I - A$  is seen as an infinite-dimensional integral operator on  $L^2(s, \infty)$ . In this context, the determinant can only mean one thing: the product of its eigenvalues. Let  $\lambda_1, \dots$ , be the eigenvalues of  $A$  as an operator in  $L^2(s, \infty)$ . Then, applying the Bernoulli representation of HKPV [5],

$$\begin{aligned} F_{TW}(s) &= P(X \leq s) \\ &= P(\text{no point in } (s, \infty)) \\ &= \prod_{i=1}^{\infty} (1 - \lambda_i) \\ &= \det(I - A)_{L^2(s, \infty)} \end{aligned}$$

as required. □

### 3 Proof of determinantal structure for LIS

In these notes we will present arguments that make the connection between the longest increasing subsequence of a random permutation and a certain determinantal process. What will be left out of these notes, is the computation that the kernel of the point process, suitably rescaled, converges to the Airy kernel. This seems to involve arid and lengthy calculations, about which I don't feel so comfortable at the moment. Instead, let's see how far we go.

#### 3.1 A sketch of the proof

First, here is a summary of the order in which the arguments unfold.

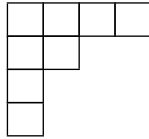
1. The RSK correspondence maps permutations to Young diagrams. The longest increasing subsequence is equal to the length of the first row of the diagram.
2. The distribution on Young diagrams that one gets by applying the RSK correspondence to a random uniform permutation, is the Plancherel measure. Hence, we need to understand the length of the first row under the Plancherel measure.
3. The Plancherel measure can be analyzed by studying a certain set of nonintersecting paths on a directed graph, and the length of the first row is equal to the highest crossing by these nonintersecting paths of a fixed vertical line.
4. Applying the LGV theorem, (which is a generalization of the classical Karlin-McGregor argument for computing the probability of nonintersection of several simple random walk paths), we obtain a determinantal form for the "weight" of nonintersecting paths. This leads to a simple relation between the Plancherel and the Schur measures.
5. A few (simple) computations show that the Schur measure is determinantal. This implies a similar result for the heights of the nonintersecting paths on a fixed vertical line. The length of the first row is the highest of these points.
6. The kernel, suitably rescaled, converges to the Airy kernel. It follows that the determinantal process converges to the Airy process. In particular, its largest point converges to the Tracy-Widom distribution.

As mentioned above, the difficult computations involved with the last part of this strategy will not be included in this document.

### 3.2 The RSK correspondence

The RSK correspondence is one of the most useful tools to study representation theory of the symmetric group. We start by recalling some basic definitions about Young diagrams and Young tableaux.

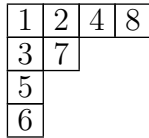
Let  $n \geq 1$ . A **partition** of  $n$  is a sequence of nonnegative integers  $\lambda = (\lambda_1, \dots, \lambda_r)$  such that  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$  and  $\sum_i \lambda_i = n$ . A **Young diagram** of size  $n$  is the shape that one obtains when we represent this partition by thinking of a number  $\lambda_i$  as a row of  $\lambda_i$  boxes of size  $1 \times 1$ . If we stack these rows on top of one another, we get a total of  $n$  boxes forming a shape that looks like this.



This is the Young diagram associated with the partition  $\lambda = (4, 2, 1, 1)$  with  $n = 8$ .

A **Young tableaux** is a particular Young diagram whose boxes have been filled by numbers from 1 to  $n$  in such a way that: on any given row, the numbers are increasing from left to right, and similarly, on any given column, the sequence of numbers is increasing from top to bottom.

Eg:



**The Robinson-Schensted-Knuth Algorithm.** The RSK correspondence is a bijective way to associate a pair of Young tableaux of the same shape to a permutation  $\sigma$ . In fact, it was first discovered by Robinson and independently by Schensted for permutation and Young tableaux, and later it was generalized by Knuth to generalized permutations and semi-standard Young tableaux. This generalized correspondence is what is useful in the context of last passage percolation. Following the excellent book of Sagan [13] Chapter 3, we denote by

$$\sigma \xleftrightarrow{\text{R-S}} (P, Q)$$

this correspondence. We start by the direction  $\xrightarrow{\text{R-S}}$ . We construct recursively a pair of partial tableaux  $(P_0, Q_0), \dots, (P_n, Q_n)$  such that  $P_k$  and  $Q_k$  have the same shape for all  $k$ . At every step we increase the size of the tableaux by one. Initially,  $P_0 = Q_0 = \emptyset$ . Now, suppose we are given a partial tableau  $P$ , and an element  $x$  which is not listed in  $P$ . The following describes how to **insert**  $x$ .

RS1 Set  $R$  to be the first row of  $P$ .

RS2 While  $x$  is less than some element of  $R$ , do:

RS2a Let  $y$  be the smallest element of  $R$  greater than  $x$ , and replace  $y$  by  $x$  in  $R$ .

RS2b Set  $x := y$  and  $R :=$  the next row down.

RS3 Now  $x$  is greater than any element of  $R$ , so we place  $x$  at the end of  $R$  and we are done.

This seems complicated but is in fact quite easy to do in practice. We try to insert an  $x$  in the tableau and this provokes a cascade. We put it as early as soon as we can so as not to violate the requirement that it is a tableau (increasing along rows and columns), and this pushes down one element to the next row, and so on.

To illustrate, suppose  $P$  is given by

1	2	5	8
4	7		
6			
9			

and we wish to insert  $x = 3$ . Then 3 can be inserted in the first row, bumping the 5. The first row now becomes 1 2 3 8, and we now have to row-insert  $x = 5$  in the three lowest rows of  $P$ . Then 5 can be inserted in the second row, bumping the 7. 7 itself can be inserted in the third row, bumping no one.

Hence the result of inserting  $x$  into  $P$  is the partial tableau  $P'$

1	2	3	8
4	5		
6	7		
9			

We note  $P' = r_x(P)$ . It is trivial to check that starting from a tableau satisfying the increasingness requirements,  $P'$  also satisfies these requirements.

Let  $\sigma$  be a permutation. We write  $\sigma$  as

$$\sigma = \begin{array}{cccc} 1 & 2 & \dots & n \\ x_1 & x_2 & \dots & x_n \end{array}$$

Let  $P_0 = Q_0 = \emptyset$ . Now, define inductively  $P_k$  by inserting  $x_k$  into  $P_{k-1}$ ,

$$P_k = r_{x_k}(P_{k-1})$$

and  $Q_k$  by adding a cell  $k$  at the position where the insertion terminated. By doing so, it is clear that  $P_k$  and  $Q_k$  have the same shape. Moreover, it is also clear from the construction that  $Q_k$  is also a partial Young tableau. As a consequence,  $P_n$  and  $Q_n$  are both Young tableaux with the same shape. We define

$$\sigma \xrightarrow{\text{R-S}} (P, Q) := (P_n, Q_n).$$

In other words, the RSK maps  $\sigma$  into the pair of Young tableaux obtained at the end of this recursive procedure. As an example, suppose  $n = 7$ , and let

$$\sigma = \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 2 & 3 & 6 & 5 & 1 & 7 \end{array}$$

Then we obtain the pair of Young tableaux:

$$\begin{array}{|c|c|c|c|} \hline 1 & 3 & 5 & 7 \\ \hline 2 & 6 & & \\ \hline 4 & & & \\ \hline \end{array}, \begin{array}{|c|c|c|c|} \hline 1 & 3 & 4 & 7 \\ \hline 2 & 5 & & \\ \hline 6 & & & \\ \hline \end{array}$$

In general, the shapes are identical but the numbers different. A remarkable fact about the RSK correspondence is that this is a bijection:

**Theorem 2.**  $\sigma \xrightarrow{\text{R-S}} (P, Q)$  defines a bijection from the symmetric group  $S_n$  onto pairs of Young tableaux with same shape of size  $n$ .

I won't include the proof of this result here, but investigate the consequences of it for what is known as the Plancherel measure.

**The Plancherel measure.** Let  $\lambda$  be a partition of  $n$ . Call  $d_\lambda$  the number of Young tableaux of shape  $\lambda$ . A consequence of the above theorem is that

$$\sum_{\lambda \text{ partition}} d_\lambda^2 = n! \tag{6}$$

a formula of primary importance in the representation theory of the symmetric group. Moreover, consider the law  $P$  on shapes of Young diagrams obtained by applying the RSK correspondence to a randomly chosen uniform permutation. Then

$$P(\lambda) = \frac{d_\lambda^2}{n!} \tag{7}$$

This is obvious from Theorem 2. Indeed, the number of pairs of tableaux with shape  $\lambda$  is  $d_\lambda^2$  by definition. Now, given a pair of tableaux  $(P, Q)$  with shape  $\lambda$ , there is a unique (by the theorem) permutation such that the RSK maps  $\sigma$

into  $(P, Q)$ . Such a permutation has probability  $1/n!$ . Summing over all pairs of tableaux, we get (7). The measure defined by (7) on Young diagrams is called the **Plancherel measure**. Note that it is indeed a probability measure by (6).

**RSK and increasing subsequences.** Let  $\sigma$  be a permutation, and let  $(P, Q)$  be the image of  $\sigma$  under RSK. Consider the first row of  $P$ .

**Theorem 3.** *The length of the first row of  $P$  is equal to the length of the LIS of  $\sigma$ . In particular,  $L_n$  is equal in distribution to  $\lambda_1$ , where,  $\lambda$  has the Plancherel measure.*

This is proved by examining closely the properties of the RSK algorithm. Note that the first row of  $P$  itself need not be an increasing subsequence. (See the example above where the longest subsequence is  $S = \{2, 3, 5, 7\}$ ).

### 3.3 Mapping diagrams to nonintersecting paths

Let  $(P, Q)$  be a pair of Young tableaux with same shape and size  $n$ . Let  $r$  be the number of rows of  $P$ . We will construct a sequence on  $r$  paths in the planar lattice  $\mathbf{Z}^2$  such that the first path starts at height 0, the second at height -1, and so on. Each path represents a row of  $P$  followed by a row of  $Q$ . If our pair  $(P, Q)$  is

$$\begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline 4 & 6 & \\ \hline 5 & & \\ \hline \end{array}, \begin{array}{|c|c|c|} \hline 1 & 3 & 5 \\ \hline 2 & 6 & \\ \hline 4 & & \\ \hline \end{array}$$

The nonintersecting paths are represented next page.

To see how the above picture is obtained from  $(P, Q)$ , we declare that the first path (corresponding to the top row) makes jumps at 1,2, 3 and then, 5,3 and 1. The second path, corresponding to the second row, makes jumps at 4,6 and then 6,2. Finally the third path, corresponding to the last row, has only one jump in each half: one at 5 and one at 4. All the jumps are equal to +1 in the left hand side of the picture and to -1 in the right hand side.

It is not hard to check that the requirement that  $(P, Q)$  is a pair of Young tableaux implies that the  $r$  paths are nonintersecting, and that the paths completely encode the information in the pair of tableaux. Moreover if  $P$  and  $Q$  have a shape equal to  $\lambda = (\lambda_1, \dots, \lambda_r)$ , then the heights of the  $r$  paths at the center of the picture are respectively,  $\lambda_1, \lambda_2 - 1, \dots, \lambda_i - i + 1, \dots$ . Indeed, in the left hand side of the picture, the first path makes a total of  $\lambda_1$  jumps, and starts at height 0. Hence when it reaches the center, it is at a height equal to  $\lambda_1$ . Similarly, the  $i^{\text{th}}$  path starts at height  $-i + 1$  and makes a total of  $\lambda_i$  jumps before it reaches the middle vertical line. Now comes the crucial remark:

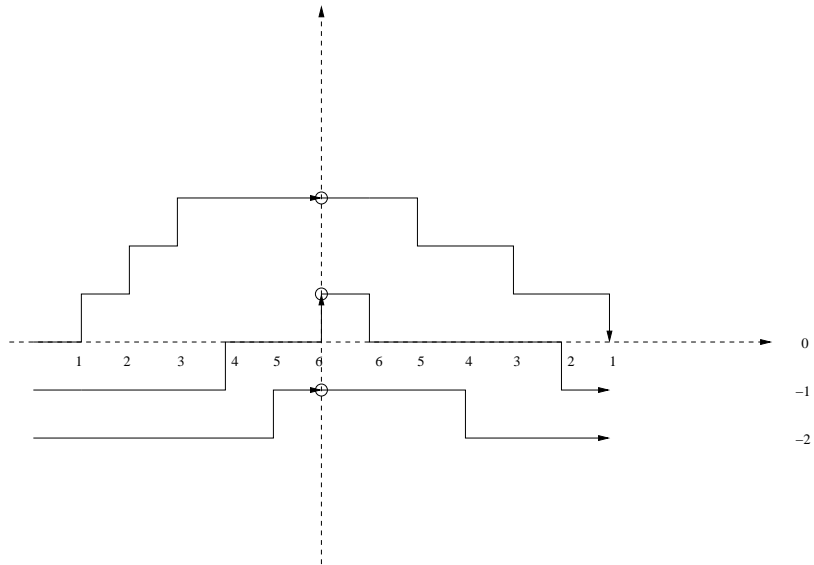


Figure 2: Nonintersecting paths associated to Young tableaux

**We may think of  $(\lambda_i - i, i = 1, \dots)$  as a point process on  $\mathbf{Z}$ .**

Note that in this representation,  $\lambda_1$ , which has the same distribution as the longest increasing subsequence of a random permutation, is the largest point of our point process.

**Modifying the paths.** In order to gain some flexibility, we want to allow the possibility that:

- Several paths can jump simultaneously
- Jumps can be larger than one.

To do so, we consider so-called *generalized Young tableaux*. These are Young diagrams whose boxes are filled with numbers in such a way that repetitions are allowed and satisfying the same “increasingness requirements” as before, but in the large sense. E.g.,

2	4	4
3	5	
3		

Given a pair of generalized Young tableaux  $(P, Q)$ , we can still construct a set of nonintersecting paths representing the tableaux. The fact that repetitions are allowed implies that several jumps can occur simultaneously, and that jumps can

be larger than one. Eg, for the path corresponding the first row, there will be two jumps of size 1 at 4, or equivalently, one jump of size 2 at 4. To see what this looks like, consider the two tableaux:

$$\begin{array}{|c|c|c|c|c|c|c|c|c|} \hline 1 & 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 \\ \hline 2 & 3 & & & & & & & \\ \hline \end{array}, \quad \begin{array}{|c|c|c|c|c|c|c|c|c|} \hline 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 3 \\ \hline 3 & 3 & & & & & & & \\ \hline \end{array}$$

Then the paths (still nonintersecting) will look like that.

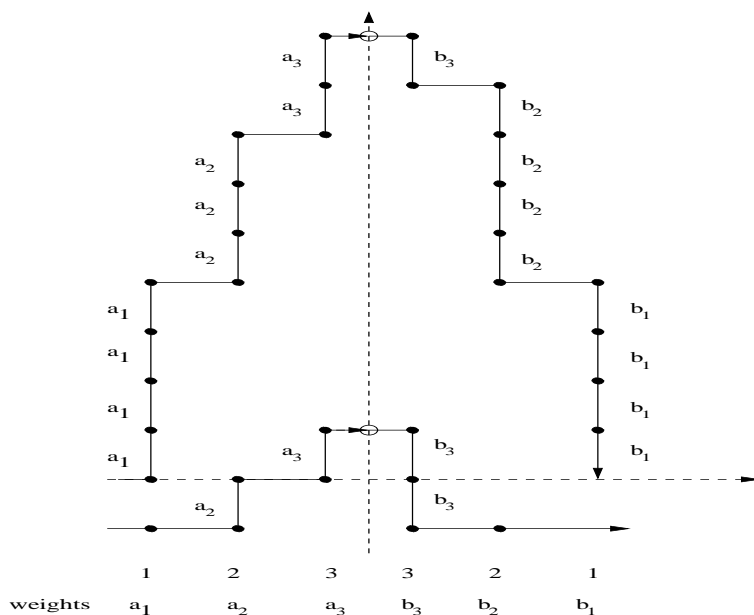


Figure 3: Nonintersecting weighed paths associated to generalized Young tableaux

Now, let  $a_1, a_2, \dots, a_n$  and  $b_1, b_2, \dots, b_n$  be two sequences of positive distinct numbers. We will use these numbers as weights of jumps. Each jump (of size +1) occurring at time  $i$  in the left hand side will have a weight equal to  $a_i$ , and each jumps of size  $-1$  at time  $i$  in the right-hand side of the picture will have a weight equal to  $b_i$ . This allows us to represent

generalized Young tableaux



set of nonintersecting paths on a directed weighted graph.

The directed graph on which these paths lie is represented in the next figure.

The fact that we can work with paths on a directed graph will allow us to use the LGV theorem a bit below.

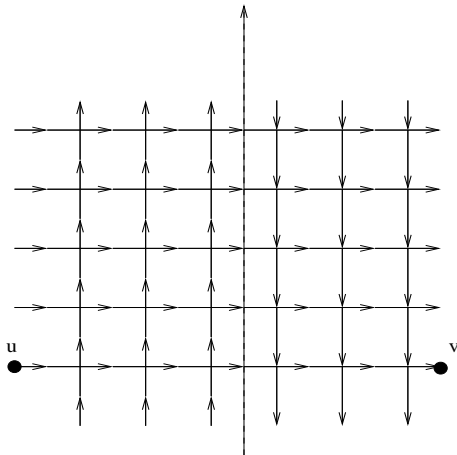


Figure 4: The directed graph for the LGV theorem.

**Recuperating the Plancherel measure from the paths.** As we just mentioned, there is a beautiful result in Combinatorics, the LGV theorem, which allows us to analyze the nonintersecting paths very well. We show how to get back to the Plancherel measure. Define the weight of a path to be the product of the weight of the edges along that path (horizontal edges have weight 1), and if the weight of a set of path is the product of the weight of the individual paths, then we get the following fundamental representation. If  $\lambda$  is a fixed partition of  $n$ :

$$\frac{1}{n!} \sum_{\text{non-int. paths with height } \lambda \text{ at } 0} \text{weight of paths} \quad (8)$$

is a polynomial in  $a_1, \dots, a_n, b_1, \dots, b_n$ , whose  $a_1 \dots a_n b_1 \dots b_n$ -coefficient is equal to  $d_\lambda^2/n!$ , the Plancherel measure of  $\lambda$ . Let us justify this claim. What differentiates a Young tableau from a generalized Young tableau, is the fact that the numbers  $1, 2, \dots, n$  will be used exactly once in the tableau. Hence every path whose weight is  $a_1 \dots a_n$  represents a Young tableau, and similarly on the right hand side. Hence if we sum over nonintersecting paths having height  $\lambda$  at 0, and care only about the contribution of paths having weight  $a_1 \dots a_n b_1 \dots b_n$ , we obtain exactly the number of pairs of Young tableaux with shape  $\lambda$ . Dividing by  $n!$  in (8) shows that the coefficient is indeed the Plancherel measure of  $\lambda$ .

### 3.4 The LGV Theorem.

Let  $G = (V, E)$  be a directed acyclic graph, with no multiple edges. Consider paths from  $\underline{u} = (u_1, \dots, u_n)$  to  $\underline{v} = (v_1, \dots, v_n)$ , i.e., directed paths from  $u_i$  to  $v_i$ .

(Here  $n$  is fixed  $\geq 1$  and  $u_i$ 's and  $v_i$ 's are fixed vertices). As usual, we say that two directed paths intersect if they share a common vertex. Choose a weight function  $w : E \rightarrow \mathbf{C}$  on the edges of the graph. If  $\pi$  a directed path in  $G$ , define:

$$w(\pi) = \prod_{e \in \pi} w(e)$$

If  $(\pi_1, \dots, \pi_n)$  is a set path from  $\underline{u}$  to  $\underline{v}$  then we define the weight of this set of paths to be:

$$w(\pi_1, \dots, \pi_n) = \prod_{i=1}^n w(\pi_i)$$

Finally, if  $\mathcal{F}$  is a family of set of paths going from  $\underline{u}$  to  $\underline{v}$ , then we define  $W(\mathcal{F})$  to be

$$W(\mathcal{F}) = \sum_{(\pi_1, \dots, \pi_n) \in \mathcal{F}} w(\pi_1, \dots, \pi_n)$$

If  $\underline{u}$  is a collection of  $n$  vertices  $(u_1, \dots, u_n)$ , and if  $\sigma$  is a permutation, call  $\underline{u}_\sigma$  the collection  $(u_{\sigma(1)}, \dots, u_{\sigma(n)})$ . The next result, the LGV theorem (for Lindström, Gessel and Viennot) computes the weight of all nonintersecting paths from  $\underline{u}$  to  $\underline{v}$  and expresses it as a determinant.

**Theorem 4.** (LGV). *Assume that there are no non-intersecting paths from  $\underline{u}$  to  $\underline{v}_\sigma$  unless  $\sigma =$  the identity. Let  $\mathcal{F}$  be the family of non-intersecting paths from  $\underline{u}$  to  $\underline{v}$ .*

$$W(\mathcal{F}) = \det(q(u_i, v_j)_{i,j=1}^n) \tag{9}$$

where

$$q(u, v) = \sum_{\pi \text{ a path from } u \text{ to } v} w(\pi) \tag{10}$$

The proof follows an old argument of Karlin and McGregor [9] which expresses the probability that  $n$  independent simple random walk paths starting at  $\underline{x}$  and ending at  $\underline{y}$  are nonintersecting, as a certain determinant. (Cf. also a related identity of Fomin and its continuous generalization in terms of Schramm-Loewner evolutions by Kozdron [11].) The proof is fairly simple and intuitive so we sketch it.

*Proof.* Let  $\Pi(\underline{u}, \underline{v})$  be the set of paths from  $\underline{u}$  to  $\underline{v}$ . Consider the determinant in

the right-hand side of (9). Using the definition of a determinant, we may expand:

$$\begin{aligned}
\det q(\underline{u}, \underline{v}) &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n q(u_i, v_{\sigma(i)}) \\
&= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n \left( \sum_{\pi \in \Pi(u_i, v_{\sigma(i)})} w(\pi) \right) \\
&= \sum_{\sigma \in S_n} \sum_{(\pi_1, \dots, \pi_n) \in \Pi(\underline{u}, \underline{v}_\sigma)} \operatorname{sgn}(\sigma) w(\pi_1, \dots, \pi_n)
\end{aligned}$$

We split this last sum according to paths with no intersection (n.i.) and path with intersections (w.i.):

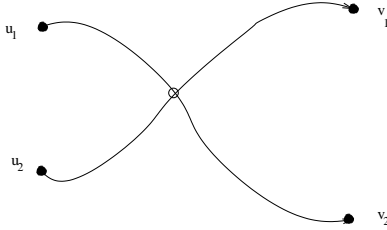
$$\begin{aligned}
&= \sum_{\sigma \in S_n} \sum_{(\pi_1, \dots, \pi_n) \in \Pi_{\text{n.i.}}(\underline{u}, \underline{v}_\sigma)} \operatorname{sgn}(\sigma) w(\pi_1, \dots, \pi_n) \\
&\quad + \sum_{\sigma \in S_n} \sum_{(\pi_1, \dots, \pi_n) \in \Pi_{\text{w.i.}}(\underline{u}, \underline{v}_\sigma)} \operatorname{sgn}(\sigma) w(\pi_1, \dots, \pi_n)
\end{aligned}$$

Let  $S_1$  and  $S_2$  be these two sums respectively. Observe that  $S_1$  is what we want:

$$S_1 = W(\Pi_{\text{n.i.}}(\underline{u}, \underline{v}))$$

because by assumption, the only  $\sigma$  for which there are nonintersecting paths between  $\underline{u}$  and  $\underline{v}_\sigma$  is  $\sigma = id$ .

It thus suffices to show that  $S_2 = 0$ . To see this, consider a path from  $\underline{u}$  to  $\underline{v}$  with intersection. Suppose it looks like this.



We may redefine the paths so that  $u_1 \rightarrow v_1$  instead of  $v_2$ , and  $u_2 \rightarrow v_2$  instead of  $v_1$ . When we do such an operation for every path from  $\underline{u}$  to  $\underline{v}$  with intersections, this does not change  $w(\pi_1, \dots, \pi_n)$ . However the *sign* of the permutation is changed to its opposite. Since this operation is an involution, we conclude

$$S_2 = -S_2 \quad \text{or, } S_2 = 0.$$

This concludes the proof of the LGV theorem.  $\square$

### 3.5 Applying the LGV theorem.

What does the LGV theorem say about our situation ? Let  $x_i = \lambda_i - i + n$ . Let  $p_n(\underline{x})$  be the probability of particular configuration of points for the point process  $\underline{x}$ .

**Theorem 5.**

$$p_n(\underline{x}) = [a_1 \dots a_n b_1 \dots b_n] \frac{1}{n! Z_n \Delta_n(\underline{a}) \Delta_n(\underline{b})} \det(a_i^{x_j})_{i,j=1}^n \det(b_i^{x_j})_{i,j=1}^n$$

Here  $[x]f(\underline{x})$  denotes the coefficient of  $x$  in  $f(\underline{x})$ ,  $Z_n$  is a normalizing factor and  $\Delta_n(\underline{a})$  denotes the Vandermonde determinant:

$$\Delta_n(\underline{a}) = \prod_{1 \leq i < j \leq n} (a_i - a_j)$$

*Proof.* We first compute the transition weight  $q(u, v)$ . If  $u, v \in \mathbf{Z}$ , let  $k = v - u$ , then

$$q(u, v) = \sum_{1 \leq i_1 \leq \dots \leq i_k \leq n} a_{i_1} \dots a_{i_k} \quad (11)$$

if  $v > u$ ,  $q(u, v) = 1$  if  $v = u$ , and  $q(u, v) = 0$  if  $v < u$ . From the above formula, it follows that if  $v > u$ , then

$$q(u, v) = h_k(a_1, \dots, a_n) \quad (12)$$

where  $h_k(a_1, \dots, a_n)$  is the  $k^{\text{th}}$  complete symmetric polynomial. which is defined by

$$\prod_{j=1}^n \frac{1}{1 - a_j z} = \sum_{k=-\infty}^{\infty} h_k(a_1, \dots, a_n) z^k \quad (13)$$

To see why (12) is true, recall that  $1/(1 - a_j z) = \sum_{k \geq 0} a_j^k z^k$ . Expanding the product into a sum, we get that the coefficient of  $z^k$  in the sum is equal to the right hand side of (11). It follows from the LGV theorem that the weight of non intersecting paths from  $(0, -1, \dots, -N + 1)$  to  $(\lambda_1, \dots, \lambda_n - n + 1)$  is equal to

$$s_\lambda(\underline{a}) = \det(h_{\lambda_i - i + j}(\underline{a}))_{i,j=1}^n \quad (14)$$

and similarly for the right hand side of the picture. The polynomial  $s_\lambda(\underline{a})$  is called the **Schur polynomial** labelled by  $\lambda$ . It follows that the weight of all nonintersecting paths (i.e., the right hand side of (8)) is equal to

$$s_\lambda(\underline{a}) s_\lambda(\underline{b})$$

Let  $Z_n = \sum_{\lambda} s_{\lambda}(\underline{a})s_{\lambda}(\underline{b})$ . The measure

$$p_S(\lambda) = \frac{1}{Z_n} s_{\lambda}(\underline{a})s_{\lambda}(\underline{b}) \quad (15)$$

is a probability distribution on Young diagrams called the **Schur measure**. (It was first introduced by Okunkov). A well-known result in symmetric function theory tells us that the determinant which defines the Schur polynomial in (14) is equal to

$$s_{\lambda}(\underline{a}) = \frac{\det(a_j^{\lambda_i - i + n})_{i,j}}{\det(a_j^{n-i})}$$

Permuting the rows in the matrix of the denominator, we see that the determinant is the Vandermonde determinant  $\Delta_n(\underline{a})$ . Hence it follows that

$$p_S(\lambda) = \frac{1}{Z_n \Delta_n(\underline{a}) \Delta_n(\underline{b})} \det(a_i^{x_j})_{i,j} \det(b_i^{x_j})_{i,j} \quad (16)$$

Moreover, it is clear by (8) that one has:

$$p_n(\underline{x}) = \frac{1}{n!} [a_1 \dots a_n b_1 \dots b_n] Z_n p_S(\lambda)$$

and rewriting  $Z_n$  as  $n!Z_n$  concludes the proof of the theorem.  $\square$

So far,  $p_n(\underline{x})$  and  $p_S(\underline{x})$  are defined on  $x_1 > \dots > x_n$  but since the right-hand side of (16) is symmetric, we can define it on  $(x_1, \dots, x_n) \in \mathbf{N}$  by the same formula, which is why we introduced an extra factor  $n!$  in the formula. (so that  $Z_n$  represents the partition function of the unordered system).

### 3.6 The analogy with GUE.

The analytical expression (16) for the Schur measure reminds us of the probability density of eigenvalues of the GUE. In that context we had

$$p_{GUE}(\underline{x}) = \frac{1}{n! Z_n} \Delta_n(\underline{x})^2 \prod_{i=1}^n e^{-x_i^2} \quad (17)$$

Both the Schur measure (16) and the GUE measure (17) can be written as:

$$p(\underline{x}) = \frac{1}{n! Z_n} \det(\phi_j(x_k))_{j,k} \det(\psi_j(x_k)) \quad (18)$$

with  $\phi_j(x) = a_j^x$  and  $\psi_j(x) = b_j^x$  for the Schur measure, and  $\phi_j(x) = \psi_j(x) = x^{j-1} e^{-x^2/2}$  for the GUE.

### 3.7 The Schur measure is determinantal

Since the GUE has a determinantal structure, it is natural to expect that the Schur measure can be expressed as a determinantal process. In fact the following result treats both the GUE and the Schur case at once. Suppose that we are given a space  $X$  and a measure  $\mu$  (not necessarily a probability measure) on  $X$ , and a density  $p_n(\underline{x})$  of the form (18). Assume that the measure  $\mu$  is such that

$$\int_{X^n} p_n(\underline{x}) d\mu(x_1) \dots d\mu(x_n) = 1$$

(In the Schur case,  $X = \mathbf{N}$  and  $\mu$  is the counting measure, and in the GUE case,  $X = \mathbf{R}$  and  $\mu$  is the Lebesgue measure.)

Then we have the following result.

**Theorem 6.** *Let  $A = (A_{i,j})$  be the matrix such that*

$$A_{i,j} = \int_X \phi_i(x) \psi_j(x) d\mu(x)$$

*Then the point process defined by  $p_n(\underline{x})$  in (18) is a determinantal point process on  $(X, \mu)$  with kernel*

$$K_n(x, y) = \sum_{i,j=1}^n \psi_i(x) (A^{-1})_{i,j} \phi_j(y)$$

*Furthermore,  $Z_n = \det(A)$ .*

*Proof.* The proof is based on the following identity:

$$\frac{1}{n!} \int_{X^n} \det(\phi_i(x_j))_{i,j} \det(\psi_i(x_j))_{i,j} d^n \mu(\underline{x}) = \det \left( \int_X \phi_i(x) \psi_j(x) d\mu(x) \right)_{i,j} \quad (19)$$

Johansson ([8], Proposition 2.10) calls this the generalized Cauchy-Binet identity, which arises in some very particular case. It's a simple calculation, which can be looked up in the above reference.

From the generalized Cauchy Binet identity (19), we immediately deduce that since  $\int_{X^n} p_n(\underline{x}) d^n \mu(\underline{x}) = 1$ ,

$$\begin{aligned} Z_n &= \frac{1}{n!} \int_{X^n} \det(\phi_j(x_k))_{j,k} \det(\psi_j(x_k))_{j,k} d^n \mu(\underline{x}) \\ &= \det A \end{aligned}$$

Johansson remarks that it is necessary and sufficient for a point process to be determinantal with kernel  $K$  that for all functions  $g : X \rightarrow \mathbf{C}$ ,

$$\int_{X^n} \prod_{j=1}^n (1+g(x_j)) p_n(\underline{x}) d\mu^n(x) = \sum_{k=0}^{\infty} \frac{1}{k!} \int_{X^k} \prod_{j=1}^k g(x_j) \det K(x_i, x_j)_{k \times k} d\mu^k(\underline{x}) \quad (20)$$

This is essentially Theorem 2 of Soshnikov [14]. Applying (19) with  $\phi$  replaced by  $(1+g)\phi$  shows that the left hand side of (20) is equal to

$$\begin{aligned} \int_{X^n} \prod_{j=1}^n (1+g(x_j)) p_n(\underline{x}) d\mu^n(x) &= \frac{1}{\det A} \det \left( \int_X (1+g(x)) \phi_i(x) \psi_j(x) d\mu(x) \right)_{i,j} \\ &= \frac{1}{\det A} \det \left( A_{i,j} + \int_X g(x) \phi_i(x) \psi_j(x) d\mu(x) \right)_{i,j} \\ &= \det \left( I + \sum_{l=1}^n (A^{-1})_{i,l} \int_X g(x) \phi_l(x) \psi_j(x) d\mu(x) \right)_{i,j} \end{aligned} \quad (21)$$

Let  $(B_{ij})$  be the second matrix in this last expression. Inverting the integral and summation signs,

$$B_{i,j} = \int_X f_i(x) g(x) \psi_j(x) d\mu(x)$$

where  $f_i(x) = \sum_{l=1}^n (A^{-1})_{i,l} \phi_l(x)$ . We claim that we have the identity:

$$\det(I + B) = \sum_{k=0}^n \frac{1}{k!} \sum_{i_1, \dots, i_k=1}^n \det(B_{i_r, i_s})_{r,s=1}^k$$

which is obtained by multilinearity of the determinant, expanding e.g. along the columns. Now, looking at (21), the right-hand side is equal to

$$\begin{aligned} \det(I + B) &= \sum_{k=0}^n \frac{1}{k!} \sum_{i_1, \dots, i_k=1}^n \det(B_{i_r, i_s})_{r,s=1}^k \\ &= \sum_{k=0}^n \frac{1}{k!} \sum_{i_1, \dots, i_k=1}^n \det \left( \int_X f_{i_r}(x) g(x) \psi_{i_s}(x) d\mu(x) \right) \end{aligned}$$

Applying (19), and using the multilinearity of the determinant this is also equal

to:

$$\begin{aligned}
\det(I + B) &= \sum_{k=0}^n \frac{1}{k!} \sum_{i_1, \dots, i_k=1}^n \int_{X^k} \det(f_{i_r}(x_s))_{r,s=1}^k \det(\psi_{i_r}(x_s))_{r,s} d^k \mu(\underline{x}) \\
&= \sum_{k=0}^n \frac{1}{k!} \int_{X^k} \det \left( \sum_{i=1}^n f_i(x_r) \psi_i(x_s) \right)_{1 \leq r, s \leq k} d^k \mu(\underline{x}) \\
&= \sum_{k=0}^n \frac{1}{k!} \int_{X^k} \prod_{r=1}^n g(x_r) \det(K_n(x_r, x_s))_{1 \leq r, s \leq k} d^k \mu(\underline{x})
\end{aligned}$$

This finishes the proof of the Theorem.  $\square$

### 3.8 Towards the Airy process

Here we give a flavor of the difficult remaining difficult computations that show that the Schur measure is asymptotically governed by the Airy process. (For the Plancherel measure, the calculations are very similar, but need a few additional arguments that I haven't been able to fill in.)

Let  $Z'_n = Z_n \Delta_n(\underline{a}) \Delta_n(\underline{b})$ , so that

$$p_S(\lambda) = \frac{1}{n! Z'_n} \det(a_i^{x_j})_{i,j} \det(b_i^{x_j})_{i,j}.$$

Applying the above theorem, we find that  $Z'_n = \det A$  where (recalling that  $\mu$  is the counting measure on  $X = \mathbf{N}$ ),

$$A_{i,j} = \int_X \phi_i(x) \psi_j(x) d\mu(x) = \sum_{x=0}^{\infty} a_i^x b_j^x = \frac{1}{1 - a_i b_j}$$

This is a Cauchy determinant, and we conclude that

$$Z'_n = \det A = \frac{\Delta_n(\underline{a}) \Delta_n(\underline{b})}{\prod_{i,j=1}^n (1 - a_i b_j)}$$

It follows by Theorem 5 that

$$p_n(\underline{x}) = [a_1 \dots a_n b_1 \dots b_n] \frac{p_S(\lambda)}{\prod_{i,j=1}^n (1 - a_i b_j)} \quad (22)$$

The difficulty in using Theorem 6 is to find an expression for  $A^{-1}$  which makes asymptotic computations possible. In the GUE case, it is possible to express

(18) using the determinant of orthogonal polynomials (more precisely, the Hermite polynomials), which make  $A$  a diagonal matrix, trivial to inverse. This is an interesting exercise, and asymptotic of the Hermite polynomials used jointly with the Christoffel-Darboux formula yield directly the result that the largest eigenvalue of a GUE matrix has asymptotically Tracy-Widom fluctuations. Here, the calculations are much heavier. However, we sketch an idea of how to get there. This is adapted from several portions of Johansson's Les Houches notes [8]. For a real-life proof, see the paper of Johansson [6] or (more analytic), that by Borodin, Okunkov, and Olshanski [3] (the two papers appeared more or less simultaneously).

To compute  $(A^{-1})_{k,l}$  we use the fact that the inverse of a matrix  $A$  is the transpose of the comatrix. Let  $A^{(k,l)}$  be the matrix obtained from  $A$  by deleting the  $k^{\text{th}}$  row and the  $l^{\text{th}}$  column. Then

$$K_n(x, y) = \sum_{k,l=1}^n (-1)^{k+l} b_k^y \frac{\det A^{(kl)}}{\det A} a_l^x$$

Note that  $\det A^{(kl)}$  is still a Cauchy determinant, and so we can use the same formula as above. We get:

$$K_n(x, y) = \sum_{k,l=1}^n b_k^y a_l^x \frac{\prod_{j=1}^n (1 - b_j a_l)(1 - b_k a_j)}{(1 - b_k a_l) \prod_{j \neq k} (b_k - b_j) \prod_{j \neq l} (a_j - a_l)}$$

To this we may apply the *residue theorem*. Assuming  $|a_i| < 1$ ,  $|b_i| < 1$ , and letting  $\gamma_1 = \{|z| = 1 + \varepsilon\}$ , and  $\gamma_2 = \{|w| = 1\}$ , we obtain:

$$K_n(x, y) = \frac{1}{(2i\pi)^2} \int_{\gamma_1} \frac{dz}{z} \int_{\gamma_2} \frac{w}{z-w} \frac{w^x}{z^y} \prod_{j=1}^n \frac{(1 - wa_j)(z - b_j)}{(1 - za_j)(w - a_j)}$$

Take all the  $a_k$  and the  $b_j$  to be equal to a constant, say  $a_n = \sqrt{\alpha}/n$ , and  $\alpha$  will be chosen large later. In those conditions,

$$K_n(x, y) = \frac{1}{(2i\pi)^2} \int_{\gamma_1} \frac{dz}{z} \int_{\gamma_2} \frac{w}{z-w} \frac{w^{x-n}}{z^{y-n}} \frac{(1 - wa_n)^n (1 - a_n/z)^n}{(1 - za_n)^n (1 - a_n/w)^n}$$

Call  $\underline{x}' = \underline{x} - n = (\lambda_i - i, i = 1, \dots, n)$ . Writing

$$w/(z-w) = 1/(1-w/z) = \sum_{k=0}^{\infty} (w/z)^k,$$

we obtain the asymptotics

$$\begin{aligned}
K_n(x', y') &\rightarrow \frac{1}{(2i\pi)^2} \int_{\gamma_1} \frac{dz}{z} \int_{\gamma_2} \frac{w}{z-w} \frac{w^{x'}}{z^{y'}} e^{\sqrt{\alpha}(z-1/z) - \sqrt{\alpha}(w-1/w)} \\
&= \sum_{k=0}^{\infty} \frac{1}{2\pi i} \int_{\gamma_1} \frac{dz}{z} \frac{1}{z^{y'+k}} e^{\sqrt{\alpha}(z-1/z)} \frac{1}{2\pi i} \int_{\gamma_2} \frac{dw}{w} w^{k+x'} e^{-\sqrt{\alpha}(w-1/w)} \\
&= \sum_{k=0}^{\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-(k+x')i\theta + \sqrt{\alpha}2i \sin \theta} d\theta \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-(k+y')i\theta + \sqrt{\alpha}2i \sin \theta} d\theta
\end{aligned}$$

The function  $J_s(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-is\theta + xi \sin \theta} d\theta$  is the Bessel function of order  $s$ , so we conclude

$$K_n(x', y') \rightarrow \sum_{k=0}^{\infty} J_{k+x'}(2\sqrt{\alpha}) J_{k+y'}(2\sqrt{\alpha}) = B^\alpha(x', y')$$

$B^\alpha(x, y)$  is known as the discrete Bessel kernel. On the other hand, there is an asymptotic formula:

$$\xi^{1/6} J_{2\sqrt{\xi} + u\xi^{1/6}}(2\sqrt{\xi}) \rightarrow \text{Ai}(u)$$

uniformly for  $u$  in a compact interval as  $\xi \rightarrow \infty$ . From this and simple manipulations, one can deduce that if  $x' = \sqrt{2n} + un^{1/6}$ , and if  $y = \sqrt{2n} + vn^{1/6}$ , we have (taking  $\alpha = n$ , although this step admittedly requires some justification), after we recognize a Riemann sum:

$$\begin{aligned}
n^{1/6} K_n(x', y') &\sim n^{-1/6} \sum_{k=0}^{\infty} n^{1/6} J_{\sqrt{2n} + un^{1/6} + k}(2\sqrt{n}) n^{1/6} J_{\sqrt{2n} + vn^{1/6} + k}(2\sqrt{n}) \\
&\sim \int_0^{\infty} \text{Ai}(u+t) \text{Ai}(v+t) dt \\
&= A(u, v)
\end{aligned}$$

by (4). Here  $A(u, v)$  is the Airy kernel. As explained at the beginning, this proves that if  $S_1$  the largest point of the Schur measure with this choice of parameters ( $a_i = b_i = n^{-1/2}$ ), then

$$\frac{S_1 - \sqrt{2n}}{n^{1/6}} \rightarrow_d F_{TW}(\cdot)$$

the Tracy-Widom distribution. □

## References

- [1] D. Aldous and P. Diaconis (1999). Longest increasing subsequences: from patience sorting to the Baik-Deift-Johansson theorem. *Bull. Amer. Math. Soc.* 36, 413–432.

- [2] J. Baik, P. Deift, and K. Johansson (1999). On the distribution of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.* 12, 1119–1178.
- [3] A. Borodin, A. Okunkov, and G. Olshanski (2000). Asymptotics of Plancherel measure for symmetric groups. *J. Amer. Math. Soc.* 13, 481–515.
- [4] J.M. Hammersley (1972). A few seedlings of research. In *Proc. Sixth Berkeley Symp. Math. Statist. and Probability.* 1, 345–394. University of California Press.
- [5] B. Hough, M. Krishnapur, Y. Peres, and B. Virag. (2006). Determinantal processes and Independence. *Probability surveys* 3, 206–229.
- [6] K. Johansson (2001). Discrete orthogonal polynomial ensembles and the Plancherel measure. *Annals of Math.* 153, 259–296.
- [7] K. Johansson (2000). Shape fluctuations and random matrices. *Comm. Math. Phys.* 209, 437–476.
- [8] K. Johansson (2005). Random matrices and determinantal point processes. Notes for Les Houches summer school, available at arXiv:math-ph/0510038.
- [9] S. Karlin and J. McGregor (1959). Coincidence probabilities. *Pacific J. Math.* 9, 1141–1164.
- [10] S.V. Kerov and A.M. Vershik (1981). Asymptotic theory of the characters of the symmetric group. *Functional Anal. Appl.*, 15:246-255.
- [11] M. Kozdron (2007). The scaling limit of Fomin’s identity for two paths. Preprint, arXiv:math.PR/0703615.
- [12] B.F. Logan and L.A. Shepp (1977). A variational problem for random Young tableaux. *Advances in Math.*, 26:206-222.
- [13] B. Sagan. *The Symmetric Group. Representations, Combinatorial Algorithms, and Symmetric Functions.* Springer, Second Edition.
- [14] A. Soshnikov (2000). Determinantal random point fields. *Russian Math. Survey* 55:5, 923–975.