1 Introduction

Here is a permutation represented as a list of numbers:

$$\sigma = (4, 1, 2, 7, 6, 5, 8, 9, 3).$$

This means that $\sigma \in S_9$ acts on $\{1, \dots, 9\}$ by $\sigma(1) = 4, \sigma(2) = 1$, et cetera. You can check that the longest increasing subsequence of σ is $1, 2, 5, 8, 9$, which has length 5. There are other longest increasing subsequences, like $1, 2, 7, 8, 9$, but there are no increasing subsequences of length 6. Let $L(\pi)$ denote the length of the longest increasing subsequence in any permutation π. For our example, $L(\sigma) = 5$.

What can we say about $L(\sigma_n)$ when σ_n is drawn uniformly at random from S_n, the set of all permutations on n elements? It turns out that $L(\sigma_n)$ grows like $\sqrt{n}$. In fact,

$$\frac{L(\sigma_n)}{\sqrt{n}} \xrightarrow{\mathbb{P}} 2.$$

We will discuss some of the ideas that go into proving this result, which is due to Vershik-Kerov and Logan-Shepp. This is a "law of large numbers" for $L(\sigma_n)$. There is also a "central limit theorem" for $L(\sigma_n)$ which states that

$$\frac{L(\sigma_n) - 2\sqrt{n}}{n^{1/6}} \xrightarrow{d} F_2,$$

where F_2 is the GUE Tracy-Widom distribution. This result is due to Baik, Deift, and Johansson.

In these notes, we will mainly discuss the LLN-type result, but we might say a few words towards the end about the CLT-type result. We will also skip most of the (somewhat gory) details about the proof of the LLN-type result, but hopefully we will see some of the key ideas involved in the proof. This talk is based on Chapter 1 (and some bits of Chapter 2) from "The Surprising Mathematics of Longest Increasing Subsequences" by Dan Romik.

2 The Growth Rate is Sqrtic

2.1 Lower Bounds and Upper Bounds

The following classical theorem due to Erdős and Szekeres will allow us to show that $\mathbb{E}[L(\sigma_n)] \geq \sqrt{n}$.

Theorem 1. *Let $L(\sigma)$ denote the longest increasing subsequence of σ and let $D(\sigma)$ denote the longest decreasing subsequence of σ. If $n > sr$ for integers r and s, then either $L(\sigma) > s$ or $D(\sigma) > r$.*

Proof. For every $k \in \{1, \dots, n\}$, let $L_k(\sigma)$ (resp. $D_k(\sigma)$) denote the longest increasing (resp. decreasing) subsequence of σ that ends with $\sigma(k)$. Then the pairs $(L_k(\sigma), D_k(\sigma))$ are distinct for distinct k. Indeed, suppose $1 \leq k < j \leq n$. Then there are two cases

Case 1: $\sigma(k) < \sigma(j)$. We can append $\sigma(j)$ to a increasing subsequence ending at $\sigma(k)$, so $L_j(\sigma) > L_k(\sigma)$.

Case 2: $\sigma(k) > \sigma(j)$. We can append $\sigma(j)$ to a decreasing subsequence ending at $\sigma(k)$, so $D_j(\sigma) > D_k(\sigma)$.

Now, if if $L(\sigma) \leq s$ and $D(\sigma) \leq r$ then $(L_k(\sigma), D_k(\sigma)) \in \{1, \dots, s\} \times \{1, \dots, r\}$ for all k. But that set has cardinality $sr < n$, and there are n distinct pairs. ∎

Corollary 2. *If σ_n is a uniformly random permutation in S_n, then $\mathbb{E}[L(\sigma_n)] \geq \sqrt{n}$.*

Proof. Notice that by symmetry, $L(\sigma_n)$ has the same distribution as $D(\sigma_n)$. Thus

$$\mathbb{E}[L(\sigma_n)] = \mathbb{E}\left[\frac{L(\sigma_n) + D(\sigma_n)}{2}\right] \geq \mathbb{E}\left[\sqrt{L(\sigma_n)D(\sigma_n)}\right]$$

by linearity of expectation, and the AM-GM inequality. The Erdős-Szekeres theorem implies that (deterministically) $L(\sigma_n)D(\sigma_n) \geq n$, so the right-hand side is at least $\sqrt{n}$. ∎

Here is an upper bound (with a different constant) that we will not prove.

Proposition 3. *If σ_n is a uniformly random permutation in S_n, then*

$$\limsup_{n\to\infty} \frac{\mathbb{E}[L(\sigma_n)]}{\sqrt{n}} \le e.$$

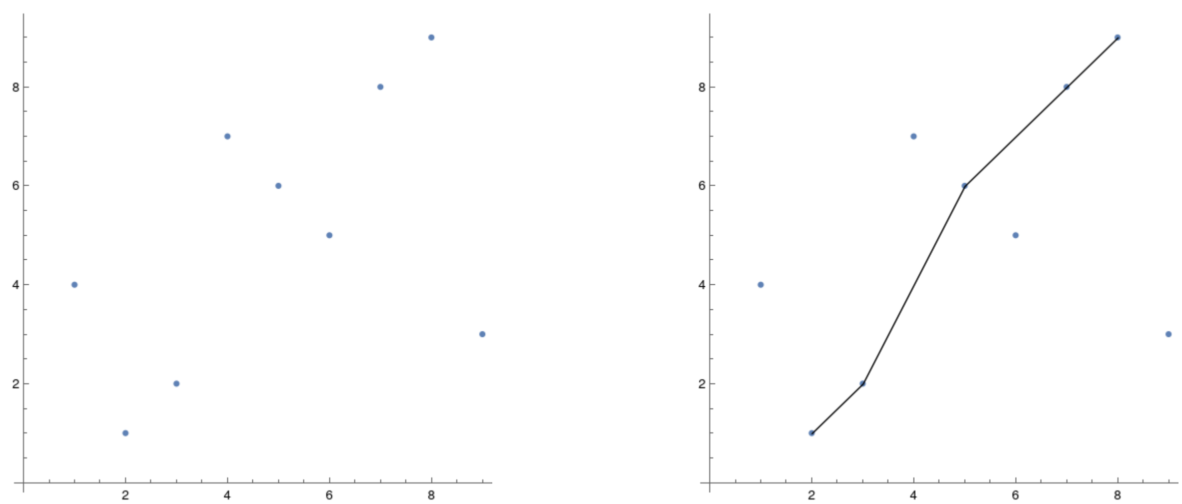
If you wanted to prove this proposition, the main idea is to examine the number of increasing sequences of a particular length, and show that the expected number is very small if the length is at least $\alpha\sqrt{n}$ for $\alpha > e$. The e appears when we apply a Stirling-esque inequality, namely $n! \ge (n/e)^n$, which holds because $e^n \ge \frac{n^n}{n!}$.

2.2 Hammersley's Theorem

We have seen that $\mathbb{E}[L(\sigma_n)] = \Theta(\sqrt{n})$, but this does not imply that $\frac{\mathbb{E}[L(\sigma_n)]}{\sqrt{n}}$ converges to anything, to say nothing of the convergence of $\frac{L(\sigma_n)}{\sqrt{n}}$ in probability. Nevertheless, here is a theorem due to Hammersley.

Theorem 4. *The limit $\Lambda = \lim_{n\to\infty} \frac{\mathbb{E}[L(\sigma_n)]}{\sqrt{n}}$ exists, and $\frac{L(\sigma_n)}{\sqrt{n}} \to \Lambda$ in probability.*

The idea is to interpret a permutation more geometrically. We plot it as follows:



On the left, we see the permutation $\sigma = (4, 1, 2, 7, 6, 5, 8, 9, 3)$ from the introduction. On the right, a longest increasing subsequence is highlighted. Increasing subsequences are the same thing as paths which travel up and to the right, made of segments that start and end at points in the plot of the permutation.

Notice that we can recover the permutation from the plot by recording the relative x-coordinates as the inputs and the relative y-coordinates as the outputs (by "relative" I mean that we only save the ordering information). This does not require that the points have integer coordinates. In fact, if we sample n points uniformly at random in the unit square, the permutation we obtain by considering the orders of the x and y coordinates is uniformly random in S_n. Of course, this will fail if any of the points lie on the same axis-parallel line, but there is zero probability for that tohappen.

At first glance, this doesn't seem to help us very much. However, this allows us to interpret a permutation as a finite portion of a larger collection of random points in the plane, called the Poisson point process. This is a random set $\Pi \subseteq \mathbb{R}^2$ with a few nice properties which we will record here without proof.

Proposition 5. *Let Π be the Poisson point process (of intensity 1). Then all of the following hold.*

(1) *Π is almost surely discrete and its points almost surely have distinct x and y coordinates.*

(2) *The distribution of Π is translation-invariant.*

(3) *If A is any deterministic compact subset of $\mathbb{R}^2$, then $N(A) = |A \cap \Pi|$, the cardinality of $A \cap \Pi$, follows a Poisson distribution with parameter $\lambda(A)$, the Lebesgue measure of A.*

(4) *If* $A_1, \ldots, A_k$ *are disjoint compact sets, then* $N(A_1), \ldots, N(A_k)$ *are independent.*

(5) *For any compact set* A *with positive Lebesgue measure, conditioned on the event* $N(A) = n$*, the set* $\Pi \cap A$ *has the same distribution as* n *independent uniformly random points in* A*.*

For any finite set B of points in $\mathbb{R}^2$ with distinct x and y coordinates, we denote by $L(B)$ the maximal number of points contained in a path which travels only up and right; i.e. this is the same as $L(\sigma)$ if σ is the permutation obtained from B. Now, for any $t > s \geq 0$, consider the random variable

$$Y_{s,t} = L\left(\Pi \cap [s,t)^2\right),$$

which is the maximal number of points contained in an up-right path through Poisson points contained in the box $[s,t)^2$. Notice that $Y_{0,m} + Y_{m,n} \leq Y_{0,n}$ since the left-hand side is the maximal number of Poisson points along an up-right path that are contained inside $[0,m)^2 \cup [m,n)^2$, which is a subset of the box $[0,n)^2$. This is (almost) the crucial hypothesis for Kingman's subadditive ergodic theorem, stated here for posterity:

Theorem 6. *Let* $(X_{m,n})_{0 \leq m < n}$ *be a family of random variables such that*

(1) $X_{0,n} \leq X_{0,m} + X_{m,n}$ *for all* $m < n$*.*

(2) *For any* $k \geq 1$*, the sequence* $(X_{nk,(n+1)k})_{n=1}^{\infty}$ *is an ergodic stationary sequence.*

(3) *For any* $m \geq 1$*, the two sequences* $(X_{0,k})_{k=1}^{\infty}$ *and* $(X_{m,m+k})_{k=1}^{\infty}$ *have the same joint distribution.*

(4) $\mathbb{E}[|X_{0,1}|] < \infty$*, and there exists a constant* $M > 0$ *such that for any* $n \geq 1$*,* $\mathbb{E}[X_{0,n}] \geq -Mn$*.*

Then

(a) *The limit* $\gamma = \lim_{n\to\infty} \frac{\mathbb{E}[X_{0,n}]}{n}$ *exists and is equal to* $\inf_{n\geq 1} \frac{\mathbb{E}[X_{0,n}]}{n}$*.*

(b) $\frac{X_{0,n}}{n} \to \gamma$ *almost surely.*

The inequality we have is backwards, but we will apply Kingman's theorem to the family $(-Y_{m,n})_{0\leq m<n}$, which satisfies condition (1). Condition (2) is satisfied because the sequence in question consists of i.i.d. random variables. Condiiton (3) is satisfied by translation-invariance of the Poisson point process. Condition (4) is satisfied by a careful application of Proposition 3, since on the event $N([0,n)^2) = k$, the random variable $Y_{0,n}$ has the same distribution as $L(\sigma_k)$, and $N([0,n)^2)$ is a Poisson random variable with mean n^2.

Thus we can apply Kingman's theorem and we find that as $n \to \infty$,

$$\frac{Y_{0,n}}{n} \xrightarrow{a.s.} \Lambda = \sup_{m\geq 0} \frac{\mathbb{E}[Y_{0,m}]}{m} = \lim_{m\to\infty} \frac{\mathbb{E}[Y_{0,m}]}{m}.$$

Note that by monotonicity, the above also holds if we consider $\frac{Y_{0,t}}{t}$, where t is a real-valued parameter. Now we will connect this back to $L(\sigma_n)$ by taking $t = T_n$, where

$$T_n = \inf\left\{t > 0 : \left|\Pi \cap [0,t]^2\right| = n+1\right\}.$$

This ensures that $\Pi \cap [0,T_n)^2$ has exactly n points. By a slight extension of property (5) of the Poisson point process, these points are uniformly distributed in $[0,T_n)^2$, and so Y_{0,T_n} has the same distribution as $L(\sigma_n)$. Therefore, we expect that T_n behaves like $\sqrt{n}$, and indeed this is true. To see why, write

$$T_n^2 = \inf\left\{s > 0 : \left|\Pi \cap [0,\sqrt{s}]^2\right| = n+1\right\}$$

and notice that

$$\left|\Pi \cap [0,\sqrt{s}]^2\right| \stackrel{d}{=} |\pi \cap [0,s]|,$$

where π is a Poisson point process in $\mathbb{R}$. The Poisson point process in $\mathbb{R}$ satisfies all the same conditions as the one in $\mathbb{R}^2$, but it has some important additional structure. Specifically, it has a natural linear ordering, and the spacing between adjacent points is distributed as an Exp(1) random variable. The spacing between different pairs of adjacent points is also independent. Therefore, T_n^2 is a sum of $n+1$ i.i.d. copies of Exp(1), which has mean 1. So by the strong law of large numbers, $\frac{T_n^2}{n} \to 1$ almost surely, and so $\frac{T_n}{\sqrt{n}} \to 1$ almost surely. Finally, observe that

$$\frac{Y_{0,T_n}}{\sqrt{n}} = \frac{T_n}{\sqrt{n}} \frac{Y_{0,T_n}}{T_n} \xrightarrow{a.s.} \Lambda.$$

Since Y_{0,T_n} has the same distribution as $L(\sigma_n)$, this also implies that

$$\frac{L(\sigma_n)}{\sqrt{n}} \xrightarrow{\mathbb{P}} \Lambda.$$

This finishes the proof.

Notice that we cannot conclude almost sure convergence of $\frac{L(\sigma_n)}{\sqrt{n}}$, since we do not know what coupling of σ_n there is between different n; indeed, we have only proved that for a specific coupling (given by the Poisson point process) we have that almost sure convergence.

As a final note, the book claims that some additional work is needed to conclude that

$$\frac{\mathbb{E}[L(\sigma_n)]}{\sqrt{n}} \to \Lambda.$$

I think this should follow from the bounded convergence theorem, but maybe I am missing something.

3 The Constant is 2

To get anywhere in proving that the constant $\Lambda = 2$, we will need to do a lot more work. Since by this point I will probably only have about 30 minutes, I will not be able to get into any of the details, but this section will provide a very high-level overview of the proof. For the details, see Chapter 1 of the book.

3.1 The Robinson-Schensted Algorithm

We will interpret a permutation in a different geometric/combinatorial way. Given a permutation, start writing the numbers in a list of rows. We will always put the next number into the top row. We want the numbers in each row to be in increasing order, so if the number we want to write is smaller than some number in the top row, we "bump" the other number down and replace it with the new number. Then we perform the same game with the old number on the second row, et cetera. Here is an example with our favorite permutation $\sigma = (4, 1, 2, 7, 6, 5, 8, 9, 3)$:

$$\begin{matrix} 4 \end{matrix} \to \begin{matrix} 1 \\ 4 \end{matrix} \to \begin{matrix} 1 & 2 \\ 4 & \end{matrix} \to \begin{matrix} 1 & 2 & 7 \\ 4 & & \end{matrix} \to \begin{matrix} 1 & 2 & 6 \\ 4 & 7 & \end{matrix}$$

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

Notice that the length of the top row is the same as the length of the longest increasing subsequence. Indeed, for any longest increasing subsequence, the next number in the subsequence will always have to be added to the end of the top row. The numbers before it might be replaced, but this shows that the length of the top row is at least as long as the longest increasing subsequence.

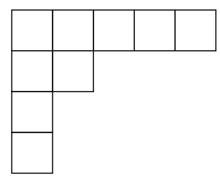
Conversely, we can construct an increasing subsequence with the same length as the top row by letting the last number in the top row be the last number in the subsequence, and then looking at the diagram at the step when the last number in the top row was added and taking the number to the left of it, repeating recursively. In our example, we get $1, 2, 5, 8, 9$.

Now, this process is actually not reversible as is; starting with the final "Young tableau" we cannot recover the permutation; which number was last to be added? However, if we keep track of the order in which positions in the diagram were filled, we can reverse the process. Specifically, if we associate to a permutation *two* different Young tableaux, P and Q, where P is the *insertion tableau* we just discussed and Q is the *recording tableau* of the same shape, we get a one-to-one correspondence between permutations and pairs of Young tableaux. Here is an example of P and Q for our permutation $\sigma = (4, 1, 2, 7, 6, 5, 8, 9, 3)$:

$$P \quad = \quad \begin{array}{|c|c|c|c|c|} \hline 1 & 2 & 3 & 8 & 9 \\ \hline 4 & 5 \\ \cline{1-2} 6 \\ \cline{1-1} 7 \\ \cline{1-1} \end{array} \qquad\qquad Q \quad = \quad \begin{array}{|c|c|c|c|c|} \hline 1 & 3 & 4 & 7 & 8 \\ \hline 2 & 5 \\ \cline{1-2} 6 \\ \cline{1-1} 9 \\ \cline{1-1} \end{array}$$

Since P and Q have the same shape, this is basically an assignment of permutations to Young diagrams, which are just empty Young tableaux, like this:

$$(4, 1, 2, 7, 6, 5, 8, 9, 3) \quad \mapsto$$



We don't need to keep track of the numbers, since the shape of the Young diagram retains all of the relevant information for our study of the longest increasing subsequence. However, we had a uniformly random permutation, but we don't get a uniformly random Young diagram—in fact, the weight of each diagram is proportional to the *square* of the number of ways to fill it with numbers such such that rows and columns are increasing. This is because we have a one-to-one correspondence between permutations and pairs of Young tableaux with the same shape.

For a Young diagram λ with n boxes (denoted $\lambda \vdash n$), let d_λ denote the number of ways to fill it with the numbers from 1 to n such that rows and columns are increasing. Then the probability measure we obtain on diagrams (called the *Plancherel measure*) assigns weight $\frac{d_\lambda^2}{n!}$ to the Young diagram λ.

3.2 Results about Plancherel Measure

On the next few pages, I have drawn two large Young diagrams sampled from the Plancherel measure, one with 1000 boxes and one with 10000 boxes. You will notice that there seems to be a definite shape appearing on the bottom-right edge of the pictures. The precise statement of this phenomenon is the content of the *limit shape* theorem due to Vershik-Kerov and Logan-Shepp.

Theorem 7. *Let $\lambda_n \vdash n$ be a random Young diagram drawn from the Plancherel measure, and let $\operatorname{set}(\lambda_n)$ denote subset of $\mathbb{R}^2$ which is the union of the boxes of λ_n (with the origin at the top-left and the y-coordinate going down in our pictures). Let $\omega = (\omega_x, \omega_y) : [-\pi/2, \pi/2] \to \mathbb{R}^2$ be the planar curve given by*

$$\omega_x(\theta) = \left(\frac{2\theta}{\pi} + 1\right) \sin\theta + \frac{2}{\pi} \cos\theta,$$
$$\omega_y(\theta) = \left(\frac{2\theta}{\pi} - 1\right) \sin\theta + \frac{2}{\pi} \cos\theta,$$

and define

$$\Delta = \{t \cdot \omega(\theta) : t \in [0, 1], \theta \in [-\pi/2, \pi/2]\}.$$

Then for any $\epsilon > 0$,

$$\mathbb{P}\left[(1-\epsilon)\Delta \subseteq \frac{1}{\sqrt{n}} \operatorname{set}(\lambda_n) \subseteq (1+\epsilon)\Delta\right] \longrightarrow 1$$

as $n \to \infty$.

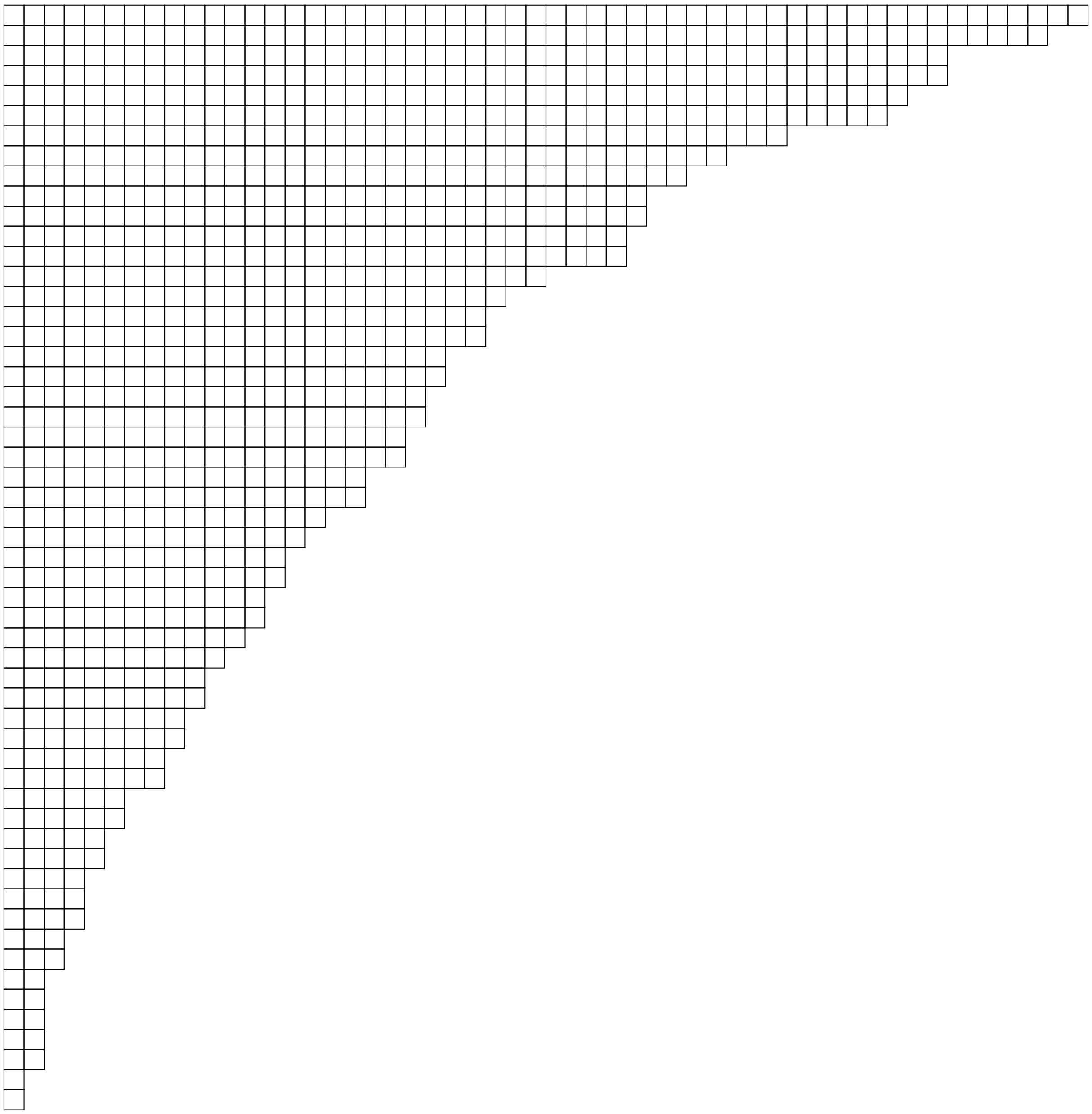


Figure 1: a random Young diagram with 1000 boxes

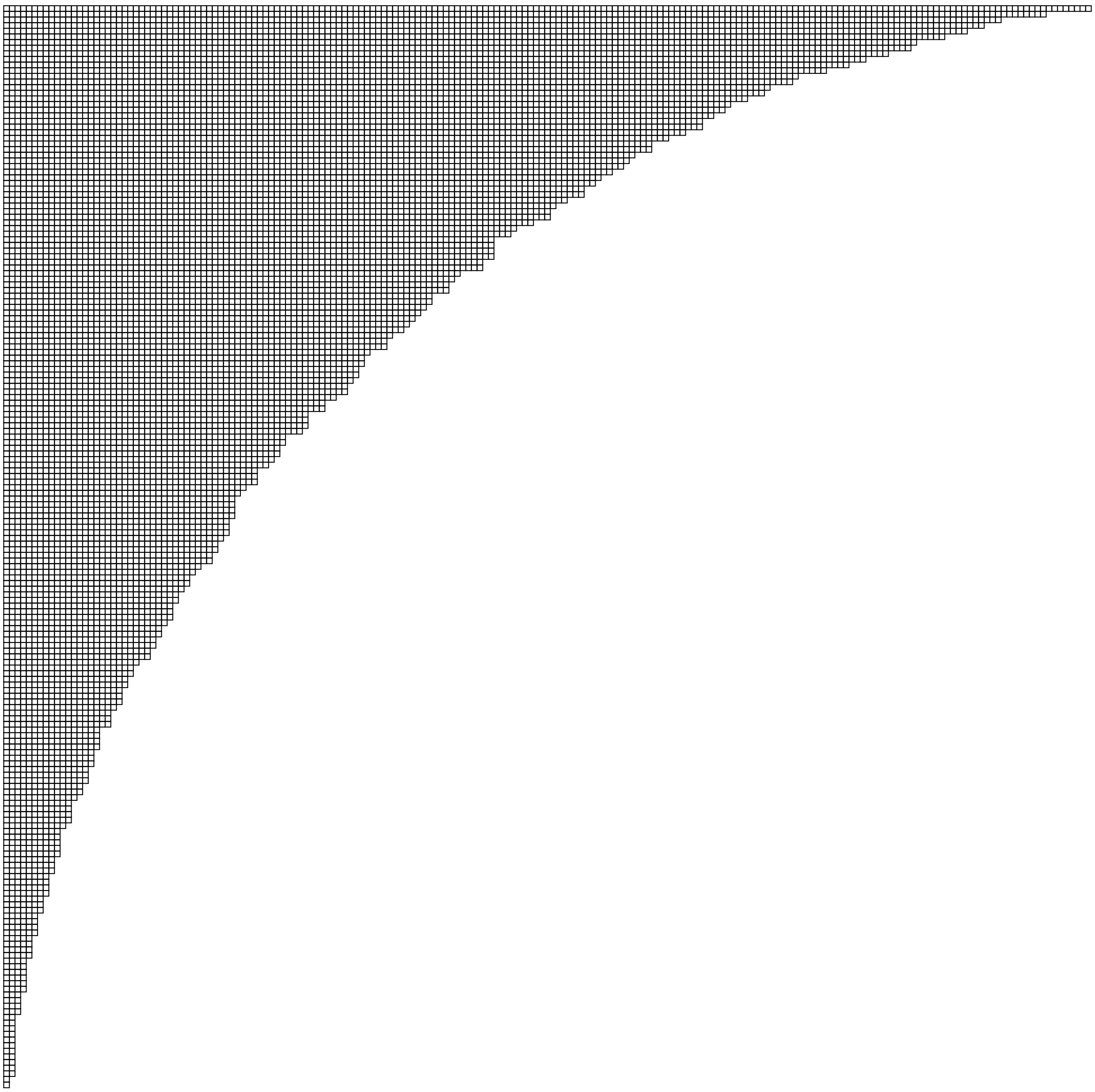
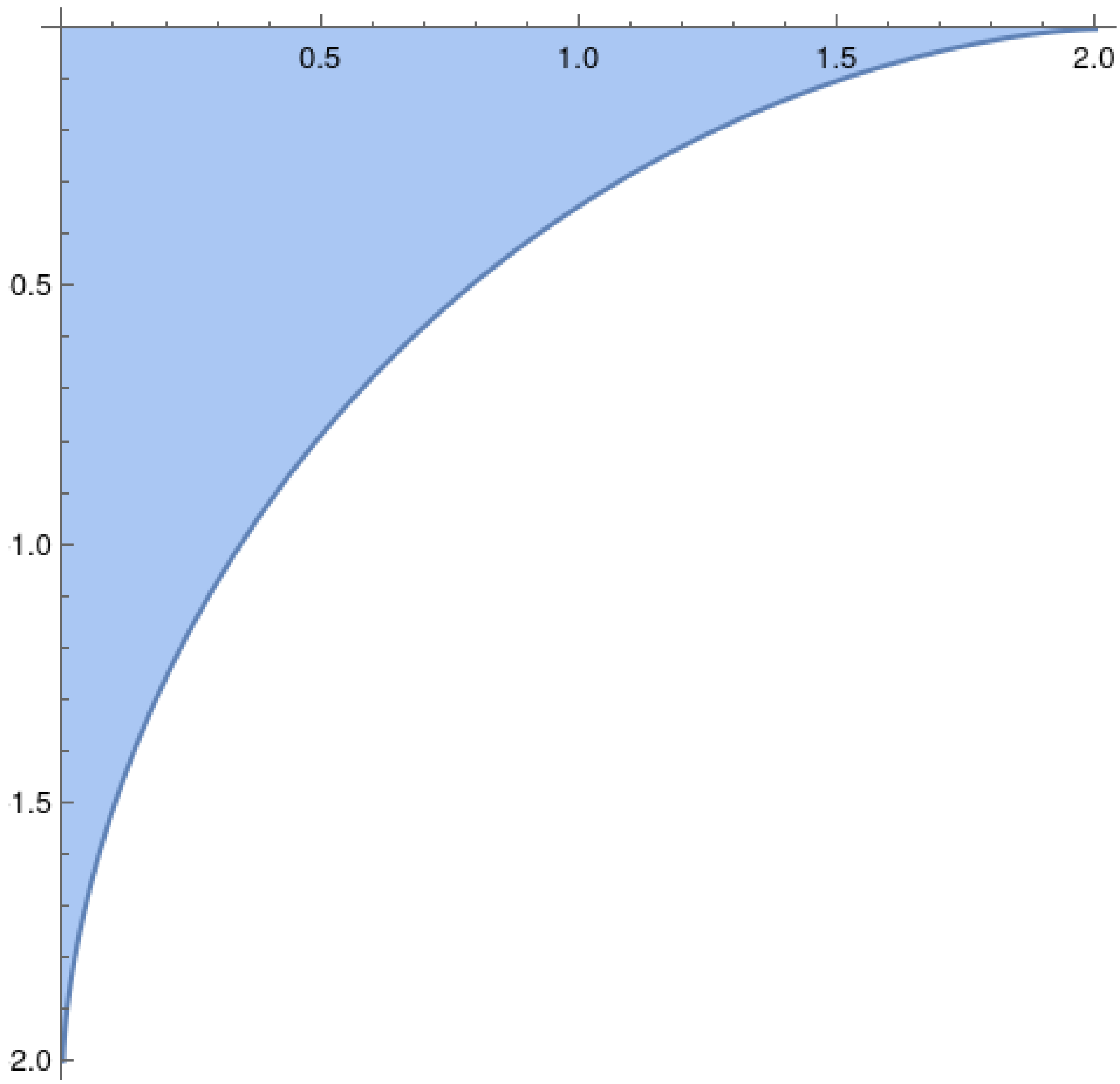


Figure 2: a random young diagram with 10000 boxes

Notice that $\frac{1}{\sqrt{n}}$ is the right scaling to use, because we want the area of the set we are considering to be constant as $n \to \infty$, and the area of $\text{set}(\lambda_n)$ is n. Here is a drawing of the limit shape Δ. Notice that the shape is contained in the box $[0,2]^2$, and the points $(0,2)$ and $(2,0)$ are in the shape. You can check this using the formulas for ω given in the theorem.



This immediately implies that

$$\frac{L(\sigma_n)}{\sqrt{n}} \xrightarrow{\mathbb{P}} 2,$$

which finishes our proof of the LLN-esque result about the length of the longest increasing subsequence in a random permutation. In fact, this result supersedes Hammersley's result (and doesn't depend on it), so we didn't really need to do that, but I think that proof is very elegant, and the proof of the limit shape theorem is much more difficult.

Before we say anything about the proof of the limit shape theorem (which may not happen at all actually), let's see the statement of the CLT-type result that involves the Plancherel measure. In essence, it says that the first few rows of a Plancherel-random Young diagram behave (after rescaling properly) very much like the highest few eigenvalues of a GUE random matrix. This result is due to Baik, Deift, and Johansson.

Theorem 8. *Let $\lambda_n \vdash n$ be a random Young diagram drawn from the Plancherel measure, and for any $i \geq 1$ let $r_i(\lambda_n)$ be the length of the ith row of λ_n. Also denote*

$$\widetilde{r}_i(\lambda_n) = \frac{r_i(\lambda_n) - 2\sqrt{n}}{n^{1/6}}.$$

For any $k \geq 1$, we have the following convergence in distribution:

$$(\widetilde{r}_1(\lambda_n), \dots, \widetilde{r}_k(\lambda_n)) \xrightarrow{d} (a_1, \dots, a_k),$$

where $a_1 > \cdots > a_k$ are the highest k points in the Airy process $\mathcal{A} = \{a_1 > a_2 > \cdots\}$. In particular,

$$\frac{L(\sigma_n) - 2\sqrt{n}}{n^{1/6}} \stackrel{d}{=} \widetilde{r}_1(\lambda_n) \xrightarrow{d} F_2,$$

where F_2 is the GUE Tracy-Widom distribution, which is the law of a_1.

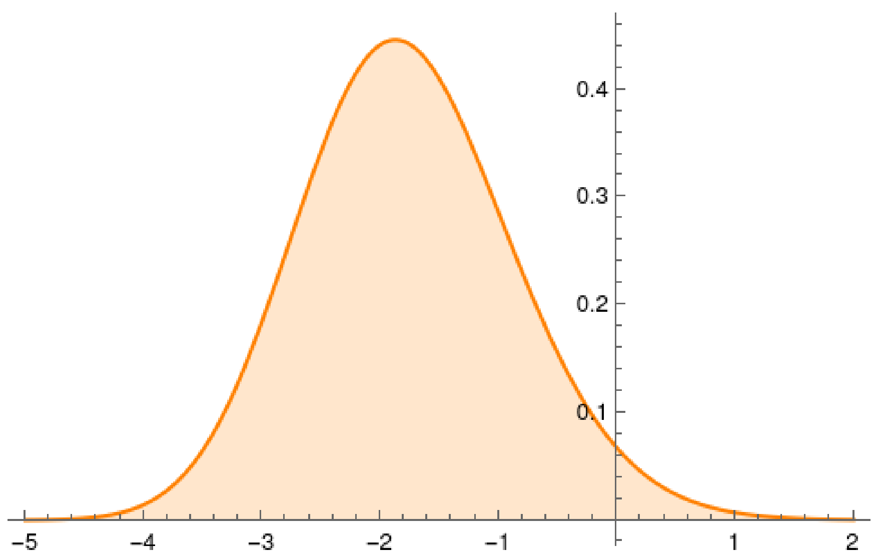
We should say a little bit about what the Airy process actually is. It is the limiting picture of the edge of the set of eigenvalues of a GUE random matrix. More precisely, let X be an $n \times n$ matrix with i.i.d. complex Gaussian entries, let $M = \frac{1}{2}(X + X^*)$, and let $\mu_1 > \cdots > \mu_n$ denote the eigenvalues of M. It is known that $\frac{\mu_1}{\sqrt{n}} \to 2$ in probability (this almost follows from the semicircle law, since 2 is the upper limit of the semicircle distribution).

In fact, for any finite $k \geq 1$, we also have $\frac{\mu_k}{\sqrt{n}} \to 2$ in probability. The Airy process (which has a more complicated definition as a "determinantal point process") arises as the rescaled fluctuations of the μ_i around $2\sqrt{n}$. Namely, for any $k \geq 1$, we have the following convergence in distribution:

$$\left(n^{1/6}(\mu_1 - 2\sqrt{n}), \dots, n^{1/6}(\mu_k - 2\sqrt{n})\right) \xrightarrow{d} (a_1, \dots, a_k),$$

where again $\mathcal{A} = \{a_1 > a_2 > \cdots\}$ is the Airy process. Notice that for the eigenvalues we actually have to *multiply* by $n^{1/6}$, because the spacing between eigenvalues gets smaller as n increases, and we want to retain a macroscopic limit (the Airy process). It's a bit mysterious to me why we get the same exponent $\frac{1}{6}$ in both theorems, when one is being multiplied and one is being divided.

Anyway, the law of a_1 is F_2, the GUE Tracy-Widom distribution, which looks like this:



Its mean is approximately -1.77, and the variance is approximately 0.81. The upper tail is heavier than a Gaussian, and the lower tail is lighter. Specifically,

$$\mathbb{P}[a_1 > s] = \exp\left(-\frac{3}{4}s^{3/2}(1 + o(1))\right) \qquad \text{and} \qquad \mathbb{P}[a_1 < -s] = \exp\left(-\frac{1}{12}s^3(1 + o(1))\right)$$

as $s \to \infty$. This reflects the fact that the highest eigenvalue of a GUE matrix is repulsed by the lower eigenvalues and so it will have an extremely low chance to be very negative, but it will have a better chance than a Gaussian to be very positive.