18.337: Eigenvalues of Random Matrices May 19, 2013

Random partitions in Julia

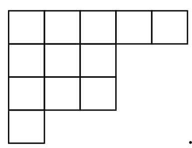
Instructor: Alan Edelman *Francisco Unda*

1 Overview

We will be working with random matrices, i.e. $M = (M_{ij})_{ij}$, where $M_{ij} \sim F_{ij}$ where F_{ij} is some distribution. We will also work with permutations, which will be denoted by σ. This work involves doing computational experiments to verify results that relate the distribution of Plancherel partitions and GUE eigenvalues. Constructing and operating on these objects is usually computationally expensive for the size required, so we need to find a way to efficiently sample them, as well as way of efficiently computing the limiting distributions. Part of the motivation for this work is trying to use the increased efficiency provided by the Julia programming language, and also we wish to find a way to illustrate said result.

1.1 Young Diagrams and SYTs

Given a number $n \in \mathbb{N}$, and a partition $\lambda_1 \geq \lambda_2 \geq \ldots$, the Young diagram associated to $\lambda = (\lambda_1, \lambda_2, \ldots)$ is the set of points (i, j), such that $j \leq \lambda_i$. These are commonly drawn as squares, and in descending order: The partition $(5, 3, 3, 1)$ of 12 is denoted by:



.

The **conjugate** partition λ' of a partition λ is the one obtained by exchanging rows and columns in the Young diagram, in our example the conjugate partition is $(4, 3, 3, 1, 1)$. A Young tableau is a completion of the boxes of a Young diagram with natural numbers, and it is called standard if these numbers are increasing by column and by row. The partition associated to a Young diagram or tableau is called the **shape** of the tableau, and the terms of a partition λ are called its **parts**.

1.2 The RSK algorithm

The Robinson-Schensted-Knuth correspondence is a constructive bijection between permutations of size n and pairs of SYTs of the same shape. Take a permutation $\sigma = (\sigma_1, \ldots, \sigma_n)$. Start with two empty tableaus P and Q, and for each $i = 1, \ldots, n$, insert σ_i the following way: Call $x = \sigma_i$, and starting of the first column, look in P for the smallest number in that column that is bigger than x; if there is none, put x at the end of the column, i in the same position in Q, and go to the next i. If such a number exists, put x in its position, put the number you removed in x, and move to the next column. As an example, the image by this bijection of the permutation (3241) is

1	2	3
4		

1	2	4
3		

.

Note that the operation in the columns can be done with binary search, since the partially built tableau has sorted columns, since it is standard. The algorithm used for generating the P tableau given a permutation can be found in appendix A.

1.3 Plancherel measure

We are interested in the function that associates to each permutation, the shape of the tableaux in the RSK correspondence. This application is clearly not a bijection and we call f_λ the number of SYT of shape λ. Since the RSK correspondence is a bijection, we must have

$$n! = \sum_\lambda f_\lambda^2.$$

Furthermore, if we choose a permutation σ with uniform probability among all permutations of $[n]$, the probability of picking a shape λ is $\frac{f_\lambda^2}{n!}$. This measure over the space of partitions of n is called the Plancherel measure. We will use the RSK correspondence then to sample partitions with this measure, and from now on unless stated otherwise, we will assume λ is drawn from this distribution. On a final note, we recall that f_λ is given by the hook-length formula:

$$f_\lambda = \frac{n!}{\Pi_{x \in T(\lambda)} h(x)}.$$

Where $T(\lambda)$ is the associated Young diagram, and $h(x)$, the hook-length of x is defined to be the number of squares to the right and below x, including x once. As an example if $\lambda = (3,1)$, $f_\lambda = \frac{4!}{4\times2\times1\times1} = 3$, which corresponds to the SYTs

$$\begin{array}{|c|c|c|}\hline 1&2&3\\ \hline 4 \\ \cline{1-1}\end{array}\quad \begin{array}{|c|c|c|}\hline 1&2&4\\ \hline 3 \\ \cline{1-1}\end{array}\quad \begin{array}{|c|c|c|}\hline 1&3&4\\ \hline 2 \\ \cline{1-1}\end{array}.$$

In particular, this formula tells us that $f_{\lambda'} = f_\lambda$, and then if λ is a random permutation following the Plancherel measure, λ' is also distributed the same way.

One of the earliest results related to this distribution are the studies of the longest increasing subsequence of a permutation. Given a permutation σ, an increasing subsequence is a sequence $i_1 < i_2 < \ldots < i_k$ such that $\sigma_{i_1} < \sigma_{i_2} < \ldots < \sigma_{i_k}$. Finding the distribution of the maximum of such lengths is called Ulam's problem. It is easy to prove, by induction for example, that if λ is the shape of the partition induced by the permutation σ, then the length of the longest increasing subsequence is equal to λ'_1. By the symmetry observation in the previous section, the distribution of λ'_1 and λ_1 is the same. It was proven by Vershik and Kerov [2] that

$$\frac{\lambda_1}{\sqrt{n}} \to 2$$

in probability. This is important for technical reasons in the RSK algorithm, since λ_1 and λ'_1 are the number of columns and rows in the corresponding Young tableau, and so will control the size of the data structure used. We use this result then to estimate the required size of the data structure, and we increase it if we need to. The algorithm for sampling partitions is given in appendix B.

2 Okounkov's result and corollary

Okounkov's result [1] is the proof of a conjecture by Baik, Deift and Johansson about the behaviour of these Plancherel partitions. This conjecture states that the sizes of the partition, properly scaled, behave like the eigenvalues of a GUE matrix. More precisely, consider the eigenvalues $E_1 \geq E_2 \geq \ldots$ of a GUE matrix H, in other words $H_{ii} \sim N(0,1)$ and $H_{ij} \sim N(0, \frac{1}{2})$ when $i \neq j$, and define

$$y_i = n^{2/3}\left(\frac{E_i}{n^{1/2}} - 2\right).$$

Now if $\lambda = (\lambda_1, \lambda_2, \ldots)$ is a partition and we define

$$x_i = n^{1/3}\left(\frac{\lambda_i}{n^{1/2}} - 2\right)$$

the conjecture states that the distribution of the x_i's coincides with the distribution of the y_i's in the limit as n goes to infinity. The following theorem proved by Okounkov proves this

Theorem 1 *In the $n \to \infty$ limit, all mixed moments of the random variables $\hat{x}(\xi)$ exist and are identical to those of the $\hat{y}(\xi)$, that is,*

$$\lim_{n\to\infty} \langle \hat{x}(\xi_1) \cdots \hat{x}(\xi_s) \rangle = \lim_{n\to\infty} \langle \hat{y}(\xi_1) \cdots \hat{y}(\xi_s) \rangle,$$

for any $s = 1, 2, \ldots$ and any numbers $\xi_1, \ldots, \xi_s > 0$.

This implies the following

Corollary 2 *In the $n \to \infty$ limit, the joint distribution of $x_1, \ldots, x_k$ is identical to the distribution of $y_1, \ldots, y_k$ for any fixed k.*

3 Numerical evaluation of limit distributions

One last ingredient to use in our numerical experiments is a way to calculate the limit distributions of the x_i's(or y_i's). This will be done using Bornemann's method [3], which involves calculating Fredholm determinants by a discretization. We won't explain this approach in detail, but define

$$E_2^{(n)}(k; J) = \mathbb{P}(\text{exactly } k \text{ eigenvalues of the } n \times n \text{ GUE lie in } J)$$

and

$$E_2(k; J) = \lim_{n\to\infty} E_2^{(n)}(k; \sqrt{2n} + 2^{-1/2} n^{-1/6} J).$$

Finally we define

$$F_2(k; s) = \sum_{j=0}^{k-1} E_2(j; (s, \infty)),$$

the probability that there are less than k eigenvalues greater than s. For example, the Tracy-Widom distribution is given by $F_2(1; s)$, the probability that there are zero eigenvalues greater than s, or equivalently, the probability that the greatest eigenvalue is less than s. The method is based on calculating $F_2(k; s)$ via the formula

$$F_2(k; s) = \sum_{j=0}^{k-1} \frac{(-1)^j}{j!} \frac{d^j}{dz^j} \det\left(I - zK_{Ai}|_{L^2(s,\infty)}\right)|_{z=1},$$

where K_{Ai} is the Airy Kernel

$$K_{Ai}(x, y) = \frac{Ai(x)Ai'(y) - Ai'(x)Ai(y)}{x - y}.$$

4 Numerical Experiments

The following experiments were conducted on a Windows machine, with an Intel Core i7-3610QM CPU @ 2.30 Ghz. Plancherel partitions were generated, and the first three parts were recorded, for sizes $n = k^6$, where $k = 3, 4, 5, 6$. The longest experiment, $k = 6$ took approximately 11 hours in Julia. The reason for having powers of size, is that it makes the scaling a division by an integer, and so the histogram bins are easily calculated by $y = min(x) : (1/n^{1/6}) : max(x)$, where x is the scaled vector of samples. Graphs are plotted in Matlab, with histograms normalized by $[f, z] = hist(x, y); bar(z, f/trapz(z, f))$.

4.1 Largest Eigenvalue

The first computation was a simple one, we study the largest part of the partition, and compare it to the largest eigenvalue of the GUE. The corresponding limiting distribution is the Tracy-Widom distribution. Figure 1 shows the histogram of 10^4 samples with $k = 5$.

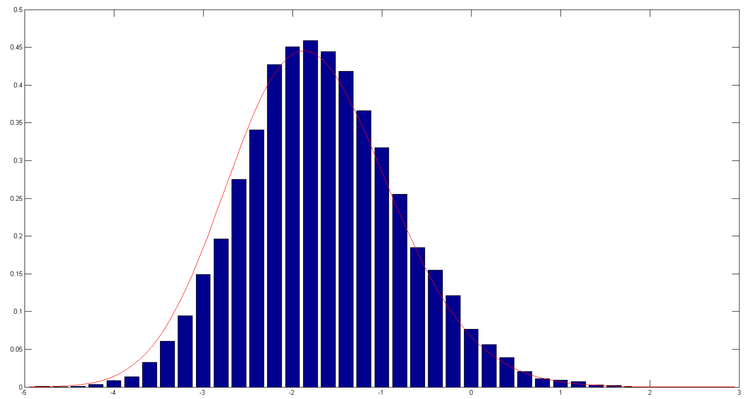


Figure 1: Largest part of plancherel partition histogram and plot of Tracy-Widom distribution.

4.2 Substraction of the first and second eigenvalues

We computed the value $r = \lambda_1 - \lambda_2$ from our samples, and the corresponding distribution

$$G(s) = P(\lambda_1 - \lambda_2 \leq s) = \int_{-\infty}^{\infty} (\partial_x F(t,t) - \partial_x F(t, t-s))\, dt$$

where F is the joint distribution of (λ_1, λ_2), was computed using F by numerical integration. Figure 2 shows $k = 5$ with 10^4 samples.

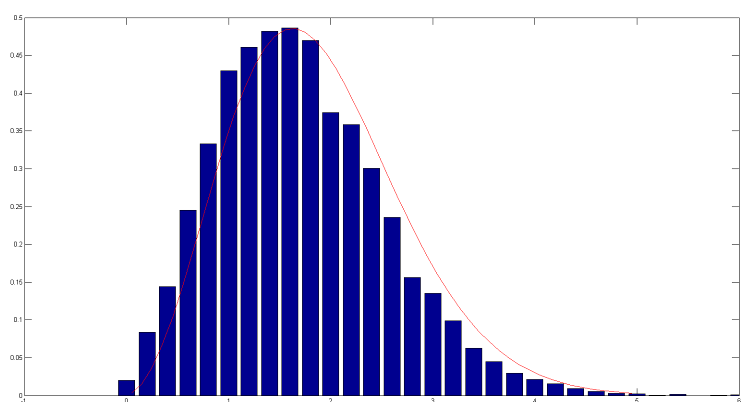


Figure 2: Histogram of the substraction of the two largest parts of a Plancherel partition and plot of the corresponding limiting distribution.

4.3 Joint distribution of the first and second eigenvalues

We plotted a two dimensional histogram of the first and second parts of the Plancherel partition with the surface $z = \partial_x \partial_y F(x, y)$. Figure 3 shows $k = 6$, and 10^4 samples.

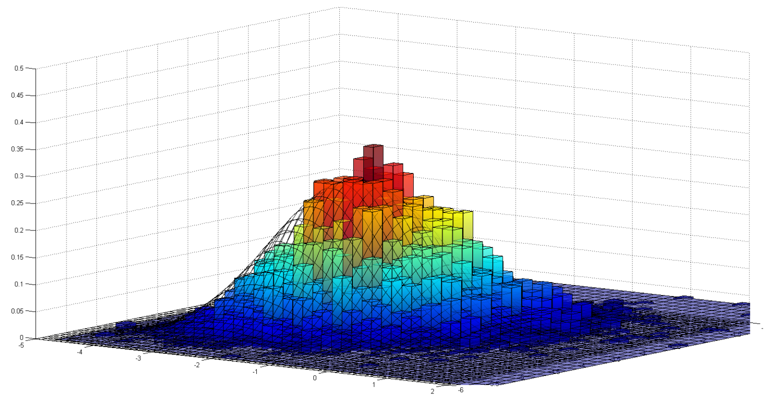


Figure 3: Two dimensional histogram of the two largest part of plancherel partition, along with the limiting distribution of the first eigenvalues of a GUE matrix.

4.4 Other experiments

In general we can see in these experiments that Okounkov's result is verified, although the slow convergence might mean that there is something to be said about the discrepancy. It might also be a consequence of the fact that the support of the scaled Plancherel parts is a lattice whose precision grows as $n^{1/6}$. All the code, and data used to produce these experiments is commented and adjoined.

A RSK algorithm

The following is an algorithm in Julia that computes the SYT P, given a permutation.

```
# rsk.jl

function rsk(seq)
## Input: seq ... a sequence of whole numbers (all >= 0)
## Output: A partition P capturing large scale sorting structure

n = length(seq)
m1 = convert(Int,round(2*sqrt(n)))
m2 = convert(Int,round(2*sqrt(n)))
P = NaN*ones(m1,m2)
for i=1:n
    new = seq[i]
    for j=1:n
        if j > m2
            P = [P NaN*ones(m1,m2)]
            m2 = 2*m2
        end
        k = 1
        while P[k,j] <= new
            k+=1
            if(k > m1)
                P  = [P ; NaN*ones(m1,m2)]
                m1 = 2*m1
            end
        end
        old = P[k,j]
        P[k,j] = new
        new = old
        if isnan(new)
            break
        end
    end
end
    return P
end
```

B Plancherel partition algorithm

This algorithm returns a sample of a partition of n following the Plancherel measure, given n.

```
# plancherelPartition.jl
require("rsk.jl")
require("randperm2.jl")
function plancherelPartition(n)
## Input: n...size of partition
## Output: lam...random partition of n(with plancherel measure)
p = randperm2(n)
P = rsk(p)
(u,v) = size(P)
lam = zeros(v)
for i=1:v
    if isnan(P[1,i])
        lam = lam[1:i-1]
        break
    end
    for j=1:u
        if isnan(P[j,i])
            break
        end
        lam[i] += 1
    end
end
return lam
end
```

References

[1] A. Okounkov, *Random Matrices and Random Permutations*, 2000

[2] A. Vershik and S. Kerov, *Asymptotics of the maximal and typical dimension of irreducible representations of symmetric group*, Func, Anal. Appl., **19**, 1985, no.1.

[3] F. Bornemann, *On the numerical evaluation of distributions in random matrix theory: A review*, 2000