Analyzing Higher Order Effects on Eigenvalues

Course 18.338 Project Report

Saaketh Vedantam
saakethv@mit.edu

Abstract

The statistical behavior of the spacing between eigenvalues for random Gaussian ensembles is well known. This distribution is known as the so-called "Wigner surmise" and is theorized to apply to many different classes of random matrices. One intuition for this may come from the Central Limit Theorem. I explore how higher order moments of underlying distributions, such as variance and kurtosis, influence the distribution of such statistics. In particular, I examine two problems: the problem of eigenvalue spacing and maximal eigenvalues.

1 Introduction

The convergence of eigenvalues to the Wigner semicircle is well known. Similarly, mathematicians have studied the distribution of eigenvalue spacings and extreme eigenvalues. Of course, these must be normalized in the appropriate way in order to ensure that the resulting distribution is meaningful.

1.1 Motivation

Liu [2001] cites that for a large class of random matrices, the distribution of eigenvalue spacings approaches the Wigner surmise: $P(s) = As\exp(-Bs^2)$ for certain values of A and B. For the Gaussian Orthogonal Ensemble, this distribution follows $P(s) = \frac{\pi}{2}s\exp\left(-\frac{\pi}{4}s^2\right)$. However, in order to reach that result, we need to answer a few questions.

How do we define eigenvalue spacing? At the highest level, we start with a symmetric $n \times n$ random matrix, evaluate the eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, and evaluate the distribution of $\lambda_{i+1} - \lambda_i$. Intuitively, this feels off. I confirm this by plotting the raw eigenvalue spacings for various values of n in Figure 1. We can see that the spacings simply go to 0, which is not an interesting result.

First, we need to normalize the matrix in some way. Although the eigenvalue distribution approaches a semicircle $p_r(\lambda) \propto \sqrt{r^2 - x^2}$, depending on the scale of the matrix, the value of r can vary greatly. This isn't too difficult to fix; in fact, I normalize every matrix so that the eigenvalues will be concentrated in $[-1, 1]$. Second, the spacing distribution actually isn't fixed! Empirical evidence suggests that for eigenvalues normalized in the range $[-1, 1]$, the edge eigenvalues are spaced apart on the order of $n^{-1/6}$, while the ones in the center are closer to $n^{-1/2}$. Since the goal of this analysis is to see limiting behavior as $n \to \infty$, we don't want to observe different distributions for $\lambda_2 - \lambda_1$ and $\lambda_{100} - \lambda_{99}$. To this end, I "unfold" the eigenvalues.

1.2 Unfolding Eigenvalues

Since the eigenvalue distribution is not uniform, the spacings between consecutive eigenvalues won't be distributed according to the same distribution. How do we homogenize the spacing distribution? We make the distribution uniform! It is a well known result in probability theory that if X is a random variable with CDF F, then $F(X)$ is distributed according to $\mathcal{U}([0, 1])$. The following is a proof of this fact.

$$\Pr[F(X) \leq c] = \Pr[X \leq F^{-1}(c)] = F(F^{-1}(c)) = c$$

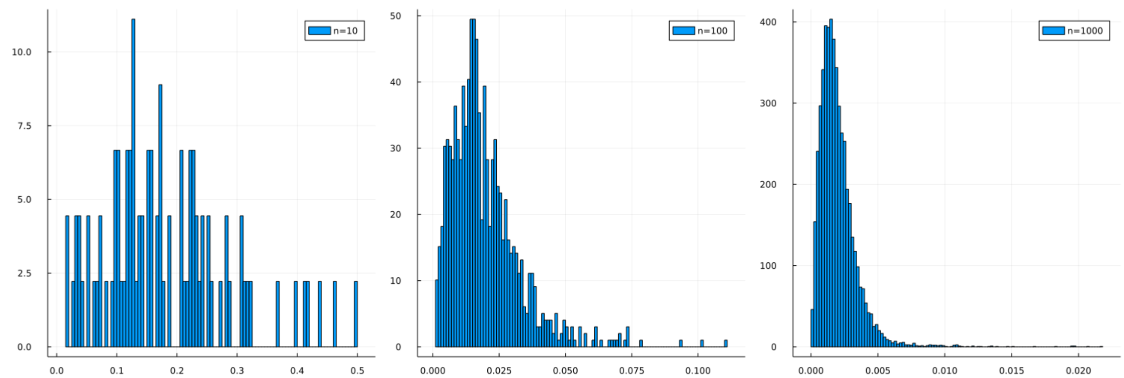

Figure 1: Raw spacings. As n increases, they're increasingly concentrated at 0.

where the first equality follows since F is monotonically increasing. The PDF of the eigenvalues is approximately semicircular. This follows $p_w(x) = \frac{2}{\pi}\sqrt{1-x^2}$. Therefore, the CDF is

$$\int_{-1}^{\lambda} \frac{2}{\pi}\sqrt{1-x^2}\,\mathrm{d}x = \frac{1}{2} + \frac{\lambda\sqrt{1-\lambda^2} + \lambda \sin^{-1}\lambda}{\pi}$$

Next, we look at the spacings in $\mathcal{U}([0,1])$. Let v_k be the kth smallest value out of n randomly chosen values in this range. Then it is well known that $E[v_k] = \frac{k}{n+1} \implies E[v_k - v_{k-1}] = \frac{1}{n+1}$. Finally, we have a value that's independent of k. However, this still goes to 0. We remedy this by multiplying by $n+1$, giving a final transformation of

$$F_1(\lambda) = \begin{cases} 0 & \lambda < -1 \\ (n+1)\left[\frac{1}{2} + \frac{\lambda\sqrt{1-\lambda^2}+\lambda\sin^{-1}\lambda}{\pi}\right] & -1 \leq \lambda < 1 \\ n+1 & \lambda \geq 1 \end{cases}$$

I placed the 1 subscript because this formula actually depends on β. For my entire analysis, however, I use $\beta = 1$ because it's the simplest and I don't think plotting the other values will add much insight.

At this point, we have a function such that $F_1(\lambda_k) - F_1(\lambda_{k-1})$ is distributed identically for all k, with mean 1. At this point, one might think that I removed all the challenge from the problem because this new value is distributed uniformly. To disprove this, I first plotted the density of spacings for n values selected from $\mathcal{U}([0, n+1])$. This is seen in Figure 2. As we see, the limiting distribution approaches the exponential density $P(s) = e^{-s}$.

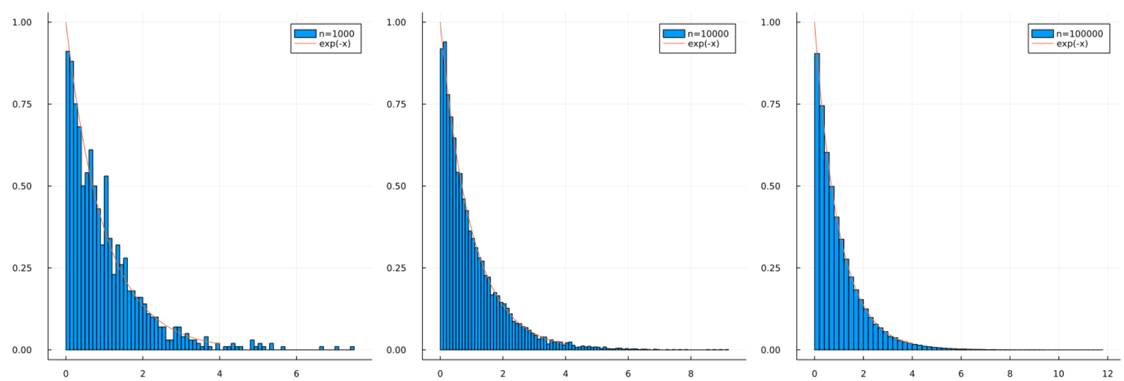

Figure 2: Spacings for values selected uniformly between 0 and $n+1$

Next, I plotted the distribution of unfolded eigenvalue spacings for the GOE (Figure 3). As we see, this actually approaches the Wigner surmise, not an exponential density. This suggests that unfolding isn't the whole story. Rather, the distinction between the eigenvalue distribution for finite n and for infinite n is enough for the limiting spacing distributions to differ. One explanation for this could be that the higher order moments of the Gaussian perturb the eigenvalues enough to alter their spacing. It is this hypothesis that I want to test.

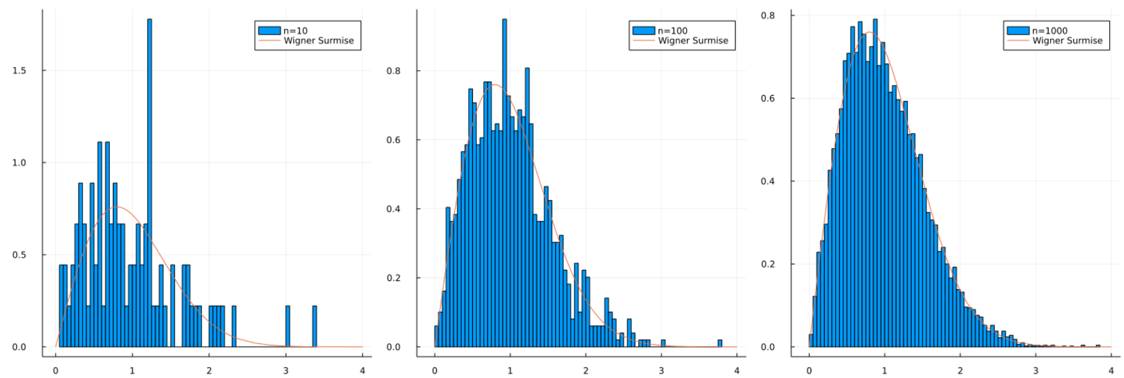

Figure 3: Spacings for unfolded eigenvalues sampled from GOE

Another interesting point covered by Ben Arous and Bourgade [2013] is the fact that the eigenvalues seem to "repel" each other; i.e. the probability that they are very close approaches 0. This might diverge from our intuition, which suggests that the spacings decrease as shown in Figure 2. In any case, this isn't the focus of this paper, but I felt that it's worth mentioning for future research.

2 Eigenvalue Spacing

Above, I exemplified the Wigner surmise for a very specific class of random matrices, namely the GOE. In order to assess higher order effects, I will carry out the same process for various other distributions. The goal will be to observe a difference in the spacing distribution. If this isn't possible, which is likely since most literature agrees that the Wigner surmise is a pretty good approximation for any random matrices, I will also look at first order errors, or errors that depend on $\frac{1}{n}$.

2.1 Procedure

First, I generate a random matrix A where each entry is i.i.d from some distribution f. Then, to make it symmetric, I calculate $\frac{A+A^T}{2}$. Another way to make it symmetric would be to calculate AA^T, but these result in singular values, which have their own interesting properties. Afterwards, I compute the eigenvalues λ_i and their unfolded forms $F_1(\lambda_i)$. With these spacings, I plot the histogram to see its similarity to the surmise. Since these are likely to be close, I also plot the first order error, defined as follows. Let $p(s)$ be the spacing distribution for finite n.

$$\text{Assumption}: p_n(s) = \frac{\pi}{2} s \exp\left(-\frac{\pi}{4}s^2\right)\left[1 + \frac{1}{n}e(s)\right] + O\left(\frac{1}{n^2}\right)$$

Under this assumption, the distribution of spacings does approach the surmise, but for finite n, there are error terms that depend on $1/n$. We can continue the Laurent series to get higher order errors, but for this analysis I'm primarily interested in the first order. We can reorganize this to get

$$e(s) \approx n\left[\frac{p_n(s)}{\frac{\pi}{2}s\exp\left(-\frac{\pi}{4}s^2\right)} - 1\right]$$

In my experiment, I replace $p_n(s)$ with $\hat{p}_n(s)$, the empirical PDF as determined by a sample of spacings.

2.2 Constraints

Let X be a random variable sampled from distribution f. If $\mathbb{E}_f[X] \neq 0$, then $\mathbb{E}_f[\text{Tr}\, A] \neq 0 \implies \mathbb{E}_f[\sum_{\lambda_i}] \neq 0$. However, the limiting distribution of eigenvalues is still a semicircle, which has average value 0. What ends up happening is that $n-1$ eigenvalues are distributed on the semicircle, while the last one takes all the weight to bring the mean to its expected value. This outlier could mess up experiments. Of course, a simple fix would be to remove the outlier. I avoid this issue entirely by restricting the distribution f so that $\mathbb{E}_f[X] = 0$.

The next issue is normalization. If we simply find the eigenvalues of $\frac{A+A^T}{2}$, these will grow as n grows, and the unfolding process won't work as expected. To this end, we divide by $\sqrt{2n}$, similar to the GOE. There's still an issue of

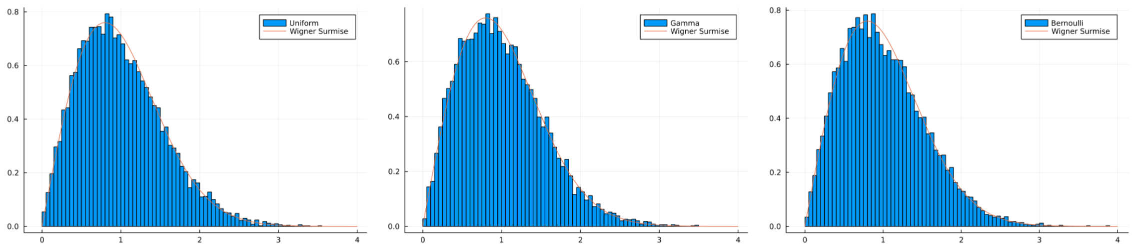

Figure 4: Comparison of various distributions to surmise. For these plots, $n = 1000$ with 30 trials.

variance, however. For distributions with higher variance, the semicircle will have a higher radius that scales with the standard deviation. To avoid this, I only deal with distributions f such that $\mathrm{Var}_f[X] = 1$. Then, the matrix we want is $\frac{A+A^T}{2\sqrt{2n}}$.

With these constraints in place, I defined three distributions:

- Uniform: $f = \mathcal{U}([-\sqrt{3}, \sqrt{3}])$
- Gamma: $f = \Gamma(1, 1)$, subtract 1 from entries to make mean 0
- Bernoulli: 1 with probability 1/2, -1 with probability 1/2

Distribution	Mean	Variance	Kurtosis
Uniform	0	1	-1.2
Gamma	0	1	6
Bernoulli	0	1	-2

As shown above, the mean and variance of these are the same, but the kurtoses differ. I define the kurtosis as $\mathbb{E}_f[X^4] - 3$. This is valid since the means are 0. I subtract 3 so that the kurtosis of a Gaussian is 0. Notably, the kurtosis of the Gamma distribution has a different sign than that of the other two.

2.3 Results

As shown in Figure 4, the spacing distribution still looks like the Wigner surmise. This is to be expected, since mathematicians know they're close. The first order errors are shown in Figure 5.

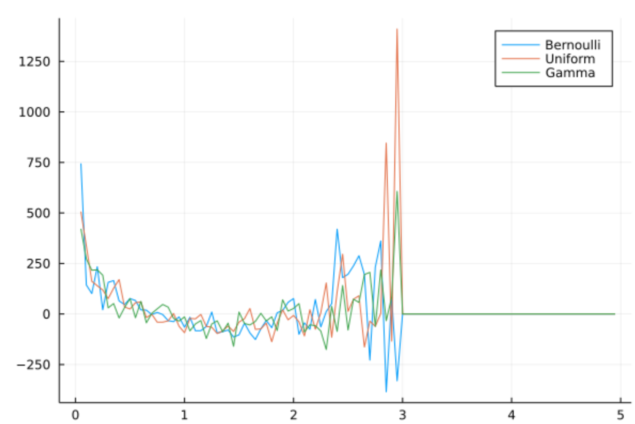

Figure 5: First order errors for each distribution

Unfortunately, it's difficult to see a relationship with the first order errors. There's a lot of noise, but most of it is concentrated at 0. The errors do spike at the tails of the distribution, but I believe this simply occurs since the Wigner surmise goes to 0 at these two ends. If even 1 sample gave an eigenvalue spacing in this range, the ratio of the empirical PDF to the surmise would be very high. It's also possible that I didn't use a high enough n. Due to an inefficient algorithm or my limited compute resources, I could only run $n = 1000$ in a realistic amount of time.

3 Maximum Eigenvalue

The next set of experiments I ran related to the problem of finding a distribution for the maximum eigenvalue. Again, this is a well known problem with a limiting distribution of the Tracy-Widom distribution. The most common occurrence of Tracy-Widom is $\beta = 2$, for the GUE. However, I continue to work with $\beta = 1$.

Most of the procedure here is the same as above, but a key difference is normalization. Here, I calculate the matrix $\frac{A+A^T}{2}$. The largest eigenvalue, which I call λ_{max}, is around $\sqrt{2n}$. Similar to above, this poses a convergence problem if we use the raw value. Therefore, we normalize according to the following formula:

$$T(\lambda) = (\lambda - \sqrt{2n})\sqrt{2}n^{1/6}$$

Tracy-Widom is defined as the theoretical limiting distribution of $T(\lambda_{\text{max}})$ for the GOE. We observe the empirical distribution of λ_{max} for the GOE and Tracy-Widom in Figure 6.

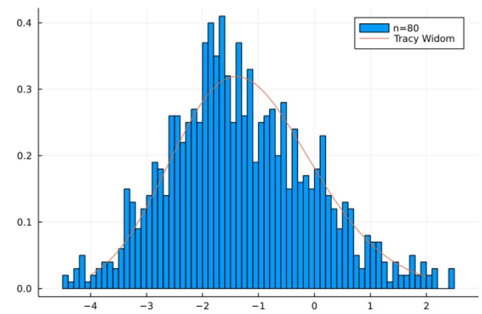

Figure 6: Visualization of basic Tracy-Widom

The distribution looks slightly off, since there's more weight in the left half than there should be. I believe this is due to the small value of n. Due to limited compute resources, I kept n at 80 and decided to go for 10000 data points. Increasing n any further would've required me to reduce the sample, which I believe would be worse for the model. I believe this skewness is caused by the value of n because as I raised n from 20 to 40 to 80, the PDF got increasingly closer. Of course, there remains the possibility that it converges to something slightly off, but since this is a proven theoretical result, I trust that this is not an issue.

I also changed the calculation of the first order error. Since the Tracy Widom distribution doesn't have a nice exponential form, I modeled the error as $p_n(s) = tw(s) + \frac{1}{n}e(s) + O\left(\frac{1}{n^2}\right) \implies e(s) \approx n(p_n(s) - tw(s))$. Instead of dividing, I subtracted the two distributions to get an error term.

3.1 Results

As shown in Figure 7, the λ_{max} distribution looks somewhat like the Tracy-Widom distribution. Notably, the Gamma plot is skewed a different way from the other two. This is an encouraging sign because as mentioned before, the Gamma distribution has a positive kurtosis while the other two have a negative kurtosis. The first order errors are shown in Figure 8.

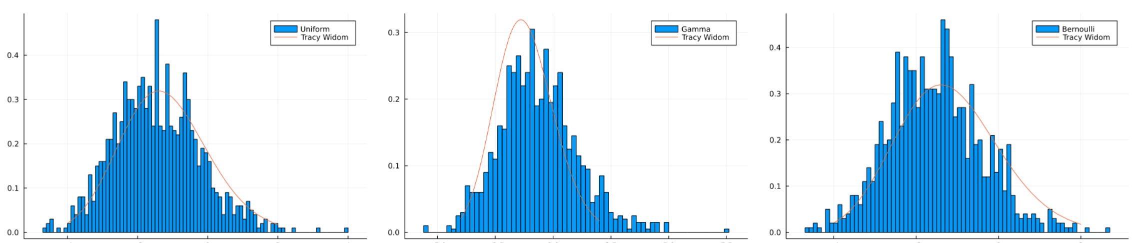

Figure 7: Comparison of various distributions to surmise. For these plots, $n = 1000$ with 30 trials.

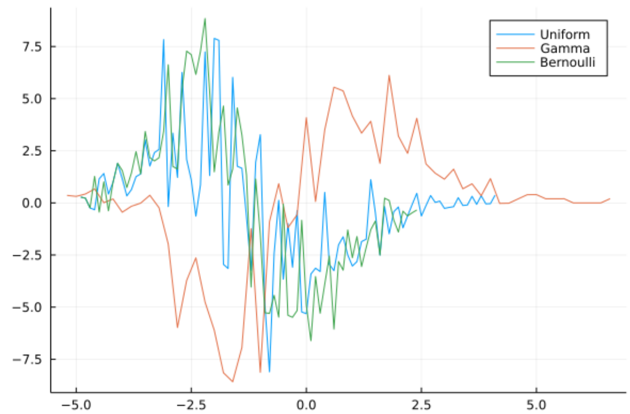

Figure 8: First order errors for each distribution

Again, there's no clear correlation between the magnitude of the kurtosis and the size of the first order error. However, I noted that there's a clear change in the sign. The Gamma distribution has a noticeably flipped error when compared to the other two. The shape is still similar, perhaps providing evidence that this is a good way to model the error function. Due to all the noise, it's difficult to actually come up with a function to approximate the error. However, this is promising news that one does exist and has the form $\kappa e(s)$, where κ is kurtosis.

4 Conclusion

After studying the Wigner semicircle law, it's easy to become complacent and assume that we can generalize many results about arbitrary random matrices. However, these two problems indicate that even simple-looking idea have complex solutions. While the local eigenvalue statistics above have been studied widely in literature, neither have nice results for a general random matrix. Other authors try to assess similar patterns in matrices. Edelman et al. [2016] ran experiments on the distribution of the minimum singular value of symmetric matrices and found that the first order errors do follow a pattern. I actually obtained motivation from this paper to extend the results to extreme eigenvalues. From the start, I believed that extreme eigenvalues had a better change of producing interesting results than eigenvalue spacing, since the literature supports this. However, I had hoped to see some correlation in the first order errors for eigenvalue spacing. My primary unanswered question is the reason for the Wigner surmise. Why do raw spacings sampled from a uniform distribution produce an exponential distribution, while unfolded spacings form the surmise? The distinction would make sense if higher order moments were involved, but we see that the surmise still exists for unfolded eigenvalues from any distribution.

Ultimately, I believe that some parts of my process were misguided. At first, I assumed that the above distinction had to mean that differences in finite distributions and infinite ones played a key role in the problem. While this may be the case, it certainly doesn't carry over to differences in finite matrices with distinct underlying distributions. The Central Limit Theorem seems to be robust enough to handle higher order moments. It's quite difficult to obtain any results mathematically, so the best chance I had way to observe unusual characteristics in the histograms. However, my methodology was flawed, either due to the low dimensionality of my data or due to an incorrect error model. I used the error model from Edelman et al. [2016], but it's feasible that others exist. In any case, the work done here shows that there are a lot of questions waiting to be answered in the field of determining universal random eigenvalue statistics.

Acknowledgements

I would like to thank Prof. Edelman for teaching such an interesting course and offering advice on this project. I would also like to thank the TAs for helping out on Piazza.

References

Y. Liu. Statistical behavior of the eigenvalues of random matrices. Spring 2001. URL http://web.math.princeton.edu/mathlab/projects/ranmatrices/yl/randmtx.PDF.

Gérard Ben Arous and Paul Bourgade. Extreme gaps between eigenvalues of random matrices. *The Annals of Probability*, 41(4), Jul 2013. ISSN 0091-1798. doi:10.1214/11-aop710. URL `http://dx.doi.org/10.1214/11-AOP710`.

Alan Edelman, A. Guionnet, and S. Péché. Beyond universality in random matrix theory. *The Annals of Applied Probability*, 26(3), Jun 2016. ISSN 1050-5164. doi:10.1214/15-aap1129. URL `http://dx.doi.org/10.1214/15-AAP1129`.