FSTM: Bachelor of Mathematics

Supervisor: Simon Campese

Random matrices

Tim Seuré, Max Thill

Experimental mathematics 4

Winter semester 2020/21

Abstract

We analyse some special random matrices and the distribution of the corresponding eigenvalues as the matrices' dimensions tend towards infinity. This will be done from an experimental point of view with the use of SageMath.

Based on the observations we can make from our experiments' outcomes, we shall establish some conjectures.

Contents

1 Marčenko–Pastur Law

In this section we consider random matrices X of the form

$$X = (x_{ij}) \in \mathcal{M}_{N\times M}(\mathbb{R}),$$

whose entries x_{ij} are all independent and identically distributed (i.i.d.), as well as the corresponding random matrices

$$Y := X \cdot X^\top \in \mathcal{M}_N(\mathbb{R}).$$

N and M are natural numbers that should be thought of as tending towards infinity.

Let $\lambda_1 \ldots, \lambda_N \in \mathbb{C}$ be the N eigenvalues of Y. Roughly speaking, our goal will be to predict the distribution of these N eigenvalues of Y.

Let us start by noticing that the λ_i are actually all non-negative, real numbers. This claim is justified by the following proposition.

Proposition 1. *Let A be a real, symmetric matrix. Then the eigenvalues of A are real. Moreover, if A equals $B \cdot B^\top$ for a real matrix B, then the eigenvalues of A are non-negative real numbers.*

Proof. Let $\lambda \in \mathbb{C}$ be an eigenvalue of A, corresponding to a complex, non-zero eigenvector v. We denote the norm of v by $|v|$. We obtain

$$\begin{aligned}
\left(\lambda - \overline{\lambda}\right) |v|^2 &= \left(\lambda - \overline{\lambda}\right) \overline{v}^\top v \\
&= \left[\lambda\overline{v}^\top - \overline{\lambda}\overline{v}^\top\right] v \\
&= \left[\overline{v}^\top \lambda - \overline{\lambda v}^\top\right] v \\
&= \left[\overline{v}^\top A - \overline{Av}^\top\right] v && (\text{as } \lambda v = Av) \\
&= \left[\left(A^\top \overline{v}\right)^\top - (A\overline{v})^\top\right] v \\
&= \left[(A\overline{v})^\top - (A\overline{v})^\top\right] v && (\text{as } A^\top = A) \\
&= 0,
\end{aligned}$$

and since $|v|^2 > 0$, we conclude that $\lambda - \overline{\lambda} = 0$, that is, λ is real.

Now, assume A equals $B \cdot B^\top$ for some real matrix B. Let λ be an eigenvalue of A. Clearly $A^\top = A$, therefore λ is real by above. Next, let v be a non-zero eigenvector corresponding to λ. Since A and λ are real, we can also assume that v is real. Now

$$\begin{aligned}\lambda|v|^2 &= \lambda v^\top v \\ &= v^\top \lambda v \\ &= v^\top A v \\ &= v^\top B B^\top v \\ &= \left(B^\top v\right)^\top \left(B^\top v\right) \\ &= \left|B^\top v\right|^2 \\ &\geq 0.\end{aligned}$$

We showed $\lambda|v|^2 \geq 0$. As v is a non-zero vector, $|v|^2$ is positive. We can conclude that $\lambda \geq 0$, as desired. $\square$

1.1 When the dimensions' ratio tends towards a constant

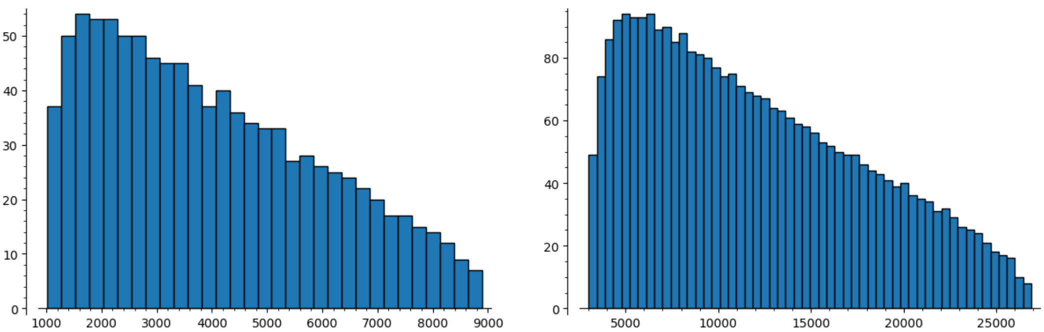
We have already mentioned above that we want to make N and M tend towards infinity. Firstly, we will assume that M and N tend to infinity at roughly the same speed, in the sense that

$$\frac{M}{N} \to c$$

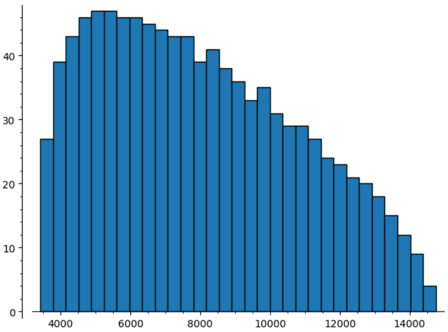
as $M, N \to \infty$, where c is some positive real number. This can be obtained by making M depend on N, for example by choosing $M = \lfloor cN \rfloor$. This choice of M will hold throughout this whole section 1.1.

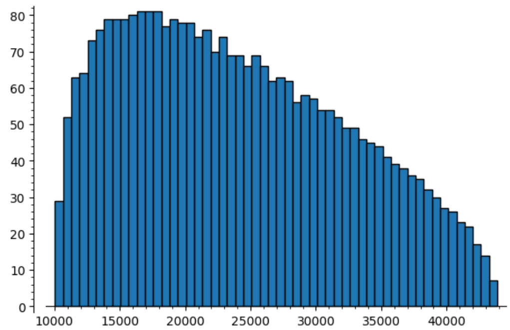
What we now do is generate some random matrices X, then compute the eigenvalues of $Y = X \cdot X^\top$ and represent them in a histogram. This is possible because Y is a symmetric matrix, and therefore only has real eigenvalues. Below the first examples we briefly explain how a histogram should be read.

For now, we choose the entries x_{ij} of X to be standard normally distributed (that is, $x_{ij} \sim \mathcal{N}(0,1)$) and select a constant c equal to 4, as well as N equal to 1 000 respectively 3 000. Recall that we want to simulate $N \to \infty$, and should therefore choose great values for N. The corresponding histograms are the following:



Observe that even though we chose different values for N, the histogram's shapes remain roughly the same. The same phenomenon can be observed with the following histograms, corresponding to the value $c = 8$ and $N = 1\,000, 3\,000$.





In section 1.1.2, we will study more closely the histograms' shapes. For the moment, let us only predict that it can be described solely in terms of c. We will later on see that the variation σ^2 of the x_{ij} also plays a (minor) role.

Let us now explain how to read one of above's histograms. For this, take a look at the first example. The initial bin (bar) has a height of 37, and ranges on the x-axis over the interval $[1\,000, 1\,250]$ (roughly). This means that 37 out of Y's 1 000 eigenvalues λ_i lie in the mentioned interval. The remaining bins can be interpreted analogously.

Notice that none of the eigenvalues of Y is negative, as we deduced from proposition 1. However, considering the histograms, the lower bound of 0 is not very satisfying: The eigenvalues actually lie far away from 0, and their distance to the origin appears to increase as N gets bigger. Driven by this dissatisfaction, we seek to find an estimation for the extreme (smallest and greatest) eigenvalues of Y.

1.1.1 Estimating the extreme eigenvalues

We shall start by estimating Y's greatest eigenvalue $\lambda_{\max}$. For this, let us first fix the constant c and vary N. As above, we choose the x_{ij} to be standard normally distributed for our current tests. The following gives an insight on the dependence of $\lambda_{\max}$ on N, where c has (arbitrarily) been chosen to equal 4.

N	$\lambda_{\max}$
500	4 514.1
1 000	9 005.1
1 500	13 393.0
2 000	17 999.7
2 500	22 441.3
3 000	26 969.4

We notice that $\lambda_{\max}$ is roughly proportional to N. This observation can also be made when choosing different values for c. Phrased differently, $\frac{\lambda_{\max}}{N}$ should not depend on N anymore. Just to be on the safe side, let us check that this assumption holds, this time choosing $c = 8$, and slightly different values for N.

N	$\frac{\lambda_{\max}}{N}$
3 000	5.79
5 000	5.82
7 000	5.81

Indeed, our conjecture appears to be legitimate (just keep in mind that this is of course still not a valid proof). For the following, we will therefore not analyse $\lambda_{\max}$ anymore, but rather $\frac{\lambda_{\max}}{N}$, while choosing some fixed value for N, say $N = 2\,000$.

Next, we vary c.

c	$\frac{\lambda_{\max}}{N}$
1	3.96
2	5.82
3	7.48
4	8.97

One might observe that inputting perfect squares c results in $\frac{\lambda_{\max}}{N}$ being close to a perfect square as well. Let us further investigate.

c	$\frac{\lambda_{\max}}{N}$	$\left\lceil \frac{\lambda_{\max}}{N} \right\rfloor$
1^2	3.96	2^2
2^2	8.97	3^2
3^2	16.02	4^2
4^2	24.94	5^2

The last column rounds the values of the second column to the nearest integer. We shall make the following observation: If $c = a^2$, then $\frac{\lambda_{\max}}{N} \approx (a+1)^2$. In other words,

$$\frac{\lambda_{\max}}{N} \approx \left(\sqrt{c}+1\right)^2,$$

that is,

$$\lambda_{\max} \approx N\left(\sqrt{c}+1\right)^2.$$

To see that this approximation holds even for values of c that are not perfect squares, let us choose $c = 5$ and $N = 3\,000$. We predict that $\lambda_{\max}$ lies close to $3\,000\left(\sqrt{5}+1\right)^2 \approx 31\ 416$. Indeed, the experimental value equals approximately 31 341, which corresponds to a relative mistake of around 0.25%.

With exactly the same methods as above, we can also estimate the difference in the extreme eigenvalues, that is, $\lambda_{\max} - \lambda_{\min}$, where $\lambda_{\min}$ is, as expected, the smallest eigenvalue of Y. We obtain the approximation

$$\lambda_{\max} - \lambda_{\min} \approx 4N\sqrt{c}.$$

Isolating $\lambda_{\min}$ in above estimation leads to

$$\begin{aligned}\lambda_{\min} &\approx \lambda_{\max} - 4N\sqrt{c}\\ &\approx N\left(\sqrt{c}+1\right)^2 - 4N\sqrt{c}\\ &= N\left(\sqrt{c}-1\right)^2.\end{aligned}$$

Let us summarize our observations so far. The greatest respectively smallest eigenvalue of Y are approximated as follows:

$$\lambda_{\max/\min} \approx N\left(\sqrt{c}\pm 1\right)^2.$$

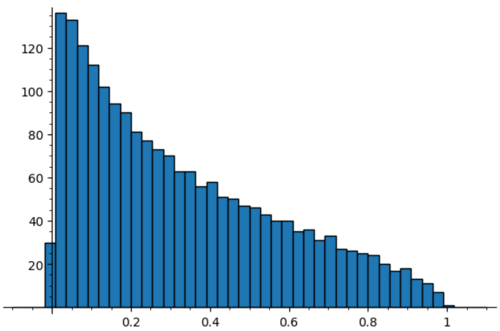
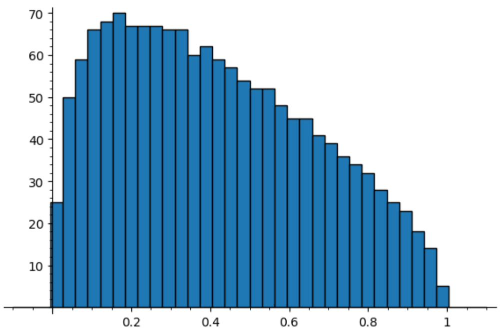
Keep in mind that this is valid for when the x_{ij} are standard normally distributed. We will generalise this result later in 1.1.3.

1.1.2 Describing the histograms' shapes

For this section, let us again choose the x_{ij} to be standard normally distributed. Our goal will be to describe the shape of the histograms by the curve of a function. For this, let us recall that the eigenvalues of Y (mostly) lie in the range $[\lambda_{\min}, \lambda_{\max}]$. If we subtract this interval's smallest value from Y's eigenvalues, then divide the obtained values by the length of the interval, we obtain values in the range $[0, 1]$, while not changing the overall shape of the corresponding histogram (since the eigenvalues are only transformed by an affine function). For simplicity, let us give these values a special name.

Definition 2. If λ is an eigenvalue of Y, we call $\frac{\lambda - N\left(\sqrt{c}-1\right)^2}{4N\sqrt{c}}$ a *normalized eigenvalue* of Y.

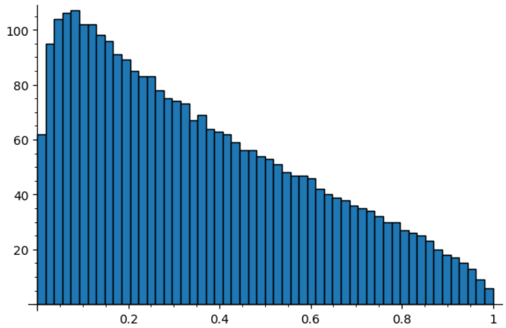
We shall take a look at the histograms of the normalized eigenvalues of Y. Below are examples for $(N, c) \in \{(2\,000, 2), (1\,500, 8)\}$.

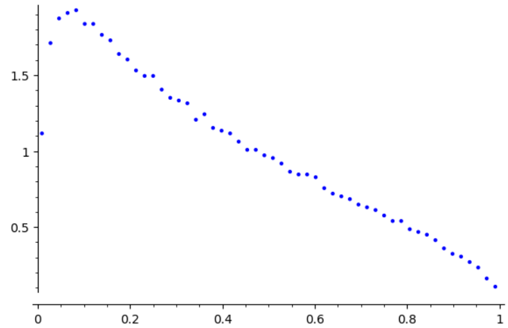



As expected, the histograms are mostly contained between $x = 0$ and $x = 1$.

Next, in order to simplify the problem even further, we want the histograms to have a constant area equal to 1. For this, it is useful to know that in all of above examples, the histograms consisted of $\lfloor\sqrt{N}\rfloor$ bins, where $\lfloor\cdot\rfloor$ denotes the floor function. Hence, every normalized eigenvalue of Y contributes to an area of $\frac{1}{\lfloor\sqrt{N}\rfloor}$, that is, 1 divided by the width of a bin in the histogram. Knowing that Y has N eigenvalues, we conclude that the histograms corresponding to the normalized eigenvalues of Y have a total area of $N \cdot \frac{1}{\lfloor\sqrt{N}\rfloor} = \frac{N}{\lfloor\sqrt{N}\rfloor}$. Therefore, by dividing the height of each bin by $\frac{N}{\lfloor\sqrt{N}\rfloor}$, the obtained histograms has an area of 1, as desired.

Equivalently, we can transform the histograms of Y's normalized eigenvalues into a set of points (namely the two-dimensional points given by the centre of the top segment of each bin), then divide the height of these points by $\frac{N}{\lfloor\sqrt{N}\rfloor}$ to obtain a list $L(Y)$. Here is an example where we start with a histogram of Y's normalised eigenvalues (left), then transform it into the corresponding list $L(Y)$, whose points have been plotted (right).



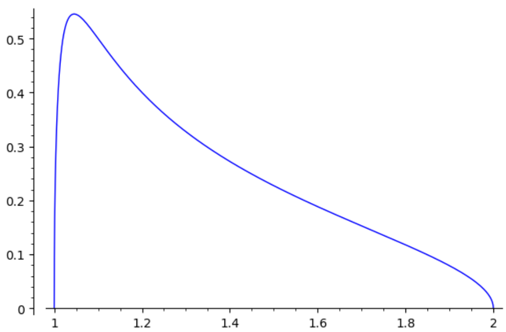


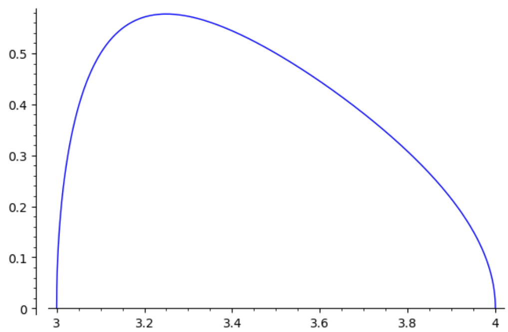
The next step is to notice that the obtained points appear to lie on the curve of a function

$$x \mapsto \frac{\sqrt{P_2(x)}}{P_1(x)},$$

where $P_2(x)$ is a polynomial of degree 2, and $P_1(x)$ is a polynomial of degree 1.

For example, let us consider the curves of the functions $x \mapsto \frac{\sqrt{-x^2+3x-2}}{4x-\frac{19}{5}}$ respectively $x \mapsto \frac{\sqrt{-x^2+7x-12}}{x-\frac{5}{2}}$.





Both appear to have great similarities with above histograms (or equivalently, with the corresponding plots of $L(Y)$).

Therefore, let us assume that the points of $L(Y)$ indeed lie on the curve of a function of the form $x \mapsto \frac{\sqrt{P_2(x)}}{P_1(x)}$. Notice that our histograms suggest that these curves should have vertical half-tangents at $x = 0$ and $x = 1$, that is, $P_2(x)$ should have $x = 0$ and $x = 1$ as zeroes, or equivalently, $P_2(x) = \alpha x(1-x)$ for some $0 \neq \alpha \in \mathbb{R}$. Notice that $\alpha > 0$, otherwise $\sqrt{P_2(x)} = \sqrt{\alpha x(1-x)}$ wouldn't be defined for $0 < x < 1$. Now

$$\frac{\sqrt{P_2(x)}}{P_1(x)} = \frac{\sqrt{\alpha}\sqrt{x(1-x)}}{P_1(x)} = \frac{\sqrt{x(1-x)}}{\alpha^{-1/2}P_1(x)},$$

where the denominator is again just a polynomial of degree 1. Therefore, we can assume that the points of $L(Y)$ lie on the curve of a function $x \mapsto \frac{\sqrt{x(1-x)}}{P_1(x)}$, where $P_1(x) = ax + b$ for some $a, b \in \mathbb{R}, a \neq 0$.

At this point, we want to mention that a, b and $P_1(x)$ of course depend on the parameter c (as we can deduce from the histograms). Therefore, let us change the notation slightly: Write a_c and b_c instead of a and b, and

$$g_c(x) := \frac{\sqrt{x(1-x)}}{a_c x + b_c}.$$

Remember that our goal is to find an exact expression for $g_c(x)$ in dependence of $c > 0$. For this, it would be useful if we only had to work with a single parameter, instead of two parameters a_c and b_c.

Fortunately, this can be done by using the following. We transformed the histograms so that the range on the x-axis corresponds to the interval $[0, 1]$, and the new area equals 1. This means that the points of $L(Y)$ should lie on a curve g_c whose integral from 0 to 1 equals 1, that is,

$$\int_0^1 g_c(x)\,\mathrm{d}x = 1.$$

Under assumption that $0 < a_c < 1$ and $b_c > 0$ (which turns out to be alright for most cases), we can use SageMath to get an exact expression of the left-hand side of above equality, namely,

$$\int_0^1 g_c(x)\,\mathrm{d}x = \frac{\pi}{2a_c^2}\left(a_c + 2b_c - 2\sqrt{(a_c + b_c)b_c}\right).$$

As this integral should be equal to 1, we obtain an easy expression for b_c in terms of a_c.

Proposition 3. *If* $\int_0^1 \frac{\sqrt{x(1-x)}}{a_c x + b_c}\,\mathrm{d}x = 1$ *for* $0 < a_c < 1$ *and* $b_c > 0$*, then*

$$b_c = \frac{(\pi - 2a_c)^2}{8\pi}.$$

Proof. We set the expression given by SageMath equal to 1, and solve it for b_c:

$$\begin{aligned}
&\frac{\pi}{2a_c^2}\left(a_c + 2b_c - 2\sqrt{(a_c + b_c)b_c}\right) = 1 \\
\Longrightarrow\ & a_c + 2b_c - 2\sqrt{(a_c + b_c)b_c} = \frac{2a_c^2}{\pi} \\
\Longrightarrow\ & 2\sqrt{(a_c + b_c)b_c} = a_c + 2b_c - \frac{2a_c^2}{\pi} \\
\Longrightarrow\ & 4a_c b_c + 4b_c^2 = a_c^2 + 4b_c^2 + \frac{4a_c^4}{\pi^2} + 4a_c b_c - \frac{4a_c^3}{\pi} - \frac{8a_c^2 b_c}{\pi} \\
\Longrightarrow\ & 0 = 1 + \frac{4a_c^2}{\pi^2} - \frac{4a_c}{\pi} - \frac{8b_c}{\pi} \\
\Longrightarrow\ & 8\pi b_c = \pi^2 + 4a_c^2 - 4\pi a_c \\
\Longrightarrow\ & 8\pi b_c = (\pi - 2a_c)^2 \\
\Longrightarrow\ & b_c = \frac{(\pi - 2a_c)^2}{8\pi},
\end{aligned}$$

as claimed. □

Therefore, we can now write

$$g_c(x) = \frac{\sqrt{x(1-x)}}{a_c x + \frac{(\pi-2a_c)^2}{8\pi}}.$$

Our goal is now to understand how a_c depends on c.

Using the *method of least squares* for curve fitting and SageMath, we now try to find a_c so that $g_c(x)$ expresses the list of points $L(Y)$ best. Note that the list $L(Y)$ of course also depends on c, but we omitted the index for ease of notation.

The following table outputs the approximated values of a_c in dependence on c, where we chose $N = 2\,000$. Note that we have already discussed above that a_c shouldn't depend on N anymore, so that the exact value for N doesn't play a role.

We first start with a table for $c \in \{1, \ldots, 10\}$:

c	1	2	3	4	5	6	7	8	9	10
a_c	1.54	1.09	0.90	0.78	0.70	0.64	0.59	0.55	0.52	0.49

Below is the corresponding table for $c \in \{11, \ldots, 20\}$:

c	11	12	13	14	15	16	17	18	19	20
a_c	0.47	0.45	0.43	0.42	0.40	0.39	0.38	0.37	0.36	0.35

Notice that multiplying c by 4 divides a_c by 2, multiplying c by 9 divides a_c by 3, and so on. This suggests that a_c is proportional to $\frac{1}{\sqrt{c}}$, that is, $a_c = \frac{\alpha}{\sqrt{c}}$ for some fixed constant α. In order to obtain the value for α, we shall consider the following table.

c	1	2	3	4	5	6	7	8	9	10
$a_c\sqrt{c}$	1.54	1.54	1.55	1.56	1.57	1.57	1.57	1.57	1.57	1.57

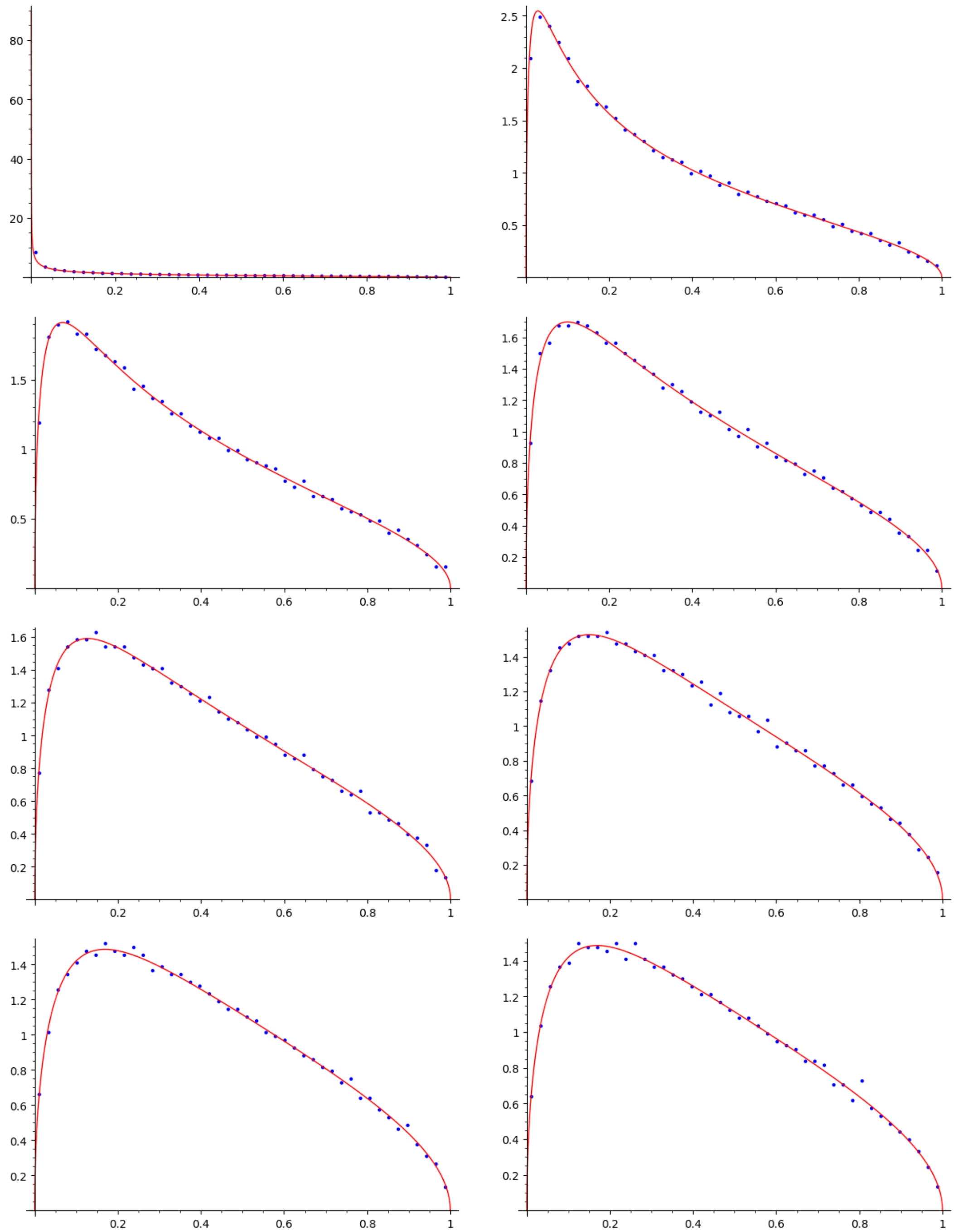
π already having appeared in the expression of b_c, we conjecture that $a_c\sqrt{c} = \frac{\pi}{2}$ (which equals approximately 1.57). In other words, we assume

$$a_c = \frac{\pi}{2\sqrt{c}},$$

and therefore

$$\begin{aligned} g_c(x) &= \frac{\sqrt{x(1-x)}}{\frac{\pi}{2\sqrt{c}}x + \frac{\left(\pi - 2\frac{\pi}{2\sqrt{c}}\right)^2}{8\pi}} \\ &= \frac{8c}{\pi} \cdot \frac{\sqrt{x(1-x)}}{4x\sqrt{c} + (\sqrt{c}-1)^2}. \end{aligned}$$

We will now see that $x \mapsto g_c(x)$ can indeed describe the list $L(Y)$ with great success. The following plottings use $N = 2\,000$, and we selected constants $c = 1, 2, \ldots, 8$. The first line of plottings corresponds to $c = 1$ and $c = 2$, the second one to $c = 3$ and $c = 4$, and so on.



Next, let us extend g_c to the whole real line, by defining $g_c(x) = 0$ for any $x \in \mathbb{R} - [0, 1]$. Notice that then, $\int_0^x g_c(t)\,\mathrm{d}t$ is, by construction, a good estimation for the ratio

$$\frac{\text{number of normalized eigenvalues of } Y \text{ below } x}{N}$$

for any $x \in \mathbb{R}$. Phrased differently, we have

$$N \int_0^x g_c(t)\,\mathrm{d}t \approx \#\left\{1 \le i \le N : \frac{\lambda_i - N\left(\sqrt{c}-1\right)^2}{4N\sqrt{c}} \le x\right\}.$$

Using this, we can now find an estimation for the number of eigenvalues of Y which lie below any given $x \in \mathbb{R}$:

$$\begin{aligned}
\#\{1 \le i \le N : \lambda_i \le x\} &= \#\left\{1 \le i \le N : \frac{\lambda_i - N\left(\sqrt{c}-1\right)^2}{4N\sqrt{c}} \le \frac{x - N\left(\sqrt{c}-1\right)^2}{4N\sqrt{c}}\right\} \\
&\approx N \int_0^{\frac{x-N\left(\sqrt{c}-1\right)^2}{4N\sqrt{c}}} g_c(t)\,\mathrm{d}t \\
&= \frac{8Nc}{\pi} \int_0^{\frac{x-N\left(\sqrt{c}-1\right)^2}{4N\sqrt{c}}} \frac{\sqrt{t(1-t)}}{4t\sqrt{c}+\left(\sqrt{c}-1\right)^2}\,\mathrm{d}t.
\end{aligned}$$

This is a way to describe the distribution of the eigenvalues of Y. We have therefore reached the original goal that we set ourselves.

The whole phenomenon is described more precisely by the "Marčenko-Pastur law." One might be interested in reading more about this in [3].
It turns out that our results are exact for $c > 1$, but need to be slightly adjusted when $0 < c \le 1$. We will however not go into more detail concerning the case $0 < c \le 1$, as the differences between the theoretical results and our experimental observations are hard to make visible.

1.1.3 Dependence on distribution

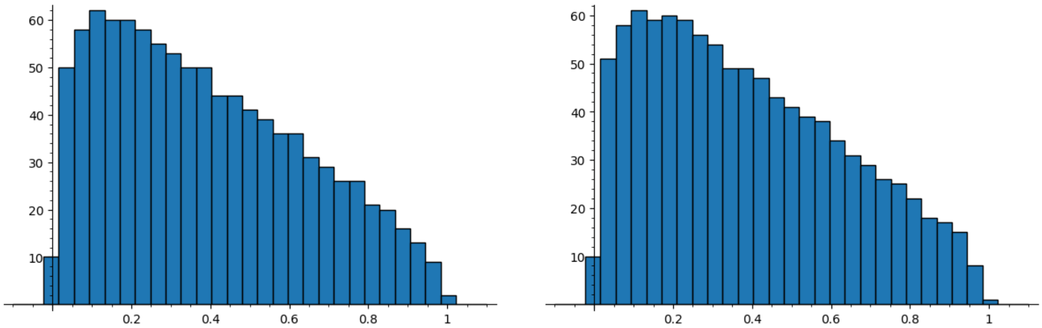
For now, we have always assumed that the entries x_{ij} of X are standard normally distributed. What happens if we change their variance, mean, or even the entire distribution?

It turns out that the only relevant information about the x_{ij} is their variance σ^2. This means that

- changing the mean of the x_{ij}, or more generally
- changing the distribution of the x_{ij}

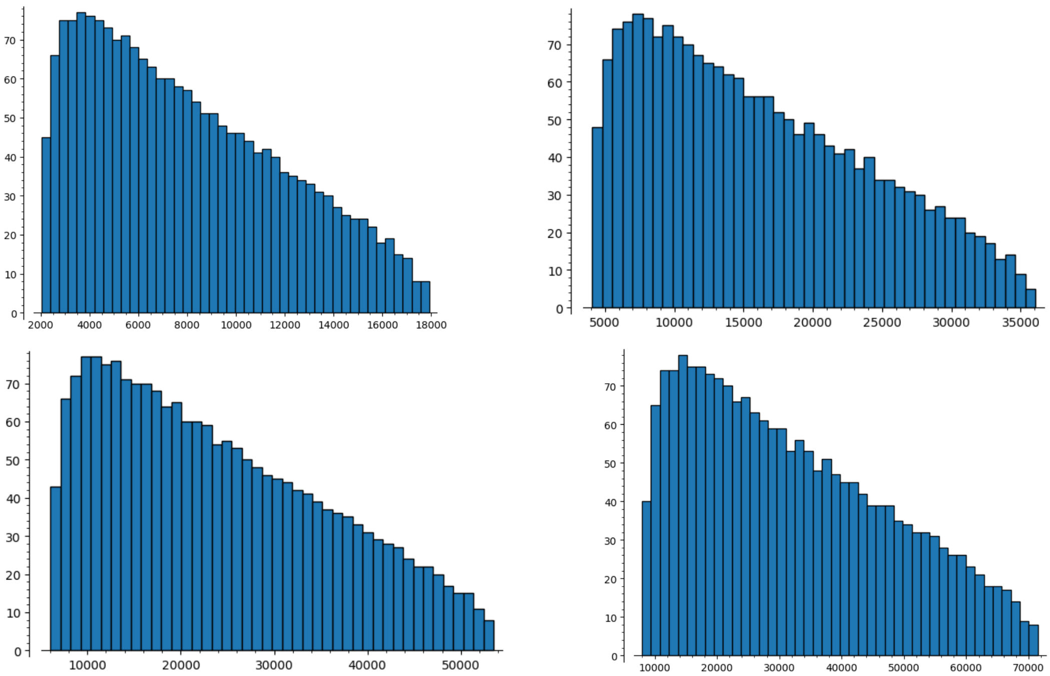
doesn't influence the histogram of Y's eigenvalues, as long as the variance of the x_{ij} remains unchanged.

For example, compare the following two histograms of Y's normalized eigenvalues. For the first one, we chose the x_{ij} to be Poisson-distributed with parameter $\lambda = 1$ (which implies a variance of $\sigma^2 = 1$ and a mean of $\mu = 1$); for the second one, we chose the x_{ij} to be standard normally distributed (with variance $\sigma^2 = 1$ and mean $\mu = 0$). We selected $N = 1\,000$.



No noteworthy difference can be observed, as described above. Many other tests showed similar results.

Next, we are interested in changing the variance. For this, consider the histograms of Y's eigenvalues, where we choose the x_{ij} to be normally distributed with mean $\mu = 0$ and variances $\sigma^2 = 1$ (top left), $\sigma^2 = 2$ (top right), $\sigma^2 = 3$ (bottom left), $\sigma^2 = 4$ (bottom right). We select $N = 2\,000$ and $c = 4$, but similar observations hold for other values of course.



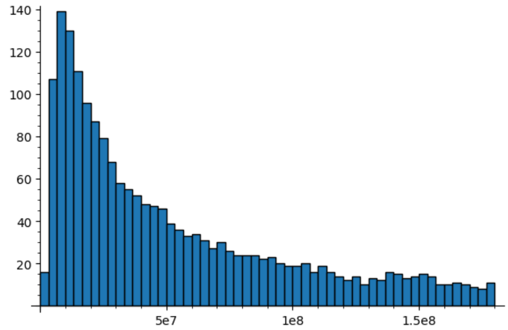
Observe that the overall shape remains the same, while $\lambda_{\min}$ and $\lambda_{\max}$ change. Notice that $\lambda_{\max/\min}$ appear to be proportional to σ^2. This leads to a more general definition of normalized eigenvalues for Y.

Definition 4. If λ is an eigenvalue of $Y = XX^\top$, and X's entries have a variance of σ^2, we call $\frac{\lambda - \sigma^2 N(\sqrt{c}-1)^2}{4\sigma^2 N\sqrt{c}}$ a *normalized eigenvalue* of Y.

This also leads to a more general approximation of the number of eigenvalues of Y lying below some $x \in \mathbb{R}$:

$$\#\{1 \leq i \leq N : \lambda_i \leq x\} \approx \frac{8Nc}{\pi} \int_0^{\frac{x-\sigma^2 N(\sqrt{c}-1)^2}{4\sigma^2 N\sqrt{c}}} \frac{\sqrt{t(1-t)}}{4t\sqrt{c} + (\sqrt{c}-1)^2} \,\mathrm{d}t.$$

It remains to see what happens when the x_{ij} don't have a variance. Let us choose the x_{ij} to be standard Cauchy distributed. This makes sure that the variance of the x_{ij} is inexistent. Below we choose $c = 3$ and $N = 3\,000$.



This suggests that the eigenvalues of Y are in this case not bounded and their distribution cannot be described using above methods.

To finish the section 1.1, we shall recapitulate our results in a final conjecture.

Conjecture 5. *Let $X = (x_{ij}) \in \mathcal{M}_{N\times M}(\mathbb{R})$ be a random matrix with i.i.d. entries x_{ij} of variance σ^2. Let $\lambda_1, \dots, \lambda_N$ be the eigenvalues of $Y = XX^\top$. If $M, N \to \infty$ such that $\frac{M}{N} \to c$ for $c \in \mathbb{R}_{>}0$, then*

$$\#\{1 \leq i \leq N : \lambda_i \leq x\} \approx \frac{8Nc}{\pi} \int_0^{\frac{x-\sigma^2 N(\sqrt{c}-1)^2}{4\sigma^2 N\sqrt{c}}} \frac{\sqrt{t(1-t)}}{4t\sqrt{c} + (\sqrt{c}-1)^2}\, \mathrm{d}t.$$

1.2 When the dimensions' ratio tends towards infinity or zero

We shall now be interested in what happens when $\frac{M}{N} \to \infty$ respectively $\frac{M}{N} \to 0$. For this, we should have a closer look at

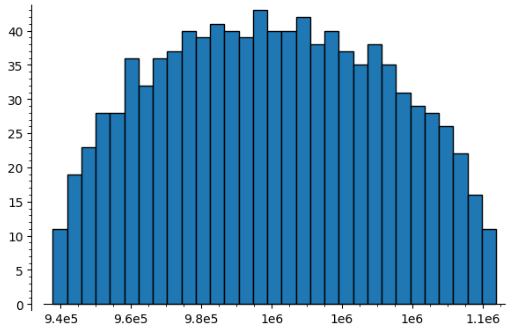
$$g_c(x) = \frac{8c}{\pi} \cdot \frac{\sqrt{x(1-x)}}{4x\sqrt{c} + (\sqrt{c}-1)^2}.$$

When $\frac{M}{N} \to \infty$, we might think of it as considering the previous case of $\frac{M}{N} \to c$, but letting c tend towards ∞. Now consider

$$\begin{aligned} g_\infty(x) &:= \lim_{c\to\infty} g_c(x) \\ &= \lim_{c\to\infty} \frac{8c}{\pi} \cdot \frac{\sqrt{x(1-x)}}{4x\sqrt{c} + (\sqrt{c}-1)^2} \\ &= \frac{8}{\pi} \lim_{c\to\infty} \frac{\sqrt{x(1-x)}}{\frac{4x}{\sqrt{c}} + \left(1 - \frac{1}{\sqrt{c}}\right)^2} \\ &= \frac{8}{\pi}\sqrt{x(1-x)}. \end{aligned}$$

This curve corresponds to a half-ellipse, suggesting that the histogram of Y's eigenvalues corresponds to a half-ellipse as well.

This is also what we observe experimentally. For example, one might choose $M = N^2$ (implying $\frac{M}{N} \to \infty$ when $N \to \infty$) and the entries of X to be standard normally distributed. Now choosing $N = 1\,000$ gives us the following histogram, which does indeed resemble a half-ellipse.



Similarly, one can consider $g_c(x)$ when $c \to 0^+$:

$$\begin{aligned} g_{0^+}(x) &:= \lim_{c\to 0^+} g_c(x) \\ &= \lim_{c\to 0^+} \frac{8c}{\pi} \cdot \frac{\sqrt{x(1-x)}}{4x\sqrt{c} + (\sqrt{c}-1)^2} \\ &= 0. \end{aligned}$$

Now notice that g_c is a function whose integral from 0 to 1 equals 1, and therefore, as $g_{0^+}(x) = 0$, the maximal height of g_c on the interval $[0, 1]$ should tend towards infinity when $c \to 0^+$. We see this as an indication that the distribution of the eigenvalues of Y does not converge.

One should always keep in mind that all the arguments in this section were based on experiments and intuition. They should be seen as an explanation of some potential phenomenon, but surely not as rigorous proofs.

2 Circular Law

In this chapter, we consider random matrices A of the form

$$A = (a_{ij}) = X + \mathrm{i}Y,$$

where $X = (x_{ij})$ and $Y = (y_{ij})$ are real random matrices of size N. Additionally, we assume that $\{x_{ij}\}_{i,j=1}^{N}$ and $\{y_{ij}\}_{i,j=1}^{N}$ are two families of i.i.d. random variables, respectively. As before, we want to think of N as tending towards infinity.
The goal will be to see how the eigenvalues $\lambda_1, \dots, \lambda_N$ of A are distributed. A being complex, there is no reason to assume that any of its eigenvalues are real. We therefore wish to obtain the distribution of A's eigenvalues *in the complex plane.*

To get experimental results, we use many of the methods from chapter 1. Therefore, in order to avoid repetition, we will focus on the outcomes rather than the detailed description of the way we proceed.

Before we start experimenting, we shall generalize the notion of expectation and variance to complex random variables.

Definition 6. (i) If x and y are real random variables (defined on the same probability space), then $a = x + \mathrm{i}y$ is called a *complex random variable.*

(ii) If $a = x + \mathrm{i}y$ is a complex random variable and $\mu_x = \mathbb{E}[x], \mu_y = \mathbb{E}[y]$ exist, then the *expectation* of a is defined as

$$\mathbb{E}[a] := \mu_x + \mathrm{i}\mu_y.$$

(iii) If it exists, the *variance* of a complex random variable a is defined as

$$\mathrm{Var}(a) := \mathbb{E}\left[|a - \mathbb{E}[a]|^2\right].$$

Proposition 7. *If $a = x + \mathrm{i}y$ is a complex random variable, then*

$$\mathrm{Var}(a) = \mathrm{Var}(x) + \mathrm{Var}(y).$$

Proof. In fact,

$$\begin{aligned}
\mathrm{Var}(a) &= \mathbb{E}\left[|a - \mathbb{E}[a]|^2\right] \\
&= \mathbb{E}\left[|x + \mathrm{i}y - \mathbb{E}[x] - \mathrm{i}\mathbb{E}[y]|^2\right] \\
&= \mathbb{E}\left[|(x - \mathbb{E}[x]) + \mathrm{i}\,(y - \mathbb{E}[y])|^2\right] \\
&= \mathbb{E}\left[(x - \mathbb{E}[x])^2 + (y - \mathbb{E}[y])^2\right] \\
&= \mathbb{E}\left[(x - \mathbb{E}[x])^2\right] + \mathbb{E}\left[(y - \mathbb{E}[y])^2\right] \\
&= \mathrm{Var}(x) + \mathrm{Var}(y),
\end{aligned}$$

as desired. $\square$

We are now ready to examine the distribution of A's eigenvalues.

2.1 Complex standard normal distribution

Our first goal is to examine the eigenvalues of A when the a_{ij} are complex standard normally distributed.

Definition 8. A complex random variable $a = x + \mathrm{i}y$ is said to be *complex standard normally distributed* if x and y are independent and normally distributed with mean 0 and variance $\frac{1}{2}$. In that case, we write $a \sim \mathcal{CN}(0, 1)$.

Notice that for a as in definition 8, we indeed have

$$\mathbb{E}[a] = \mathbb{E}[x] + \mathrm{i}\mathbb{E}[y] = 0,$$

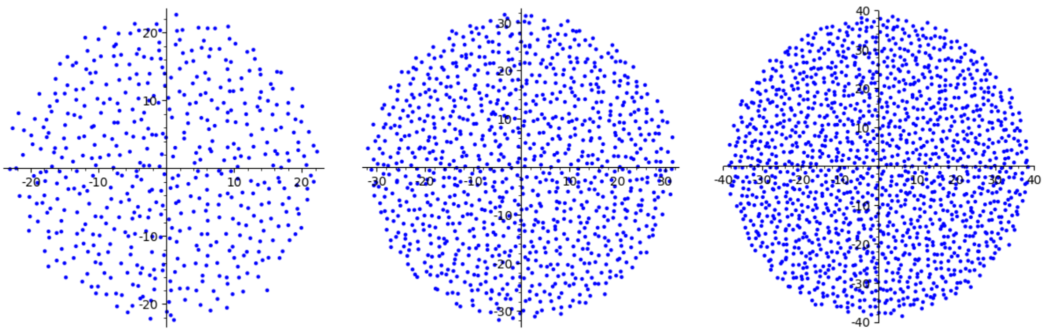
and by proposition 7,

$$\mathrm{Var}(a) = \mathrm{Var}(x) + \mathrm{Var}(y) = \frac{1}{2} + \frac{1}{2} = 1,$$

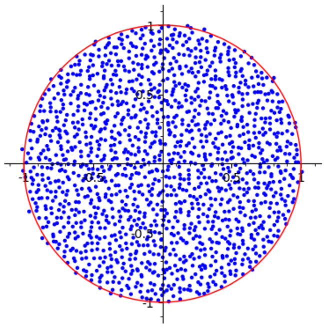
as suggested by the notation $a \sim \mathcal{CN}(0, 1)$.

We will independently generate random matrices X and Y with i.i.d. entries that are normally distributed with mean 0 and variance $\frac{1}{2}$, then compute the complex eigenvalues of $A = X + \mathrm{i}Y$ and represent them in the complex plane $\mathbb{C}$ (or equivalently, in the euclidean plane $\mathbb{R}^2$). Notice that this construction of A ensures that its entries are complex standard normally distributed.

Recall that N denotes the size of the matrices X and Y (and therefore of A). Choosing $N = 500, 1\,000, 1\,500$ gives us the following plots.



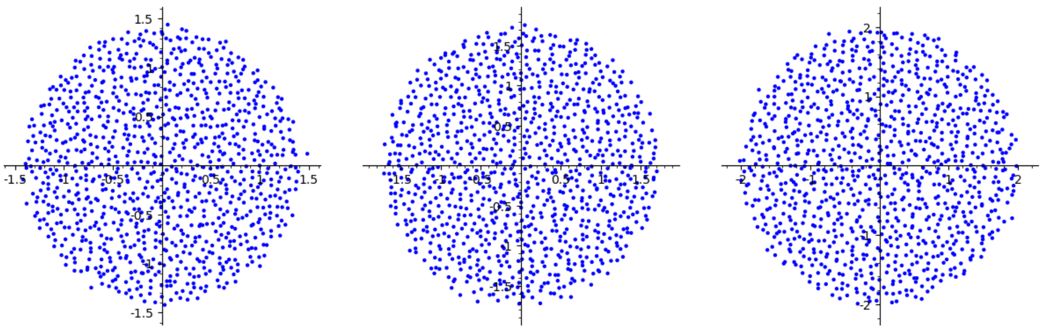
The eigenvalues of A appear to lie mostly in a circle of radius $\sqrt{N}$. Phrased differently, we expect the $\frac{\lambda_i}{\sqrt{N}}$ to lie in a circle of radius 1. Let us check this assumption for $N = 1\,500$ by plotting the $\frac{\lambda_i}{\sqrt{N}}$. In red we mark the unit circle centred at 0.



Our assumption seems to hold.

2.2 Dependence on distribution

The next step is to increase the variance σ^2 of the a_{ij}. This can be obtained by choosing the x_{ij} and y_{ij} to be normally distributed with mean 0 and variance $\frac{\sigma^2}{2}$ for varying σ^2. This ensures that that $\mathrm{Var}(a_{ij}) = \mathrm{Var}(x_{ij}) + \mathrm{Var}(y_{ij}) = \frac{\sigma^2}{2} + \frac{\sigma^2}{2} = \sigma^2$. For example, let us choose $\sigma^2 = 2, 3, 4$ and $N = 1\,000$. The corresponding plottings of $\frac{\lambda_i}{\sqrt{N}}$ are the following.



One notices that the circle's radius appears to be proportional to the square root of the variance σ^2, that is, to the standard deviation σ.

We were also able to experimentally observe that, as long as N and σ^2 remain untouched,

- changing the mean of the x_{ij} and / or y_{ij}
- changing the distribution of the x_{ij} and / or y_{ij}
- choosing y_{ij} to depend on x_{ij} (for example, when Y is a multiple of X)

have no effect on the distribution of A's eigenvalues.

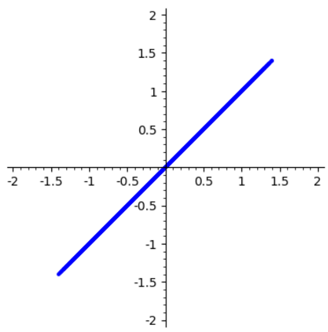
However, it seems to be important to assure that the a_{ij} remain independent (the notion of independence generalizes to complex random variables in the expected way). For example, one might want to choose $Y = X^\top$. As the x_{ij} are i.i.d., the y_{ij}

would also be i.i.d. However, notice that the a_{ij} aren't independent anymore (when $N \geq 2$), since $a_{ij} = \mathrm{i}\overline{a_{ji}}$. In fact,

$$\begin{aligned} a_{ij} &= x_{ij} + \mathrm{i}y_{ij} \\ &= x_{ij} + \mathrm{i}x_{ji} \\ &= \mathrm{i}(x_{ji} - \mathrm{i}x_{ij}) \\ &= \mathrm{i}(x_{ji} - \mathrm{i}y_{ji}) \\ &= \mathrm{i}\overline{a_{ji}}, \end{aligned}$$

where we used that $y_{ij} = x_{ji}$, since $Y = X^\top$. Also, to avoid confusion, we want to point out the difference in typefaces used for the imaginary unit i and the index i.

When choosing $Y = X^\top$, the corresponding plot of the $\frac{\lambda_i}{\sqrt{N}}$ (for $N = 1\,000$ and x_{ij} normally distributed with variance $\frac{1}{2}$) is the following.



All the points appear to lie on the line given by $\{x + \mathrm{i}x : x \in \mathbb{R}\}$. This phenomenon can easily be explained. Firstly, observe that $a_{ij} = \mathrm{i}\overline{a_{ji}}$ for all $1 \leq i, j \leq N$ is equivalent to the equality $\mathrm{i}\overline{A}^\top = A$, and therefore we can apply the next proposition.

Proposition 9. *If a complex square matrix A satisfies $\mathrm{i}\overline{A}^\top = A$, then all of A's eigenvalues have the same real and imaginary part.*

Proof. If $\mathrm{i}\overline{A}^\top = A$, then $\overline{A}^\top = -\mathrm{i}A$, so $A\overline{A}^\top = A(-\mathrm{i}A) = (-\mathrm{i}A)A = \overline{A}^\top A$, that is, A is normal. By the Spectral Theorem for normal matrices, there is a diagonal matrix D and a unitary matrix U for which

$$A = UD\overline{U}^\top.$$

Now, as U is unitary $\left(\text{that is, } U^{-1} = \overline{U}^\top\right)$, we get

$$D = \overline{U}^\top AU = \overline{U}^\top \mathrm{i}\overline{A}^\top U = \mathrm{i}\overline{\overline{U}^\top AU}^\top = \mathrm{i}\overline{D}^\top.$$

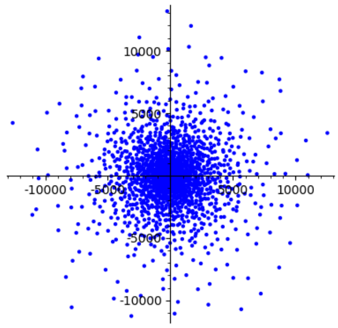
As an eigenvalue λ of A lies on the diagonal of D, there is k for which $\lambda = D_{kk}$. But

$$D_{kk} = \left(\mathrm{i}\overline{D}^\top\right)_{kk} = \mathrm{i}\overline{D_{kk}},$$

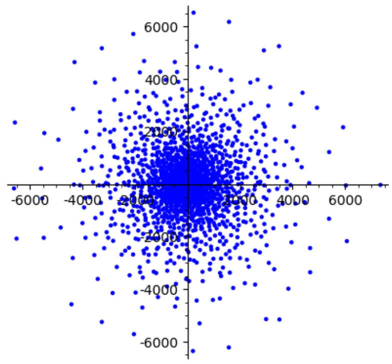
that is, $\mathrm{Im}(D_{kk}) = \mathrm{Re}(D_{kk})$, or equivalently, $\mathrm{Im}(\lambda) = \mathrm{Re}(\lambda)$, as claimed. $\square$

We therefore want the a_{ij} to be independent as well.

Lastly, we are interested in what happens when we choose distributions for x_{ij} or y_{ij} that do not have a finite variance. As an example, we choose the standard Cauchy distribution for both, and we obtain the following plot of A's eigenvalues (for $N = 1\,500$).



This doesn't follow the circular pattern from above. We obtain similar results if the entries of one matrix have finite variance, while the entries of the other one don't. For example, we might choose the x_{ij} to be standard Cauchy distributed, and the y_{ij} to be Poisson distributed with parameter $\frac{1}{2}$. We obtain the following plot of A's eigenvalues, again for $N = 1\,500$.



We end this chapter with a recapping conjecture based on our observations.

Conjecture 10. *Let $A = (a_{ij}) \in \mathcal{M}_N(\mathbb{C})$ be a random complex matrix with i.i.d. entries a_{ij} of variance σ^2. As $N \to \infty$, the distribution of the eigenvalues of $\frac{A}{\sigma\sqrt{N}}$ converges to the uniform distribution in the complex unit disc centred at* 0.

Note that dividing A by $\sigma\sqrt{N}$ has the effect of dividing the eigenvalues of A by the same factor, hence making sure that the obtained values lie (mostly) in the mentioned unit disc.

This phenomenon is known as the "circular law." One might want to read more about this in [2].

3 Wigner's Semicircle Law

We shall finish with the arguably most famous result about random matrices: Wigner's Semicircle Law.

In this chapter, we consider real *symmetric* random matrices A of the form

$$A = (a_{ij}) \in \mathcal{M}_N(\mathbb{R}),$$

where the upper triangular matrix entries of A are independent random variables. Phrased differently,

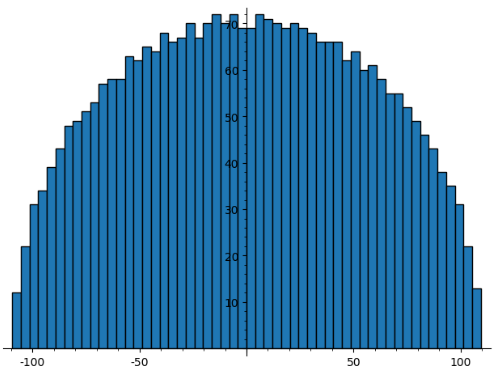
$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{1n} & \cdots & a_{nn} \end{pmatrix},$$

where the entries $a_{ij}, i \leq j$, are real independent random variables.

Again, the goal will be to analyse the distribution of A's eigenvalues as N tends towards infinity. Firstly, we notice that, as A is symmetric, all of A's eigenvalues are real by proposition 1. This allows us to represent them in a histogram, analogously to the way we did in chapter 1.

3.1 Recap of experimental results

For example, we may choose all the a_{ij} to be standard normally distributed. We obtain the following histogram for $N = 3\,000$.



Notice that the histogram appears to correspond to a half-ellipse. This observation can be made in many other cases, as we describe next.

If in addition to our original assumptions, we suppose that

- all the a_{ij} have equal variance σ^2,
- the a_{ii} (A's diagonal entries) are identically distributed,
- the $a_{ij}, i \neq j$, (A's remaining entries) are identically distributed,

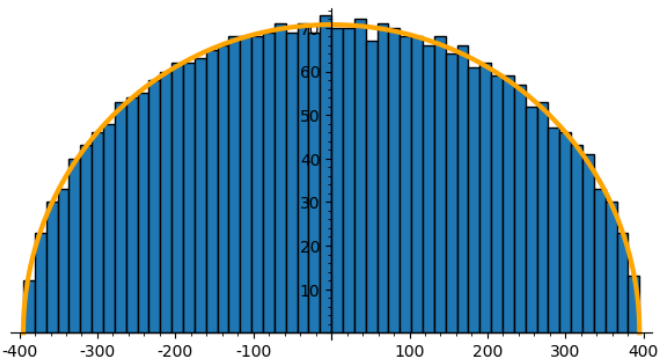
one notices that the histogram of A's eigenvalues corresponds to a half-ellipse centred at the origin, with width $4\sigma\sqrt{N}$. In particular, most of A's eigenvalues lie in the range $\left[-2\sigma\sqrt{N}, 2\sigma\sqrt{N}\right]$.
As the methods we used to derive these conjectures are analogous to those used in the previous chapters, we shall not go into more detail.

It should be pointed out that *again*, the mean of the a_{ij} does not appear to play a role.

To show how precise above conjecture is, we shall consider an example. Let us choose the a_{ii} to be Poisson distributed with parameter 13, and the $a_{ij}, i \neq j$, to be normally distributed with mean 100 and variance 13. This guarantees that all of A's entries have a variance of $\sigma^2 = 13$. Note that 13 is just some arbitrary value. Let us also arbitrarily choose $N = 3\,000$.
Now, we generate the corresponding histogram. We choose $\lfloor\sqrt{N}\rfloor$ bins for the histogram. As the width of each bin now equals $\frac{4\sigma\sqrt{N}}{\lfloor\sqrt{N}\rfloor}$ (width of histogram divided by number of bins), each eigenvalue of A contributes to an area in the histogram of exactly that amount. This implies that the histogram's total area equals $\frac{4\sigma N\sqrt{N}}{\lfloor\sqrt{N}\rfloor}$.
As the histogram should correspond to a half-ellipse of semi-width $w = 2\sigma\sqrt{N}$ and some semi-height h, its area is also given by $\frac{\pi w h}{2}$, and thus

$$\frac{4\sigma N\sqrt{N}}{\lfloor\sqrt{N}\rfloor} = \frac{\pi w h}{2} = \frac{\pi \cdot 2\sigma\sqrt{N}h}{2},$$

so that $h = \frac{4N}{\pi\lfloor\sqrt{N}\rfloor}$. This leaves us with the following histogram and half-ellipse.



3.2 Interpretation

Returning to the general case, let us denote the N eigenvalues of A by $\lambda_1, \ldots, \lambda_N$. Then we obtain by above observations for $x \in \left[-2\sigma\sqrt{N}, 2\sigma\sqrt{N}\right]$:

$$\frac{\text{number of eigenvalues of } A \text{ below } x}{N} \approx \frac{\text{area of ellipse from } -2\sigma\sqrt{N} \text{ to } x}{\text{area of ellipse from } -2\sigma\sqrt{N} \text{ to } 2\sigma\sqrt{N}}.$$

We can rescale the obtained half-ellipse horizontally and vertically, as well as translate it horizontally, to obtain the upper half unit circle, described by the curve of the function $f : [0, 1] \to \mathbb{R}$, $f(t) = \sqrt{t(1-t)}$. As the ellipse ranges from $-2\sigma\sqrt{N}$ to $2\sigma\sqrt{N}$, the horizontal rescaling and translation changes values x on the x-axis into $\frac{x}{4\sigma\sqrt{N}} + \frac{1}{2}$. Therefore

$$\frac{\text{area of ellipse from } -2\sigma\sqrt{N} \text{ to } x}{\text{area of ellipse from } -2\sigma\sqrt{N} \text{ to } 2\sigma\sqrt{N}} = \frac{\text{area of } f \text{ from } 0 \text{ to } \frac{x}{4\sigma\sqrt{N}} + \frac{1}{2}}{\text{area of } f \text{ from } 0 \text{ to } 1}.$$

Notice that we can ignore the vertical rescaling, as it would just multiply numerator and denominator by the same factor, and could therefore be cancelled out.

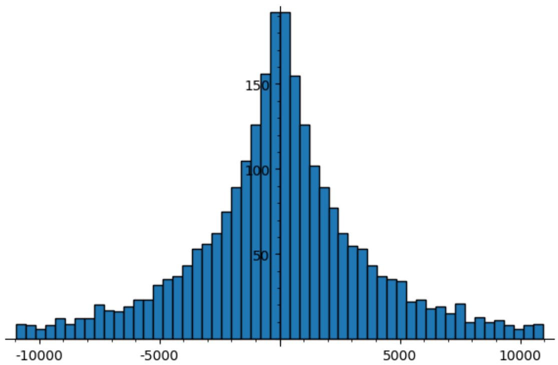
Putting everything together, we have

$$\begin{aligned} \frac{\#\{1 \le i \le N : \lambda_i \le x\}}{N} &\approx \frac{\text{area of } f \text{ from } 0 \text{ to } \frac{x}{4\sigma\sqrt{N}} + \frac{1}{2}}{\text{area of } f \text{ from } 0 \text{ to } 1} \\ &= \frac{\int_0^{\frac{x}{4\sigma\sqrt{N}}+\frac{1}{2}} f(t)\,\mathrm{d}t}{\int_0^1 f(t)\,\mathrm{d}t}. \end{aligned}$$

Using SageMath (or a trigonometric substitution), we can obtain an anti-derivative for f, namely $F : [0, 1] \to \mathbb{R}$, $F(t) = \frac{(2t-1)\sqrt{t(1-t)} - \arcsin\sqrt{1-t}}{4}$. Hence, we get

$$\begin{aligned} \frac{\#\{1 \le i \le N : \lambda_i \le x\}}{N} &\approx \frac{F\left(\frac{x}{4\sigma\sqrt{N}} + \frac{1}{2}\right) + \frac{\pi}{8}}{\frac{\pi}{8}} \\ &= \frac{8}{\pi} F\left(\frac{x}{4\sigma\sqrt{N}} + \frac{1}{2}\right) + 1. \end{aligned}$$

Note that none of the our assumptions can be relaxed. For example, if we choose the a_{ij} to be standard Cauchy distributed, we do not obtain a histogram corresponding to a half ellipse, as shown below, where we choose $N = 3\,000$.



Similarly, it is of great importance to ensure that all the a_{ij} have equal variance σ^2, otherwise we cannot conclude in the same way.

To finish this section, we shall summarize our observations in a final conjecture.

Conjecture 11. *Let $A = (a_{ij}) \in \mathcal{M}_N(\mathbb{R})$ be a random symmetric matrix, where*

- *the* $a_{ij}, i \leq j$, *are independent and have equal variance* σ^2,
- *the* a_{ii} *respectively the* $a_{ij}, i \neq j$, *are identically distributed.*

If we denote A*'s eigenvalues by* $\lambda_1, \ldots \lambda_N$, *then for any* $x \in \mathbb{R}$, *we have*

$$\frac{\#\{1 \leq i \leq N : \lambda_i \leq x\}}{N} \approx \begin{cases} 0 & \text{if } x < -2\sigma\sqrt{N} \\ \frac{8}{\pi} F\left(\frac{x}{4\sigma\sqrt{N}} + \frac{1}{2}\right) + 1 & \text{if } |x| \leq 2\sigma\sqrt{N} \\ 1 & \text{if } x > 2\sigma\sqrt{N} \end{cases}$$

where $F(t) = \frac{(2t-1)\sqrt{t(1-t)} - \arcsin\sqrt{1-t}}{4}$.

The interested reader might want to take a look at [1], which covers this topic in great detail.

Bibliography

[1] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices.* Cambridge Studies in Advanced Mathematics, CUP, 2009.

[2] T. Tao, V. Vu, and M. Krishnapur, “Random matrices: Universality of ESDs and the circular law,” 2009.

[3] V. A. Marčenko and L. A. Pastur, “Distribution of eigenvalues for some sets of random matrices,” 1967.