Assignment 4 - Numerical Linear Algebra

Arthur Rabello Oliveira[1], Henrique Coelho Beltrão[2]

20/06/2025

Abstract

We performed a computational study of $m \times n$ real Gaussian matrices whose entries are i.i.d. $N(0,1)$. First, for increasing values of the ambient dimension m (with $n = 1000$) we sampled the 2 -norm of each column and verified that the empirical histograms are well-approximated by the χ_m distribution, converging to $\sqrt{m}$. Second, fixing $m = 100$ and letting n vary from 10 to 1000, we examined all off-diagonal inner products; they converge to $N(0, m)$, confirming the classical central-limit prediction. Third, we estimated the **worst** non-orthogonality—the maximum absolute cosine similarity among every matrix's columns—over up to $K = 10^5$ independent realisations and showed that the resulting maxima follow a Gumbel law, in line with extreme-value theory. Finally, we derived and validated the algorithmic cost $O(mn^2)$ of this pipeline and quantified that $K \approx 10^3$ already yields stable statistics.

Contents

[1]Escola de Matemática Aplicada, Fundação Getúlio Vargas (FGV/EMAp), email: arthur.oliveira.1@fgv.edu.br
[2]Escola de Matemática Aplicada, Fundação Getúlio Vargas (FGV/EMAp), email: henrique.beltrao@fgv.edu.br

1. Introduction

1.1. From multivariate statistics to nuclear physics

The systematic study of Gaussian random matrices began with **John Wishart** (1928), whose sample-covariance analysis produced the distribution that still underpins multivariate testing and Bayesian inference. Two decades later, **Eugene Wigner** introduced random matrices to model the energy levels of heavy nuclei, proving the celebrated semicircle law for their eigenvalue distribution. These twin origins—statistics and physics—sparked what is now called **Random Matrix Theory (RMT)**, whose cornerstones include the Marchenko-Pastur law for singular values of rectangular Gaussian matrices and sharp bounds on their extremes.

1.2. Modern applications

Random-matrix ideas now inform topics as diverse as numerical conditioning, wireless communication, portfolio theory and, most recently, **deep learning**. Which is why we study Gaussian matrices in this assignment, focusing on their non-orthogonality and the distribution of their inner products.

2. Norm Distribution (a)

2.1. The Chi-Square Distribution

Here we construct a theoretical basis for our analysis of the histograms shown in Section 2.2

When we generate a matrix $A \in \mathbb{R}^{m\times n}$, with $A_{ij} \sim N(0,1)$ independent, each column c_i is a gaussian vector in $\mathbb{R}^m$. if

$$x = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix} \in \mathbb{R}^m \tag{1}$$

Is a column, then:

$$\begin{aligned} V = \|x\|_2 &= \sqrt{\sum_{i=1}^{m} X_i^2} \\ V^2 &= \sum_{i=1}^{m} X_i^2 \end{aligned} \tag{2}$$

Is of our interest. The expected value and variance are:

$$\begin{aligned} \mathbb{E}[V^2] &= \mathbb{E}\left[\sum_{i=1}^{m} X_i^2\right] = \sum_{i=1}^{m} \mathbb{E}[X_i^2] = m \\ \mathrm{Var}(V^2) &= \mathrm{Var}\left(\sum_{i=1}^{m} X_i^2\right) = \sum_{i=1}^{m} \mathrm{Var}(X_i^2) = 2m \end{aligned} \tag{3}$$

But we know that if $X_i \sim N(0,1)$ are independent:

$$\sum_{i=1}^{m} X_i^2 \sim \chi_m^2 \tag{4}$$

where χ_m is the chi-squared distribution with m degrees of freedom, better discussed in Section 2.1.

Taking the square root on eq. (4), we have:

$$V = \|x\|_2 = \sqrt{\sum_{i=1}^{m} X_i^2} \sim \sqrt{\chi_m^2} \sim \chi_m \tag{5}$$

The 2-norm of a vector x is distributed as a chi distribution with m degrees of freedom, in order to understand the distribution for many values of m, we can calculate the expected value and variance of this distribution as a function of m. The PDF of the chi distribution (with m degrees of freedom) is:

$$f_V(\varphi) = \frac{1}{2^{\frac{m}{2}-1} \cdot \Gamma\left(\frac{m}{2}\right)} \varphi^{m-1} e^{-\frac{\varphi^2}{2}} \tag{6}$$

So from this, the expected value is:

$$\mathbb{E}(V) = \sqrt{2} \cdot \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)} \tag{7}$$

And from this, the variance:

$$\mathrm{Var}(V) = m - \left(\sqrt{2} \cdot \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)}\right)^2 \tag{8}$$

The Stirling Approximation provides a good approximation for the expected value and variance:

$$\mathbb{E}(V) \approx \sqrt{m} \cdot \left(1 - \frac{1}{4m} + O\left(\frac{1}{m^2}\right)\right) \tag{9}$$

$$\mathrm{Var}(V) \approx \frac{1}{2} + O\left(\frac{1}{m}\right) \tag{10}$$

2.2. Histograms

The first cell of this notebook has as expected output, with input being matrices with fixed $n = 1000$ and $m \in \{10, 20, 100, 200, 1000, 2000\}$, the following plots:

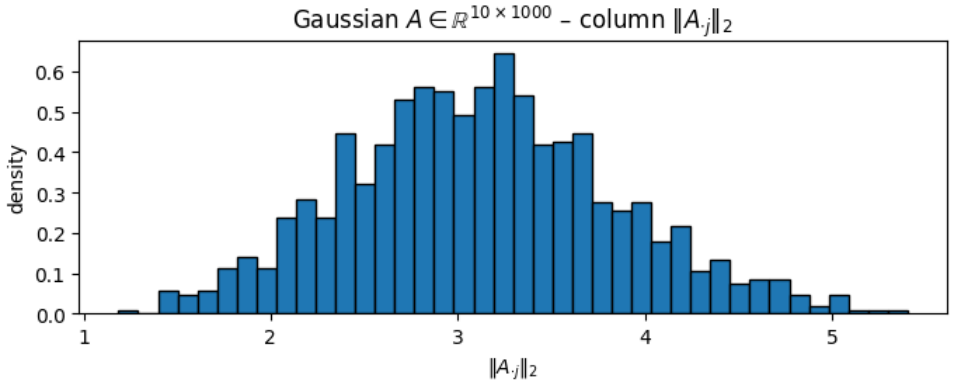


Figure 1: 10×1000 gaussian matrix

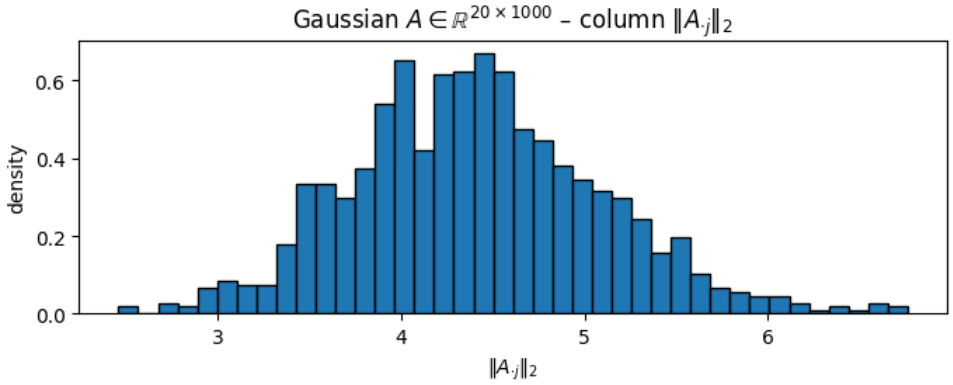


Figure 2: 20×1000 gaussian matrix

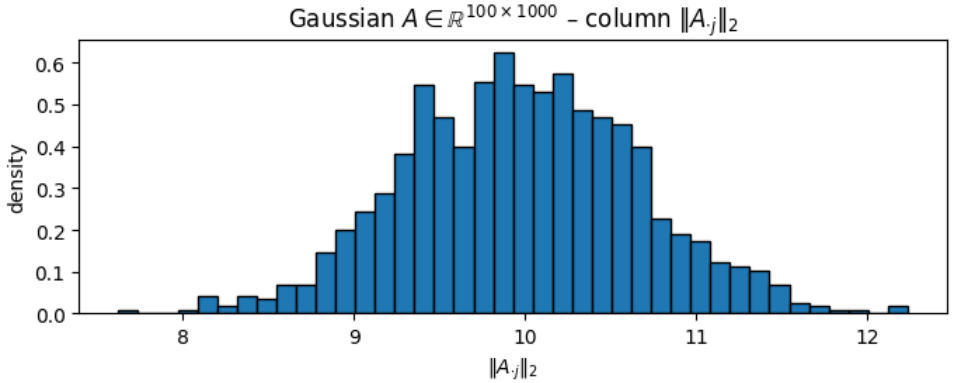


Figure 3: 100×1000 gaussian matrix

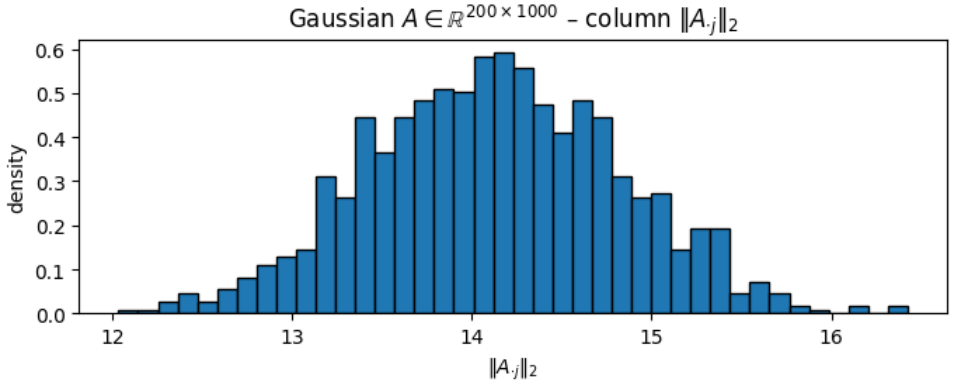


Figure 4: 200×1000 gaussian matrix

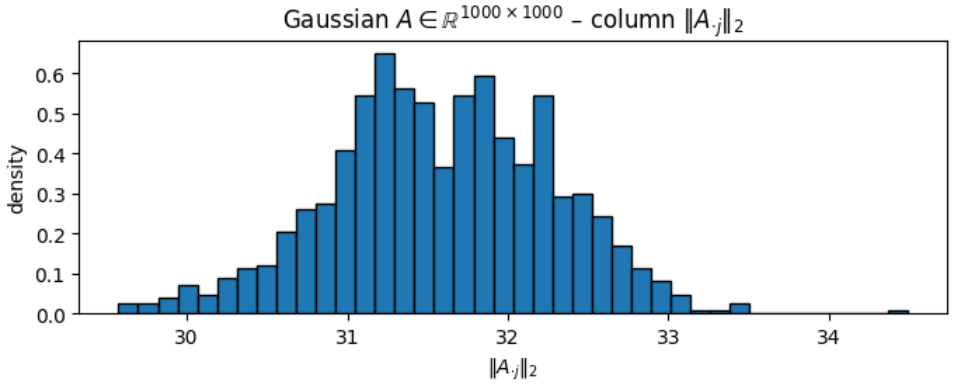


Figure 5: 1000×1000 gaussian matrix

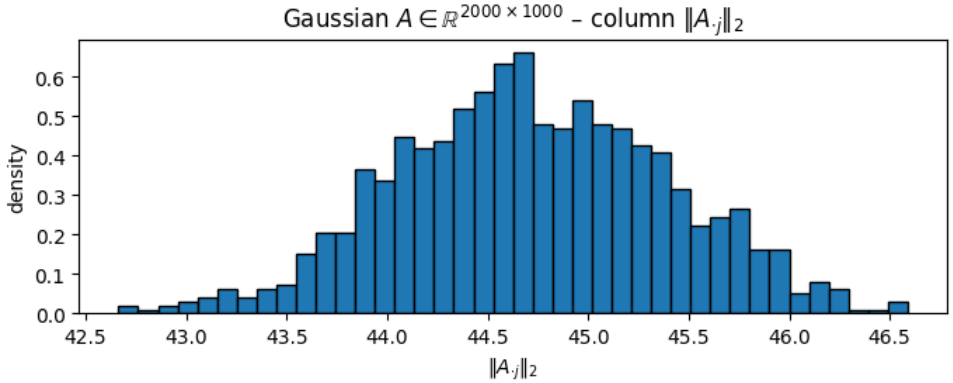


Figure 6: 2000×1000 gaussian matrix

m	**approximate μ_m (theory)**	**$[\mu \pm 3\sigma]$ (theory)**	**observed spike**	**visual range**
10	3.08	$1.0 - 5.18$	≈ 3.1	$1.2 - 5.1$
20	4.42	$2.3 - 6.52$	$\approx 4.3 - 4.4$	$2.9 - 6.4$
100	9.98	$7.9 - 12.1$	$\approx 9.9 - 10.0$	$7.8 - 12.0$
200	14.12	$12.0 - 16.2$	≈ 14.1	$12.2 - 16.2$
1000	31.61	$29.5 - 33.7$	$\approx 31.7 - 32.0$	$29.9 - 33.3$
2000	44.72	$42.6 - 46.8$	$\approx 44.5 - 45.0$	$42.8 - 46.6$

This table illustrates the expected value μ_m and the range $[\mu - 3\sigma, \mu + 3\sigma]$ for Figure 1 to Figure 6.

So apparently as m grows, the size of the gaussian vectors rapidly converge to $\sqrt{m}$, with small errors.

3. Inner Products (b)

Here we construct a theoretical basis for our analysis of the inner products shown in Section 3.1.

When we generate a matrix $A \in \mathbb{R}^{m \times n}$, with $A_{ij} \sim N(0, 1)$ independent, each column c_i is a gaussian vector in $\mathbb{R}^m$. If The inner product of two gaussian vectors $x = (X_1, ..., X_n), y = (Y_1, ..., Y_n)$ is:

$$Z = \langle x, y \rangle = \sum_{i=1}^{m} X_i Y_i \tag{11}$$

With $X, Y \sim N(0,1)$. Since X_i, Y_j are independent, we have:

$$\begin{aligned} \mathbb{E}[Z] &= \sum_{i=1}^{m} \mathbb{E}[X_i Y_i] = \sum_{i=1}^{m} \mathbb{E}[X_i]\mathbb{E}[Y_i] = 0 \\ \mathrm{Var}(Z) &= \sum_{i=1}^{m} \mathrm{Var}(X_i Y_i) = \sum_{i=1}^{m} \mathbb{E}[X_i^2]\mathbb{E}[Y_i^2] = \sum_{i=1}^{m} 1 = m \end{aligned} \tag{12}$$

If $W = X_i Y_i$, we have:

$$M_W(\varphi) = \mathbb{E}[e^{\varphi W}] = \frac{1}{\sqrt{1-\varphi^2}}, |\varphi| < 1 \tag{13}$$

Over all $W_i = X_i Y_i$:

$$M_Z(\varphi) = \mathbb{E}[e^{\varphi Z}] = (M_W(\varphi))^m = \left(\frac{1}{\sqrt{1-\varphi^2}}\right)^m = (1-\varphi^2)^{-\frac{m}{2}}, |\varphi| < 1 \tag{14}$$

And magically:

$$M_{\frac{Z}{\sqrt{m}}}(\varphi) = \left(1 - \frac{\varphi^2}{m}\right)^{-\frac{m}{2}} \Rightarrow \lim_{m\to\infty} M_{\frac{Z}{\sqrt{m}}}(\varphi) = e^{\frac{\varphi^2}{2}} \tag{15}$$

Precisely the moment generating function of a standard normal distribution, so as $m \to \infty$:

$$\frac{Z}{\sqrt{m}} \sim N(0,1) \tag{16}$$

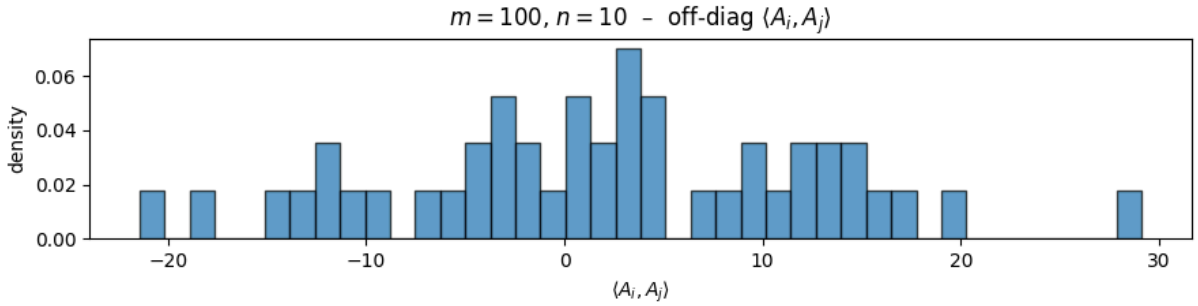
And finally:

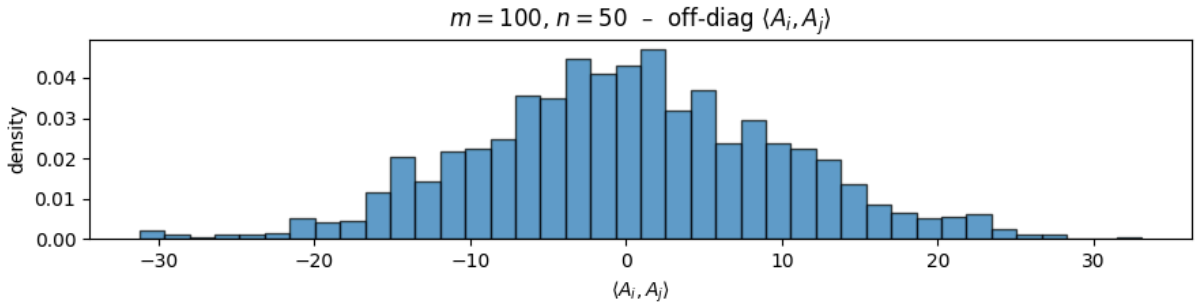
$$Z \sim N(0,m) \tag{17}$$

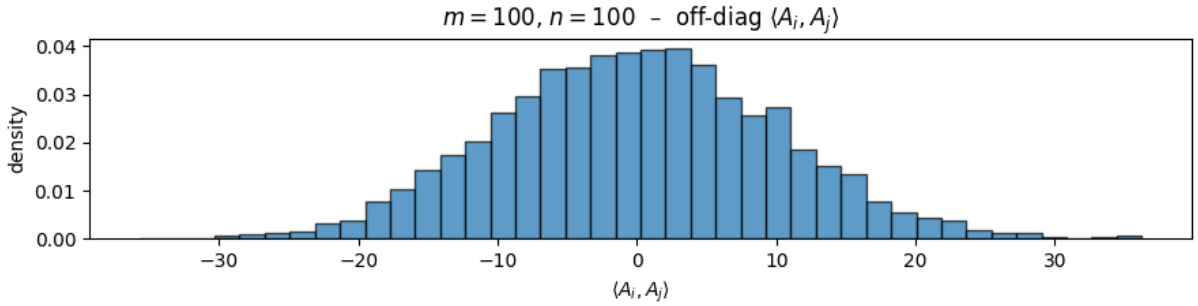
With a fixed $m = 100$, when $n \to \infty$ we can see the distribution approaching $N(0,m)$, as shown in Section 3.1

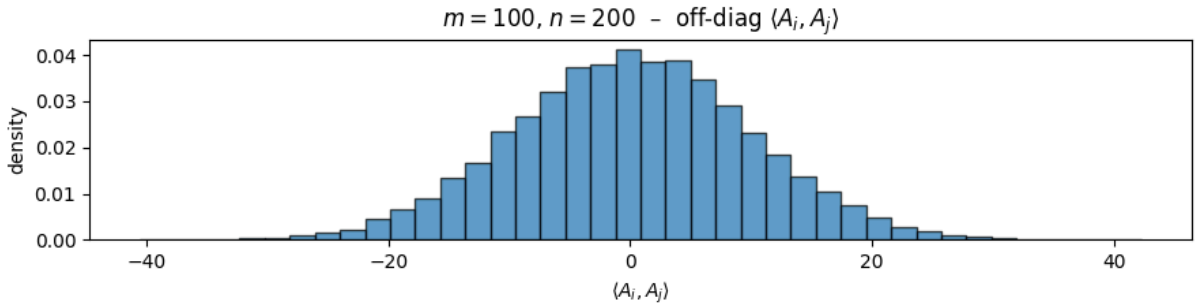
3.1. Histograms

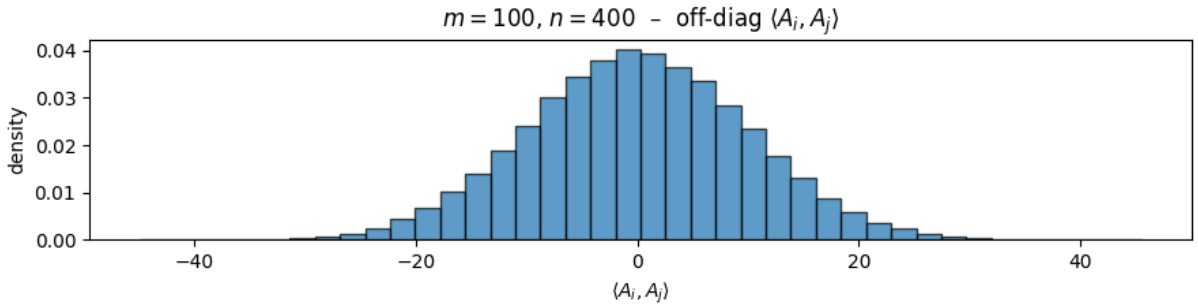
The following plots are an expected output for the second cell of this notebook, with input $m = 100, n \in \{10, 20, 30, 40, 50, 60, ..., 1000\}$:

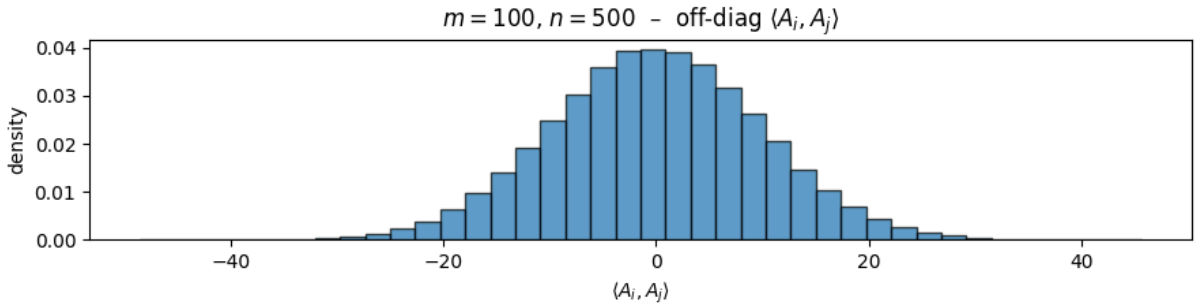



$m = 100, n = 50$ – off-diag $\langle A_i, A_j \rangle$
density
0.00
0.01
0.02
0.03
0.04
−30
−20
−10
0
10
20
30
$\langle A_i, A_j \rangle$


$m = 100, n = 100$ – off-diag $\langle A_i, A_j \rangle$
density
0.00
0.01
0.02
0.03
0.04
−30
−20
−10
0
10
20
30
$\langle A_i, A_j \rangle$


$m = 100, n = 200$ – off-diag $\langle A_i, A_j \rangle$
density
0.00
0.01
0.02
0.03
0.04
−40
−20
0
20
40
$\langle A_i, A_j \rangle$


$m = 100, n = 400$ – off-diag $\langle A_i, A_j \rangle$
density
0.00
0.01
0.02
0.03
0.04
−40
−20
0
20
40
$\langle A_i, A_j \rangle$


$m = 100, n = 500$ – off-diag $\langle A_i, A_j \rangle$
density
0.00
0.01
0.02
0.03
0.04
−40
−20
0
20
40
$\langle A_i, A_j \rangle$

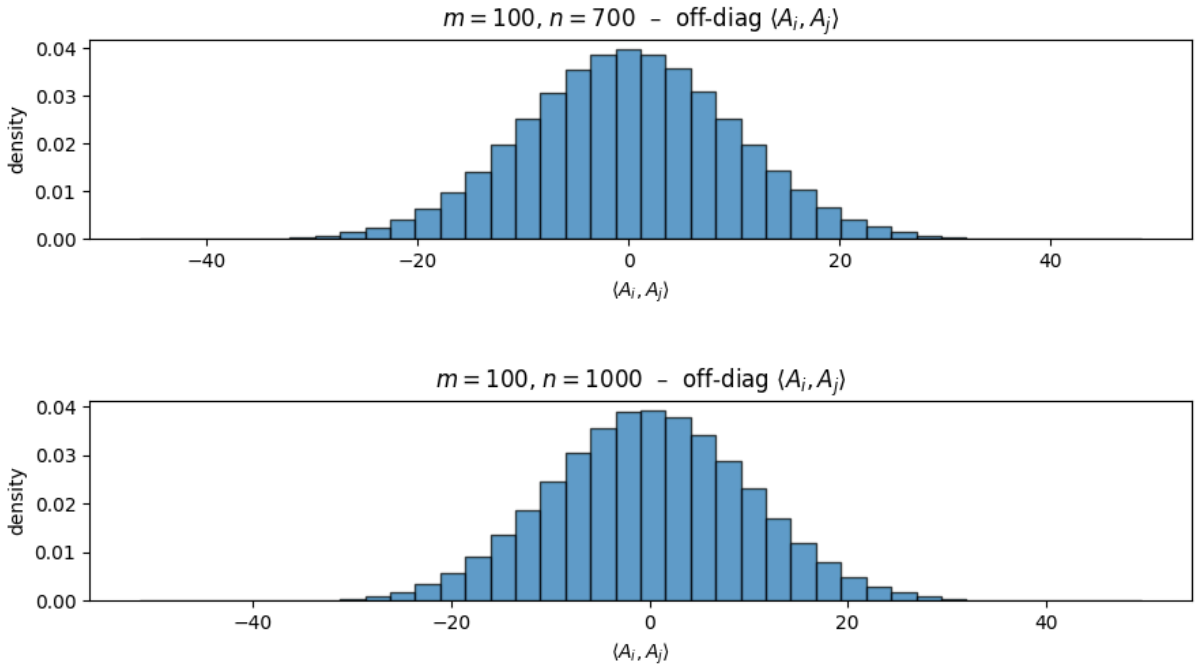


Figure 7 → Figure 14 shows that the distribution indeed approaches $N(0,1)$

4. The Maximum Distribution (c)

In this section, we analyze the distribution of the maximum non-orthogonality between columns of a Gaussian matrix. This non-orthogonality is quantified by the maximum absolute value of the cosine similarity between any two distinct column vectors. Specifically, for a matrix $A \in \mathbb{R}^{m\times n}$, we study the distribution of the random variable:

$$M = \max_{i\neq j} \frac{|\langle A_i, A_j\rangle|}{\|A_i\|\|A_j\|} \tag{18}$$

Our experiment generates K independent realizations of this value, $M_1, M_2, ..., M_K$, by creating K different Gaussian matrices of size $m = 100, n = 300$. The histograms later shown in Section 4.2 display the empirical probability density function of this collection of maxima.

4.1. Theoretical Framework and the Gumbel Distribution

Let $C_{ij} = \frac{\langle A_i, A_j\rangle}{\|A_i\|\|A_j\|}$. For a given matrix A, we are examining the maximum of $N = \frac{n(n-1)}{2}$ random variables, $\{|C_{ij}|\}_{1\leq i<j\leq n}$. For $m = 100$ and $n = 300$, this is the maximum of $N = 44850$ values.

We are interested in the maximum of $\{|C_{ij}|\}$. As established in previous sections:

- From part (a) (Section 2), for large m, $\|A_i\|$ concentrates around $\sqrt{m}$.
- From part (b) (Section 3), $Z_{ij} = \langle A_i, A_j\rangle$ is approximately $N(0, m)$.

Let's first characterize the distribution of a single variable C_{ij}.

$$C_{ij} = \frac{Z_{ij}}{\|A_i\|\|A_j\|} \approx \frac{N(0,m)}{\sqrt{m}\cdot\sqrt{m}} = \frac{N(0,m)}{m} \tag{19}$$

If a random variable $X \sim N(0,\sigma^2)$, then $\frac{X}{c} \sim N\left(0, \frac{\sigma^2}{c^2}\right)$. Thus:

$$C_{ij} \approx N\left(0, \frac{m}{m^2}\right) = N\left(0, \frac{1}{m}\right) \tag{20}$$

So, the individual correlation values are approximately drawn from a normal distribution with mean 0 and a small variance of $\frac{1}{m}$.

Our analysis, however, concerns the variable $M = \max_{i \neq j} |C_{ij}|$. The parent distribution is therefore not $N\left(0, \frac{1}{m}\right)$, but rather its absolute value, $|N\left(0, \frac{1}{m}\right)|$. This is known as a **folded normal distribution**.

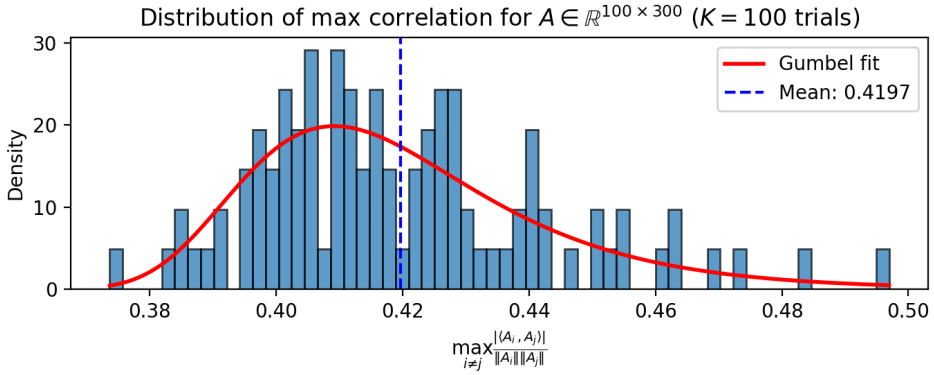
The tail of the folded normal distribution behaves identically to the tail of the underlying normal distribution. According to **Extreme Value Theory**, the limiting distribution for the maximum of many i.i.d. variables from a parent distribution with an exponential tail (like the normal distribution) is the **Gumbel distribution**.

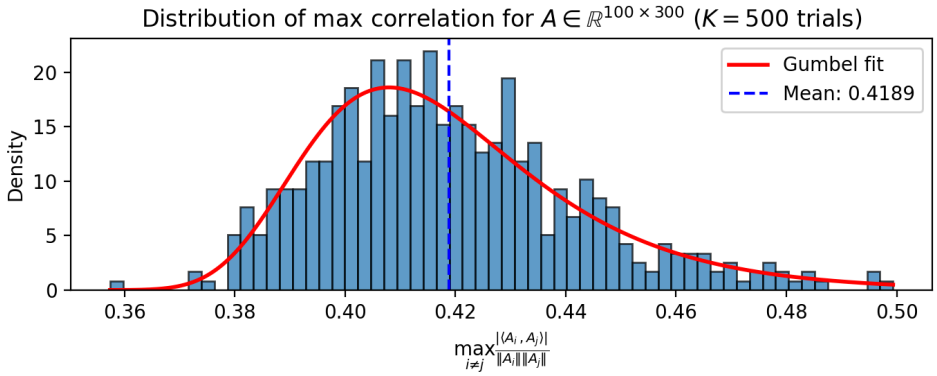
The probability density function (PDF) for the Gumbel distribution is given by:

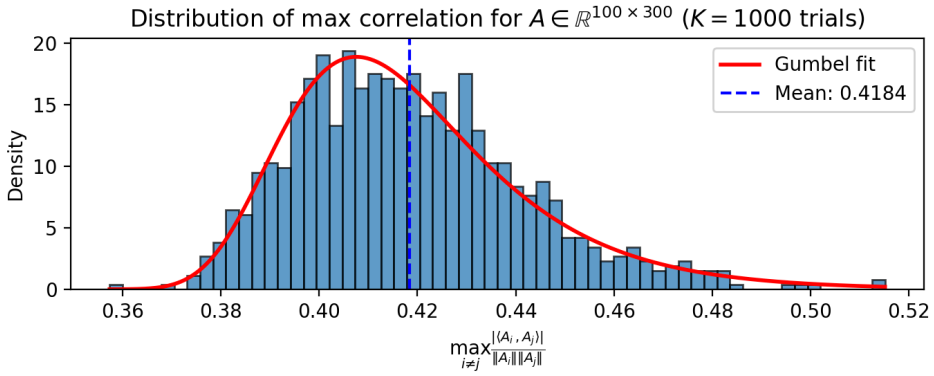
$$f(x; \mu, \beta) = \frac{1}{\beta} e^{-(z + e^{-z})}$$
$$z = \frac{x - \mu}{\beta} \tag{21}$$

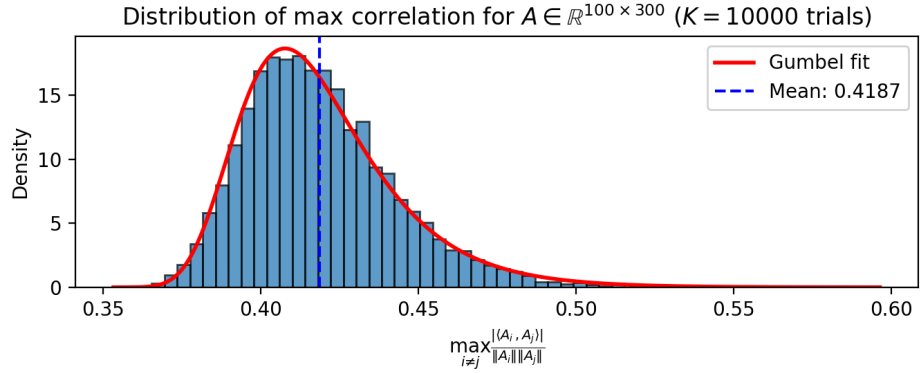
where μ is the mode of the distribution (location parameter) and β is the scale parameter (proportional to the standard deviation).

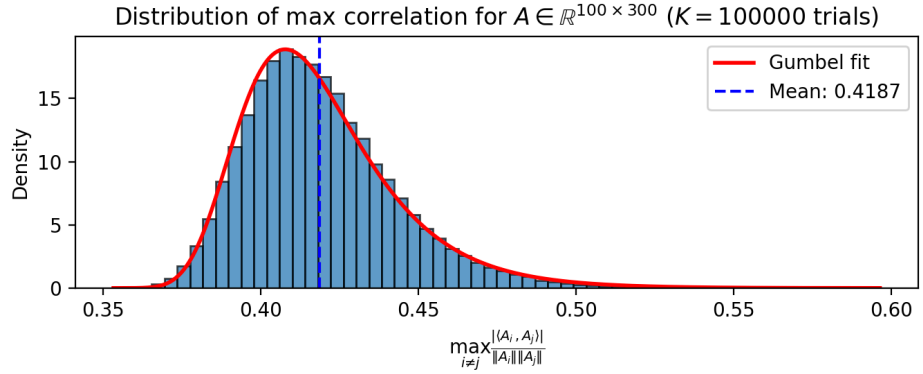
4.2. Analysis of the Histograms











The histograms generated, especially for large K (e.g., $K = 10000$ and $K = 100000$ as shown in Figure 18 and Figure 19, respectively), exhibit the distinct features of a Gumbel distribution:

- A single peak (unimodal).
- Asymmetry with a more extended tail on the right side.

As we can observe, growing K *(number of trials)* leads to a smoother plot and a clearer shape of the distribution, which aligns with the theoretical expectations of the Gumbel distribution.

The observed mode of the distribution is around 0.42, which is consistent with theoretical predictions. The location parameter μ can be approximated by:

$$\mu \approx \sqrt{\frac{2\ln(N)}{m}} = \sqrt{\frac{2\ln\left(\frac{n(n-1)}{2}\right)}{m}} \tag{22}$$

This formula arises from the well-known approximation for the expected maximum of N standard normal variables ($\sqrt{2\ln N}$), applied to our standardized variables $\{|\sqrt{m}C_{ij}|\}$.

For $m = 100$ and $n = 300$, we have $N = 44850$:

$$\mu \approx \sqrt{\frac{2\ln(44850)}{100}} \approx \sqrt{\frac{2 \cdot 10.71}{100}} = \sqrt{0.214} \approx 0.462 \tag{23}$$

This theoretical approximation gives a value in the general vicinity of the observed peak (around 0.42). The discrepancy arises, and will be more evident when discussing convergence at Section 5.2, because the variables $\{C_{ij}\}$ are not perfectly independent (for instance, $C_{1,2}$ and $C_{1,3}$ both depend on column A_1) and their distribution is only approximately normal. Nonetheless, this formula correctly shows that the peak of the distribution is determined by the dimensions m and n.

In conclusion, the observed distribution is a **Gumbel distribution**. This arises because we are plotting the maximum of a very large number of approximately independent, normally-distributed random variables (the cosine similarities).

5. Complexity

5.1. Algorithm Complexity and Runtime

The complexity of the algorithm is determined by the main operations within each of the K iterations.

The process begins by generating a Gaussian matrix of size $m \times n$, which has a time complexity of $O(mn)$. We then calculate the L2-norm for n columns of length m using `norms = np.linalg.norm(A, axis=0)`, an operation with $O(mn)$ complexity. The most computationally expensive step is the calculation of the Gram Matrix via `G = A.T @ A`. This matrix multiplication of an n x m matrix with an m x n matrix has a complexity of $O(mn^2)$. Subsequent operations, including the outer product ($O(n^2)$), element-wise division ($O(n^2)$), and maximum extraction ($O(n^2)$), are less expensive.

The total complexity for a single iteration is the sum of these steps, dominated by the Gram matrix calculation:

$$O(\text{One Iteration}) = O(mn) + O(mn) + O(mn^2) + O(n^2) = O(mn^2) \tag{24}$$

Therefore, for K iterations, the total complexity of our algorithm is $O(Kmn^2)$. This implies that the runtime should scale linearly with K and m, and quadratically with n. We can verify this empirically.

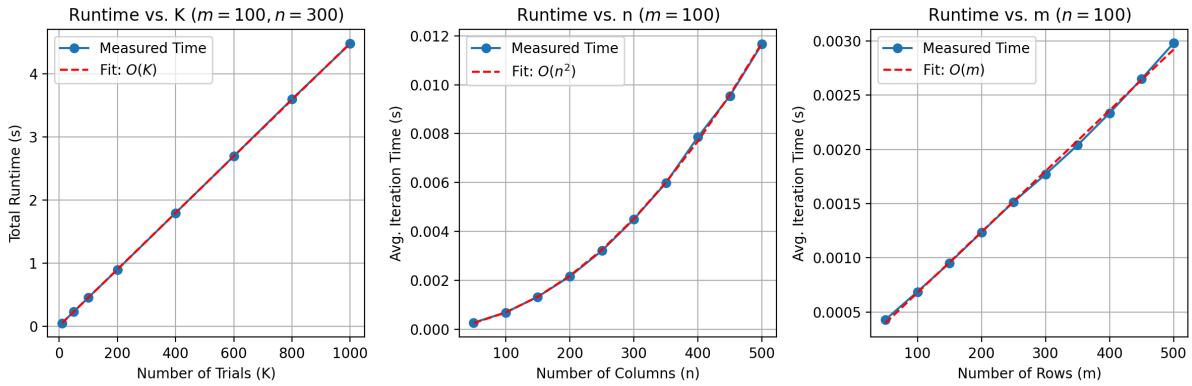


Figure 20: Runtime for varying K, n, and m

As predicted, Figure 20 confirms our theoretical model. The plots show that the runtime scales linearly with K and m, and quadratically with n. This empirically verifies the algorithm's overall complexity.

5.2. Algorithm Convergence and Choosing an Appropriate K

The question *"What value of K is good for a good estimate of the expected maximum?"* is about statistical convergence. K represents our sample size, which must be large enough for our statistics (like the mean and the histogram's shape) to be stable and reliable.

We first examine this convergence for the baseline case of $(m = 100, n = 300)$. To compute these maxima over many iterations (up to $K = 10^5$), we used Multiprocessing. A simple way to visualize convergence is to plot the running average of the maximum correlation as K increases. We expect this average to fluctuate for small K and converge to a stable value as K grows.

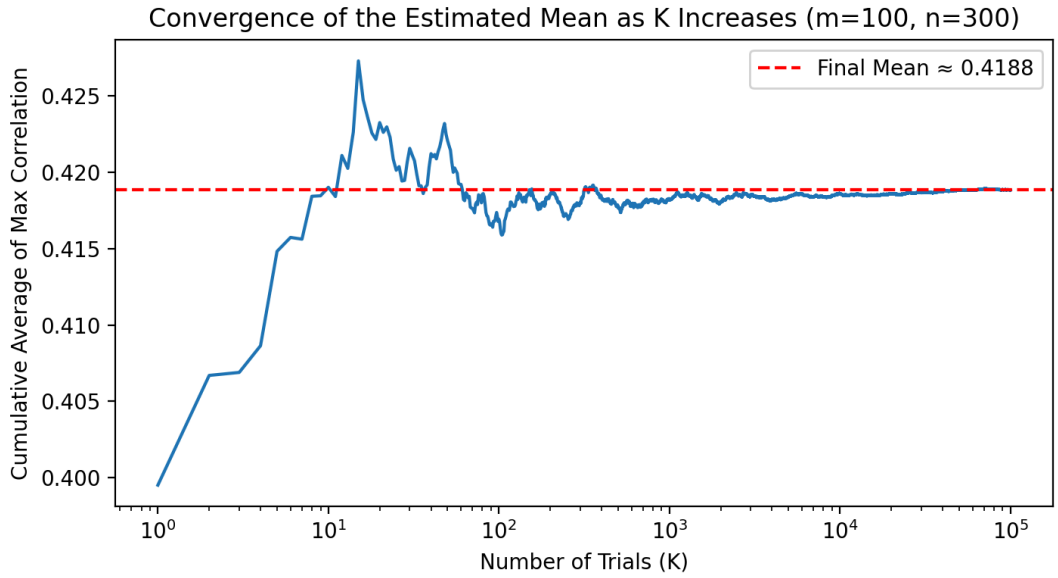


Figure 21: Convergence of eq. (18) as K grows to 10^5 with $(m, n) = (100, 300)$

From Figure 21, we can observe that, for this specific case of $(m = 100, n = 300)$, the mean of the maximum correlation converges to approximately 0.42 as K increases. The convergence can be summarized as follows:

- For $K < 100$, the estimate is very noisy and unreliable.
- For $100 \leq K < 1000$, the estimate begins to stabilize, despite some minor yet visible fluctuations.
- For $K \geq 1000$, our estimate becomes very stable and converges smoothly to ≈ 0.42.

A K value in the range of 10^3 to 10^4 is therefore a good choice for this specific problem. However, the rate of convergence for other pairs (m, n) may differ. The convergence speed depends on the variance of the underlying Gumbel distribution, which is proportional to the square of its scale parameter, β. This parameter can be derived from established results in Extreme Value Theory.

The derivation begins with the known asymptotic scale parameter for the maximum of N standard normal variables, $\beta_N \approx \frac{1}{\sqrt{2 \ln N}}$. In our case, the standardized variables are $Y_{ij} = \sqrt{m} C_{ij} \sim N(0, 1)$. Since our variable of interest, $M = \max|C_{ij}|$, is scaled by a factor of $\frac{1}{\sqrt{m}}$ relative to the maximum of the standardized variables, the scale parameter is also scaled by this factor:

$$\beta_{m,n} = \frac{\beta_N}{\sqrt{m}} \approx \frac{\frac{1}{\sqrt{2 \ln N}}}{\sqrt{m}} = \frac{1}{\sqrt{2m \ln(N)}} \tag{25}$$

This result implies that the variance is inversely proportional to $m \ln(n)$. This leads to the prediction that distributions for larger (m, n) pairs, having smaller variance, should converge more rapidly.

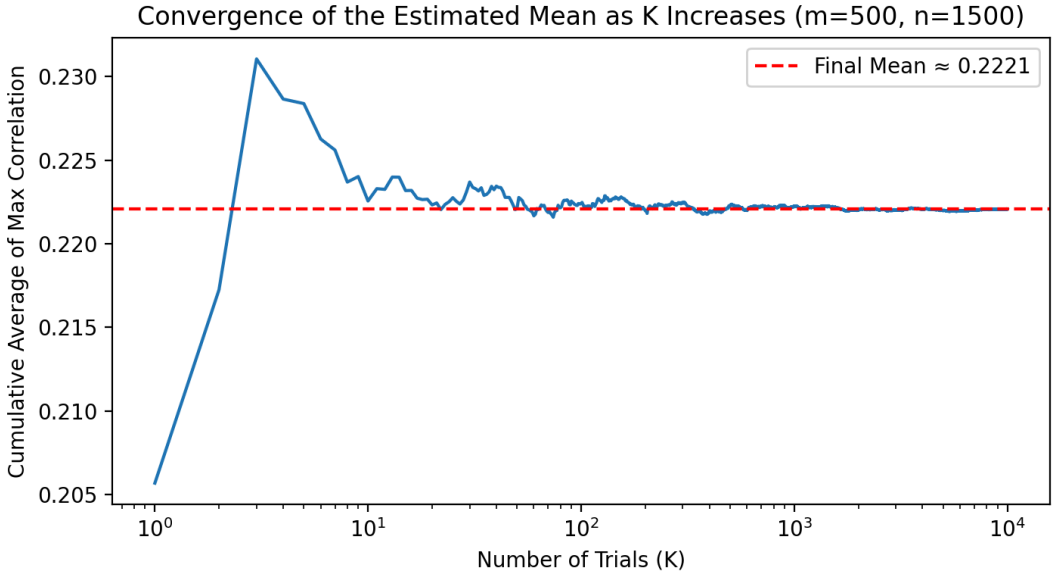


Figure 22: Convergence of eq. (18) as K grows to 10^4 with $(m, n) = (500, 1500)$

In Figure 22, we observe a similar convergence pattern for $(m = 500, n = 1500)$, but with a more rapid stabilization of the mean. The fluctuations are smaller, and the mean converges to ≈ 0.22.

There are two paths we can take to ensure convergence:

5.2.1. Method 1: Adaptive K

One could design an algorithm that stops once the running mean has stabilized. For instance, it could monitor the standard deviation of the running average over a recent window of trials and halt when this stability metric falls below a set tolerance.

- **Pros:** This approach is computationally efficient, performing only the necessary number of trials for each case. A clear advantage is that it adapts to the specific convergence behavior of each (m, n) pair, ensuring each pair has a "good enough" K with a clean stopping criterion.
- **Cons:** It introduces new arbitrary parameters (the window size and the tolerance threshold, and justifying those can become just as arbitrary as justifying a fixed K) and its sequential nature makes it difficult to parallelize efficiently.

5.2.2. Method 2: Fixed K

The second approach is to select a single, fixed, and sufficient value of K for all experiments.

- **Pros:** This method is simple to implement, easily parallelizable (as the number of tasks is known beforehand), and ensures all results are directly comparable under identical sampling conditions. By choosing a K that is sufficient for our "noisiet" case, we can guarantee that all other cases are also more than sufficiently sampled. There is no doubt about the statistical stability of any of your results.
- **Cons:** It is less computationally efficient, as it may "over-sample" cases that converge quickly.

The main goal of this project is to analyze and report on the statistical properties of random matrices, not to develop or showcase the most efficient simulation algorithm., hence why, for this project, we chose Method 2. The key is to select a K based on a "worst-case" or slower-converging scenario. Since the smaller (m, n) pairs are more variable, a K value that is sufficient for them will be more than adequate for all larger, more stable pairs. A choice of $K = 2500$ was therefore selected as a robust value that guarantees all our experiments are well-converged and our comparisons are statistically sound. It also produces results that are easier to interpret and compare across different (m, n) pairs.

This value of K is sufficient to ensure convergence for the more variable cases (like $(m = 100, n = 300)$ and $(m = 100, n = 100)$) and is more than adequate for all other pairs (m, n) we tested.

This justifies our choice of a fixed $K = 2500$ for the final experiments, as this value is sufficient for the more variable cases and thus more than adequate for all others. Nonetheless, the method of adaptive K is a valid alternative and was also implemented in the notebook, allowing users to explore convergence behavior interactively. Here is the distribution of the maximum correlation for $(m, n) = (100, 100)$, using the adaptive method with a window size of 25 and a tolerance of 10^{-5}:

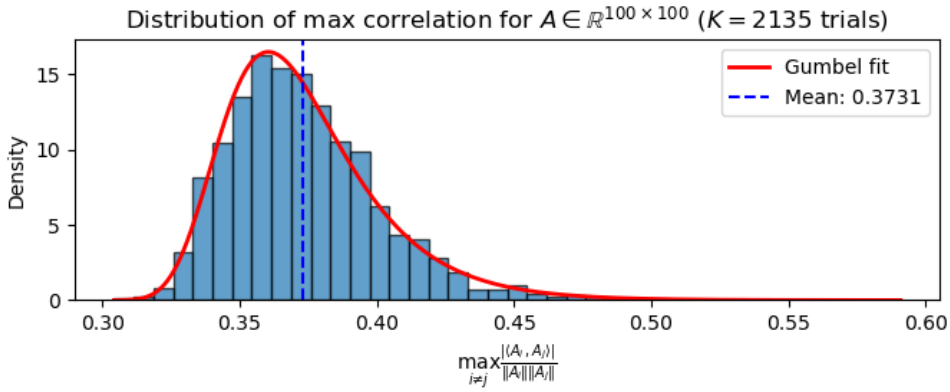


Figure 23: Distribution of eq. (18) for $(m, n) = (100, 100)$. window $= 25$ tolerance $= 10^{-5}$

For this case, the adaptive method stopped after $K = 2135$ iterations. All other pairs (m, n) should therefore converge with a similar or smaller K.

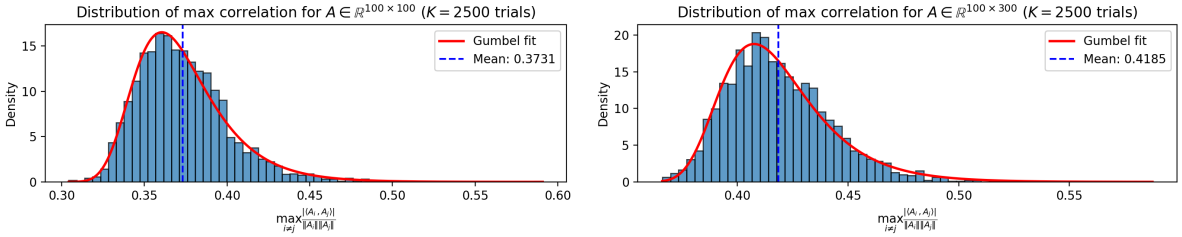
6. Analysis of Maximum Correlation for Varying Dimensions

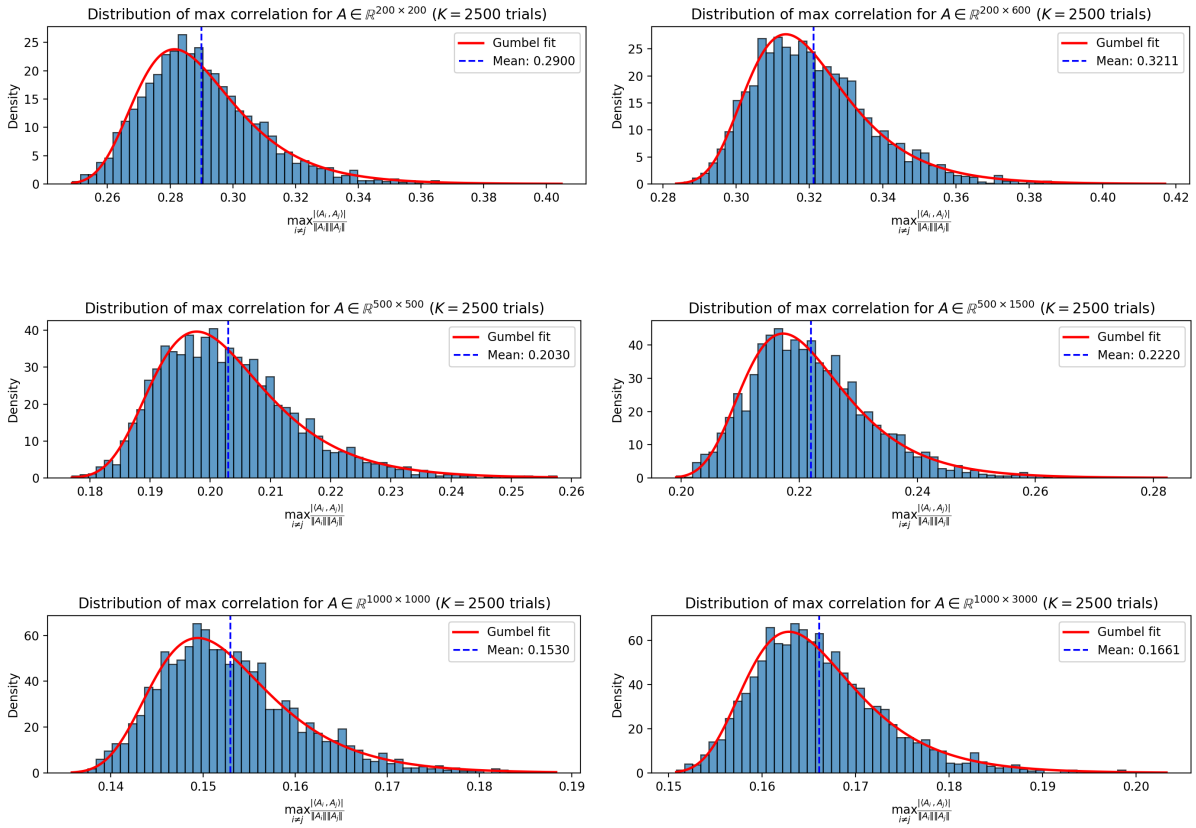
Once more with:

$$M_{m,n} = \max_{i \neq j} \frac{|\langle A_i, A_j \rangle|}{\|A_i\| \|A_j\|}, A \in \mathbb{R}^{m \times n}, A_{ij} \sim N(0, 1) \tag{26}$$

Having established that the maximum correlation follows a Gumbel distribution Section 4 and determined a robust methodology for sampling it Section 5.2, we now broaden our investigation. This section analyzes how the parameters of the Gumbel distribution change for various matrix dimensions, revealing the interplay between the ambient dimension, m, and the number of vectors, n.

For each pair (m, n), we generated $K = 2500$ i.i.d. realizations of $M_{m,n}$. As established in Section 5.2, this fixed number of trials ensures that our results are statistically robust and directly comparable across all configurations. However, if you are interested in seeing the results of the adaptive method at Section 5.2.1, they can be found at the second to last cell of the notebook. The parallel method was used for these computations, and a Gumbel distribution was fitted to each resulting histogram:





We know from eq. (20) that the distribution of $C_{ij} = \frac{\langle A_i, A_j \rangle}{\|A_i\| \|A_j\|}$ is approximately $N\left(0, \frac{1}{m}\right)$. So taking the absolute value gives a folded normal distribution that decays like $\exp\left(\frac{-m\varphi^2}{2}\right)$.

The maximum of $N = \frac{n(n-1)}{2}$ converges to a Gumbel law, as discussed in Section 4.1. Hence every histogram has the same shape, a sharp mode with a long tail, regardless of (m, n). Only the 2 Gumbel parameters change:

$$\begin{aligned} \mu_{m,n} &= \sqrt{\frac{2\ln(N)}{m}} \\ \beta_{m,n} &= \frac{1}{\sqrt{2m \log N}} \end{aligned} \tag{27}$$

where $N = \frac{n(n-1)}{2}$ is the number of distinct pairs (i, j).

The derivation of β is analogous to the one for μ.

One could note that the mode μ moves as m or n are fixed. To better understand this. Set $N = \frac{n(n-1)}{2}$ and $\sigma^2 = \frac{1}{m}$. With only the dominant terms in eq. (27):

$$\mu_{m,n} \approx \sigma\sqrt{2\log N} = \sqrt{\frac{2\log N}{m}} \tag{28}$$

These equations explain the trends visible in the plots. The mode μ is influenced by two competing factors: it increases very slowly with the number of vectors (n, as $\sqrt{\ln n}$), but decreases more significantly with the dimension of the space (m, as $\frac{1}{\sqrt{m}}$). This tells us that increasing dimensionality has a stronger effect on reducing the maximum correlation than increasing the number of vectors has on raising it.

7. Conclusion

This report systematically investigated the geometric properties of random Gaussian matrices, confirming that their non-orthogonality is governed by a competition between the number of vectors (n) and the dimension of the space (m). We showed that the L2-norms of columns concentrate sharply around $\sqrt{m}$ (Chi distribution), and that inner products between columns are approximately Normal, reflecting near-orthogonality in high dimensions. The maximum correlation, quantifying the greatest non-orthogonality, follows a Gumbel extreme value law, with parameters determined by (m, n).

1. **Increasing the number of vectors** (n) for a fixed dimension (m) increases the expected maximum correlation.
2. **Increasing the dimension of the space** (m) decreases the expected maximum correlation, even when the number of vectors n increases proportionally. That is, it has stronger effect in reducing the maximum correlation than increasing n has in raising it.

Thus, the "blessing of dimensionality" ensures that large random matrices, despite their randomness, exhibit highly predictable and quantifiable structure.