

# Projeto 4 — Álgebra Linear Numérica

Leonardo Verissimo e Jaime Willian

Monitora: Beatriz Lucia Teixeira de Souza

FGV EMAp

21 de junho de 2025

# Sumário

<b>1</b>	<b>Questão 1: Matrizes Gaussianas</b>	<b>3</b>
1.1	a) Distribuição das normas . . . . .	3
1.2	b) Produtos internos . . . . .	9
1.3	c) A distribuição do máximo . . . . .	11
1.4	d) Complexidade computacional e estimativa de $K$ . . . . .	12
1.5	e) A distribuição do máximo, parte 2 . . . . .	17
<b>2</b>	<b>Demonstrações</b>	<b>20</b>

# 1 Questão 1: Matrizes Gaussianas

## 1.1 a) Distribuição das normas

Ao realizar experimentos com matrizes aleatórias cujas entradas são independentes e identicamente distribuídas segundo  $\mathcal{N}(0, 1)$ , obtivemos os histogramas mostrados nas Figuras. Observa-se que a distribuição das normas das colunas apresenta um contorno aproximadamente em forma de sino, sugerindo convergência para uma curva gaussiana.

Para caracterizar esse comportamento de modo exato, considere a norma ao quadrado de uma coluna  $A_i$  de dimensão  $m$ :

$$\|A_i\|_2^2 = \sum_{k=1}^m a_{k,i}^2, \quad a_{k,i} \sim \mathcal{N}(0, 1).$$

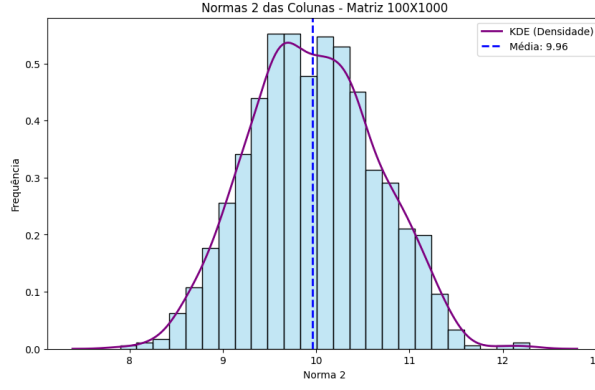
Pelo fato de cada  $a_{k,i}^2$  seguir uma distribuição  $\chi_1^2$  e de essas variáveis serem independentes, conclui-se que

$$\|A_i\|_2^2 \sim \chi_m^2,$$

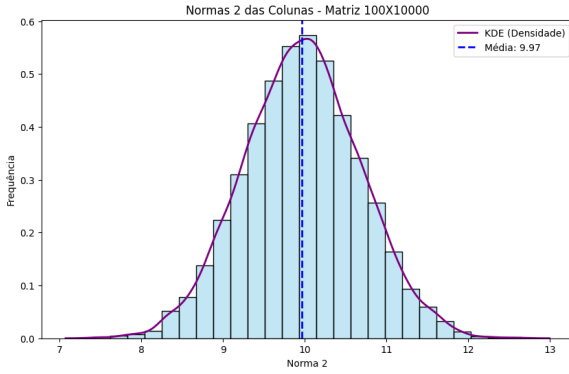
isto é, uma distribuição qui-quadrado com  $m$  graus de liberdade. Note que a distribuição qui-quadrado também pode ser representada por uma distribuição  $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ , dessa forma, teremos que  $\mathbb{E}[\|A_i\|_2^2] = m$ .

**Lema 1.1** (2). *Para valores grandes de  $m$ , se  $X \sim \chi_m^2$ , então:*

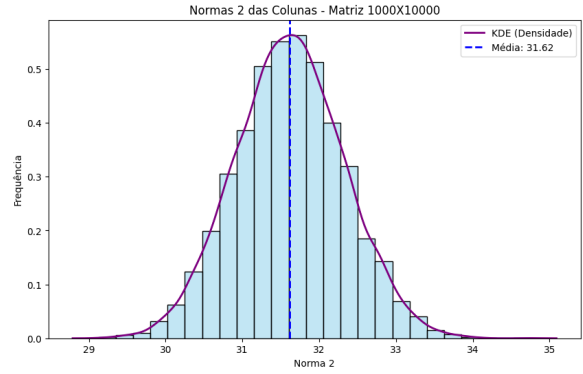
$$\mathbb{E}[\sqrt{X}] = \sqrt{2} \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)} \sim \sqrt{m}$$



(a) Figura 1: Matriz 100x1000



(b) Figura 2: Matriz 100x10000



(c) Figura 3: Matriz 1000x10000

Figura 1: Distribuição das normas em Matrizes Gaussianas onde  $M < N$ .

1. **Formato da Distribuição:** Em todos os cenários apresentados (Matrizes Gaussianas onde  $M < N$ ), a distribuição das normas-2 das colunas se assemelha a uma **distribuição normal (gaussiana)**. Dos 3 gráficos, o que deixa menos clara essa distribuição é a figura 1, referente a matriz gaussiana 100 x 1000, onde é visto que a distribuição tem 3 picos, sendo nenhum deles no *bin* da média 9.96. A distribuição fica evidente nas figuras 2 e 3, quando aumentamos o número de colunas, pela forma de sino dos histogramas e pela boa aderência das curvas de Estimativa de Densidade de Kernel (KDE <sup>1</sup>) a essa forma.
2. **Influência do Número de Linhas ( $m$ ) na Média da Norma:** Observamos uma forte correlação entre o número de linhas ( $m$ ) da matriz e a média da norma-2 das suas colunas.
  - Para as matrizes com  $m = 100$  (Figura 1: 100 x 1000 e Figura 2: 100 x 10000), a média das normas é 9.96 e 9.97, respectivamente. Note que  $\sqrt{100} = 10$ , o que é condizente com os valores observados.
  - Para a matriz com  $m = 1000$  (Figura 3: 1000 x 10000), a média das normas é 31.62. Novamente,  $\sqrt{1000} \approx 31.622$ , demonstrando que a média da norma-2 de um vetor gaussiano padrão de  $m$  dimensões se aproxima de  $\sqrt{m}$ .

<sup>1</sup>A KDE (Estimativa de Densidade de Kernel) é uma técnica não-paramétrica para estimar a função densidade de probabilidade de uma variável aleatória. No `seaborn.kdeplot`, esta curva suaviza os dados do histograma para fornecer uma representação contínua da distribuição, ajudando a visualizar a forma e a densidade dos pontos analisados.

Essa observação é uma característica fundamental das normas de vetores cujas entradas são independentes e identicamente distribuídas (i.i.d.) com média zero e variância unitária, como é o caso das entradas de uma matriz gaussiana padrão. A norma quadrática ( $\|x\|_2^2 = \sum x_i^2$ ) de um vetor gaussiano padrão de  $m$  dimensões segue uma distribuição qui-quadrado com  $m$  graus de liberdade, cuja esperança é  $m$ .

**Em conclusão,** a norma-2 das colunas de uma matriz gaussiana  $A_{m \times n}$  tende a se distribuir de forma aproximadamente gaussiana em torno de  $\sqrt{m}$ . À medida que  $m$  cresce, essa distribuição torna-se mais concentrada em torno de  $\sqrt{m}$ , enquanto  $n$  atua principalmente como o fator que determina o tamanho da amostra para a visualização da distribuição.

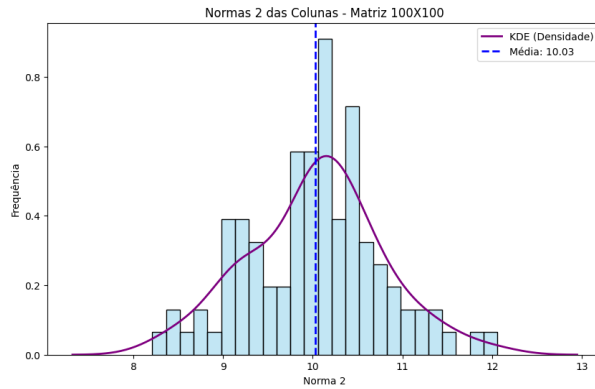


Figura 1

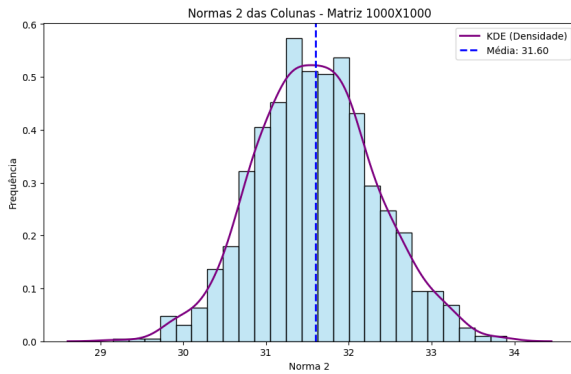


Figura 2

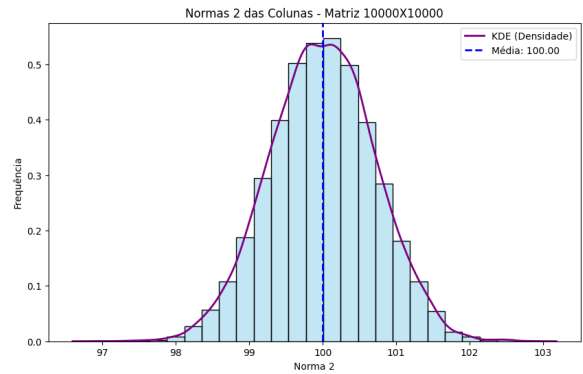


Figura 3

Figura 2: Distribuição das normas em Matrizes Gaussiana onde  $M = N$ .

Analisando agora os histogramas da norma-2 das colunas para matrizes gaussianas  $A_{m \times n}$  onde o número de linhas ( $m$ ) é igual ao número de colunas ( $n$ ), como apresentados nas Figuras 1 ( $100 \times 100$ ), 2 ( $1000 \times 1000$ ) e 3 ( $10000 \times 10000$ ), observamos as seguintes características:

1. **Formato da Distribuição:** Consistentemente com as observações anteriores, a distribuição das normas-2 das colunas mantém um formato que se aproxima de uma **distribuição normal (gaussiana)**. Na figura 1, referente a matriz  $100 \times 100$  não é tão visível devido a sua assimetria e dimensão baixa, mas quando aumentamos a dimensão da matriz (figura 2 e 3), a forma de sino do histogramas fica mais clara,

e as curvas de Estimativa de Densidade de Kernel (KDE) ajustam-se bem a essa forma nas escalas de dimensão.

2. **Média da Norma e a Relação com  $m$ :** A relação fundamental entre a média da norma-2 das colunas e o número de linhas ( $m$ ) é novamente observada:

- Para a matriz  $100 \times 100$  (Figura 1), a média observada é  $\approx 10.03$ , que é extremamente próxima de  $\sqrt{100} = 10$ .
- Para a matriz  $1000 \times 1000$  (Figura 2), a média é  $\approx 31.60$ , o que coincide perfeitamente com  $\sqrt{1000} \approx 31.62$ .
- Para a matriz  $10000 \times 10000$  (Figura 3), a média observada é  $\approx 100.00$ , que é exatamente  $\sqrt{10000} = 100$ .

Essa consistência reforça a propriedade de que a média da norma Euclidiana de um vetor aleatório cujas entradas são i.i.d. gaussianas padrão é aproximadamente  $\sqrt{m}$ , independentemente de a matriz ser quadrada ou retangular (considerando as colunas como vetores de  $m$  dimensões).

**Em conclusão,** os gráficos para o caso  $M = N$  reiteram e fortalecem as conclusões tiradas dos gráficos anteriores. A média das normas-2 das colunas de uma matriz gaussiana é robustamente predita por  $\sqrt{m}$ , e a distribuição dessas normas se torna cada vez mais concentrada em torno desse valor à medida que a dimensão  $m$  aumenta. A condição de  $M = N$  não altera o comportamento fundamental das normas das colunas; ela apenas significa que o número de vetores amostrados é igual à sua dimensão.

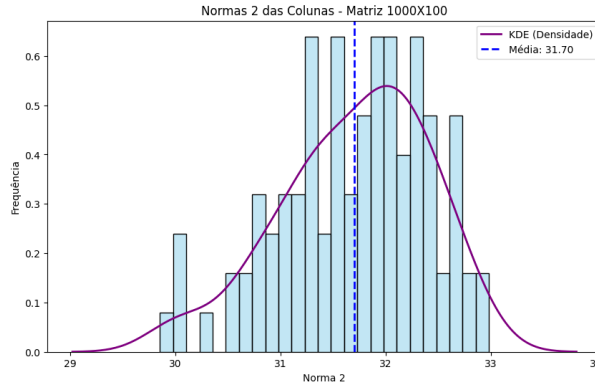


Figura 1

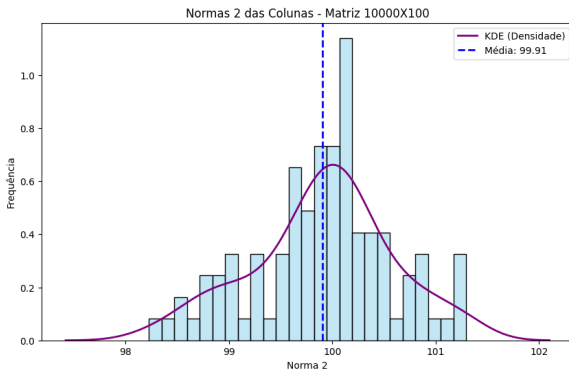


Figura 2

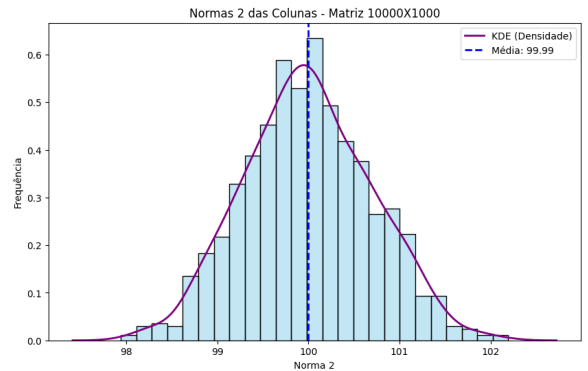


Figura 3

Figura 3: Distribuição das normas em Matrizes Gaussianas onde  $M > N$ .

Por fim, vamos analisar os casos de Matrizes Gaussianas em que  $M > N$ .

1. **Formato da Distribuição:** Em todas as configurações analisadas, as curvas de Estimativa de Densidade de Kernel (KDE) se assemelham a uma **distribuição normal (gaussiana)**. Sendo mais visível na figura 3 (matriz  $10000 \times 1000$ ).
2. **Média das Normas e sua Relação com  $m$ :** A observação mais proeminente e robusta é a forte correlação entre a média da norma-2 das colunas e a raiz quadrada do número de linhas ( $m$ ):
  - Para a matriz  $10000 \times 1000$  ( $m = 10000$ ), a média observada é  $\approx 99.99$ , o que é praticamente idêntico ao valor esperado de  $\sqrt{10000} = 100$ .
  - Para a matriz  $10000 \times 100$  ( $m = 10000$ ), a média observada é  $\approx 99.91$ , novamente muito próxima de  $\sqrt{10000} = 100$ .
  - Para a matriz  $1000 \times 100$  ( $m = 1000$ ), a média observada é  $\approx 31.70$ , condizente com o valor esperado de  $\sqrt{1000} \approx 31.62$ .

Essa consistência, observada nos cenários anteriores ( $m < n$  e  $m = n$ ), reforça a propriedade de que a média da norma Euclidiana de um vetor gaussiano padrão de  $m$  dimensões é aproximadamente  $\sqrt{m}$ , independentemente da relação entre  $m$  e  $n$ .

3. **Ao analisar a dispersão dos dados foi observado:**
  - **Variação de  $n$  para  $m$  fixo:** Ao comparar a matriz  $10000 \times 1000$  (figura 3) com a  $10000 \times 100$  (figura 2), ou seja, mesmo  $m = 10000$ , mas  $n$  diferente, observa-se que um  $n$  maior (1000) resulta em um histograma mais suave e uma estimativa de KDE mais precisa, devido ao maior número de amostras. No entanto, a largura intrínseca da distribuição (sua dispersão) para o mesmo  $m$  permanece a mesma.
  - **Variação de  $m$  para  $n$  fixo:** Ao comparar a matriz  $10000 \times 100$  com a  $1000 \times 100$  (mesmo  $n = 100$ , mas  $m$  diferente), nota-se que a distribuição para  $m = 10000$  é significativamente mais concentrada em torno de sua média ( $\approx 100$ ) do que a distribuição para  $m = 1000$  (que se concentra em  $\approx 31.70$ ). Ou seja, à medida que  $m$  (a dimensão do vetor coluna) aumenta, a distribuição das normas torna-se cada vez mais concentrada em torno de seu valor esperado de  $\sqrt{m}$ .

Em todos os cenários testados ( $M < N$ ,  $M = N$ ,  $M > N$ ), a distribuição empírica das normas-2 das colunas exibiu um contorno que se assemelha fortemente a uma **distribuição normal (gaussiana)**. Isso foi evidenciado pela forma de sino dos histogramas gerados e pela boa aderência das curvas de Estimativa de Densidade de Kernel (KDE) aos dados. Foi vista também uma grande relação entre a média da norma 2 das colunas e o número de linhas:

- Para matrizes com  $m = 100$ , as médias das normas observadas foram consistentemente próximas de 10, o que é exatamente  $\sqrt{100}$ .
- Para  $m = 1000$ , as médias observadas foram próximas de 31.62, condizente com  $\sqrt{1000} \approx 31.622$ .
- Para  $m = 10000$ , a média observada aproximou-se de 100, ou seja,  $\sqrt{10000}$ .

De forma análoga, observou-se uma forte correlação entre a dispersão das distribuição das normas e o número de linhas. À medida que  $m$  aumenta, a **distribuição das normas se torna mais concentrada** em torno de sua média ( $\approx \sqrt{m}$ ). Isso significa que, à medida que a dimensão do vetor coluna aumenta, a variabilidade relativa dos valores das normas diminui, tornando esses valores mais consistentes e previsíveis. Por fim, algumas consequências da variação dos número de amostras:

- Um valor maior de  $n$  (para um  $m$  fixo) resulta em histogramas mais **suaves e precisos** e em curvas KDE mais bem definidas. Isso ocorre porque uma amostra maior permite uma melhor aproximação da forma subjacente da distribuição.
- É importante notar que  $n$  não altera a média teórica nem a dispersão intrínseca da distribuição das normas (que dependem de  $m$ ), mas sim a **qualidade da estimativa empírica** dessas propriedades.



## 1.2 b) Produtos internos

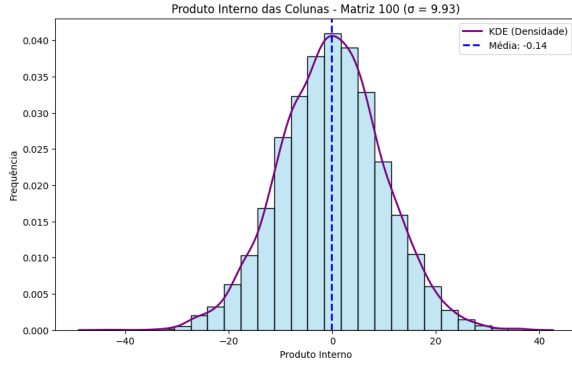


Figura 1

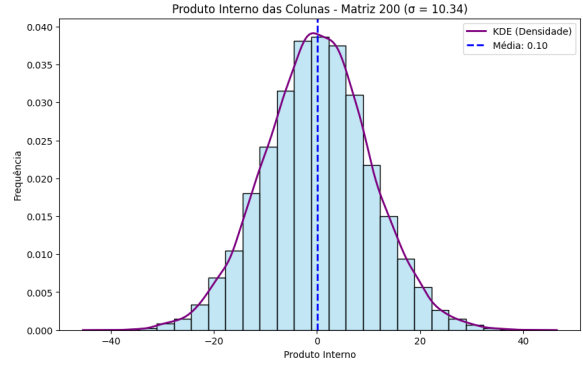


Figura 2

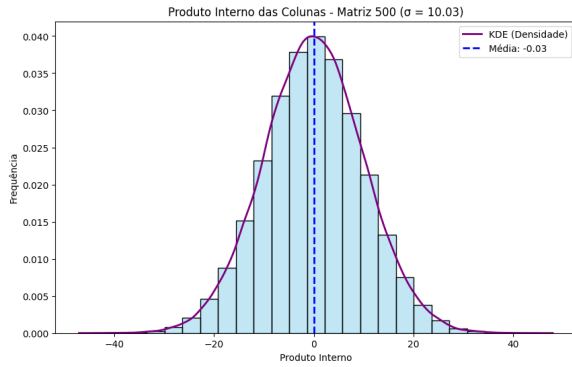


Figura 3

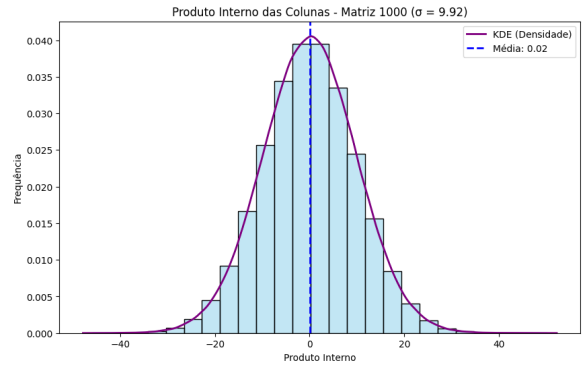


Figura 4

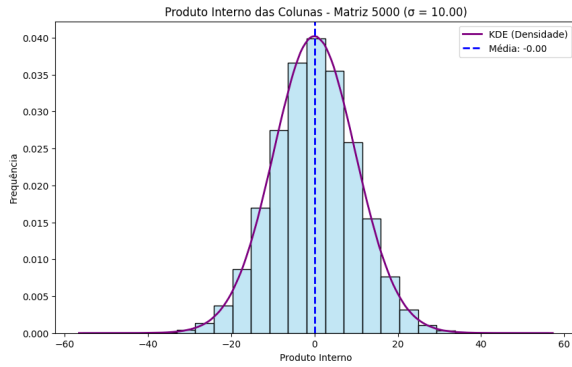


Figura 5

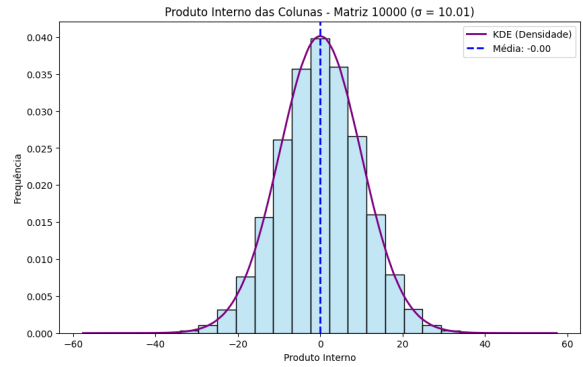


Figura 6

Figura 4: Produto interno de colunas de matrizes gaussianas, fixando ( $m = 100$ )

## Produtos Internos ( $\langle A_i, A_j \rangle$ )

1. **Formato da Distribuição:** Fixando o número de linhas  $m$  em 100 e variando o número de colunas  $n$  entre 100, 200, 500, 1000, 5000 e 10000, foi observado que em todos os casos apresentados, a distribuição dos produtos internos entre colunas distintas de uma matriz gaussiana exibiu uma forma de sino, assemelhando-se a uma **distribuição normal (gaussiana)**. As curvas de Estimativa de Densidade de Kernel (KDE) reforçam essa observação, ajustando-se bem aos histogramas.

**2. Média dos Produtos Internos:** A média dos produtos internos observados foi sempre próxima de 0:

- Para  $n = 100$ : Média  $\approx -0.14$
- Para  $n = 200$ : Média  $\approx 0.1$
- Para  $n = 500$ : Média  $\approx -0.03$
- Para  $n = 1000$ : Média  $\approx 0.02$
- Para  $n = 5000$ : Média  $\approx 0.0$
- Para  $n = 10000$ : Média  $\approx 0.0$

Este resultado já era esperado. Dado que as entradas da matriz gaussiana são *i.i.d* com uma distribuição  $\mathcal{N}(0, 1)$ , o produto interno de duas colunas distintas  $A_i$  e  $A_j$ ,  $\langle A_i, A_j \rangle = \sum_{k=1}^m A_{ki}A_{kj}$ , tem uma esperança de  $E[\langle A_i, A_j \rangle] = \sum_{k=1}^m E[A_{ki}A_{kj}]$ . Como  $A_{ki}$  e  $A_{kj}$  são independentes (para  $i \neq j$ ),  $E[A_{ki}A_{kj}] = E[A_{ki}]E[A_{kj}] = 0 \times 0 = 0$ . Portanto, a esperança do produto interno é  $m \times 0 = 0$ . As pequenas variações observadas são devidas à natureza amostral dos dados.

- Para  $n = 100$ :  $\sigma \approx 9.93$
- Para  $n = 200$ :  $\sigma \approx 10.34$
- Para  $n = 500$ :  $\sigma \approx 10.03$
- Para  $n = 1000$ :  $\sigma \approx 9.92$
- Para  $n = 5000$ :  $\sigma \approx 10$
- Para  $n = 10000$ :  $\sigma \approx 10.01$

**Lema 1.2 (2).** *A variância do produto interno entre dois vetores gaussianos padrão de dimensão  $m$  é  $m$ .*

Como as colunas das matrizes gaussianas analisadas possuem número de linhas  $m$  fixo em 100, implica que a variância do produto interno entre duas colunas distintas é 100. Consequentemente, o desvio padrão esperado é  $\sqrt{100} = 10$ , o que é condizente com os dados observados.

- **O que acontece com o aumento de  $n$ ?:** À medida que  $n$  (o número de colunas) aumenta, o número de pares distintos de colunas para os quais o produto interno é calculado aumenta significativamente ( $N(N-1)/2$ ). Isso resulta em:
  - Uma representação mais suave e precisa da distribuição subjacente no histograma e na curva KDE.
  - As estimativas da média e do desvio padrão (mostradas nos rótulos) convergem mais fielmente para seus valores teóricos de 0 e  $\sqrt{m}$  (10), respectivamente. Como pode ser visto na Figura 5 e 6.
- **Qual parece ser a distribuição para  $n \rightarrow \infty$ ?:** Com base nas observações de que a média se aproxima de zero e o desvio padrão se aproxima de  $\sqrt{m}$  (10), e que a forma da distribuição é consistentemente normal, a distribuição para  $n \rightarrow \infty$  de produtos internos entre colunas distintas de uma matriz gaussiana  $A_{m \times n}$  (com  $m = 100$ ) parece ser uma **distribuição normal com média 0 e desvio padrão  $\sqrt{m}$** , ou seja  $\mathcal{N}(0, m)$ .

### 1.3 c) A distribuição do máximo

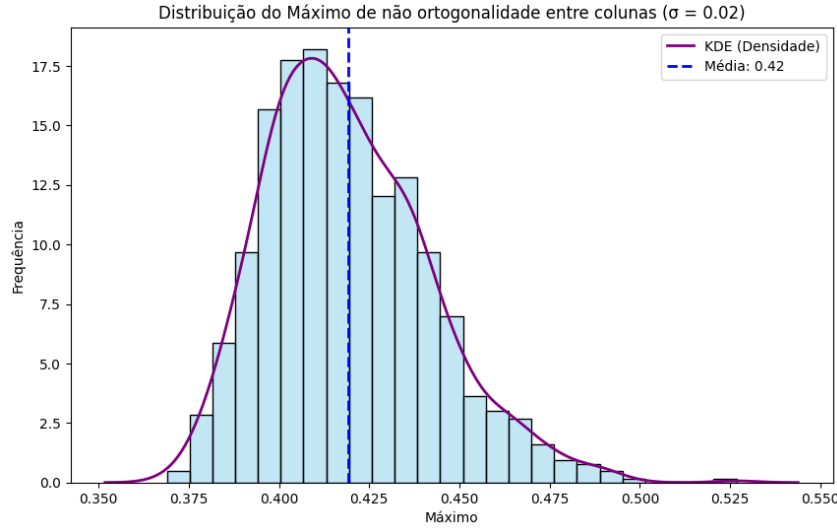


Figura 5: Distribuição do máximo da não-ortogonalidade entre colunas de matrizes gaussianas  $100 \times 300$  ( $K = 1000$  amostras).

Os valores máximos de não ortogonalidade (representados por  $\frac{|\langle A_i, A_j \rangle|}{\|A_i\| \|A_j\|}$ ) correspondem aos maiores valores absolutos do cosseno do ângulo de quaisquer pares de colunas distintas ( $A_i$  e  $A_j$ ) das matrizes gaussianas geradas. Um valor de 0 indica ortogonalidade perfeita, enquanto 1 indica colinearidade perfeita.

Tabela 1: Estatísticas Descritivas do Máximo da Não-Ortogonalidade

Métrica	Valor Observado
Média ( $\bar{x}$ )	0.4184
Moda ( $m_o$ )	0.3639
Mediana ( $m_d$ )	0.4152

**Formato da Distribuição:** Seja  $x$  a distribuição observada para o máximo da não-ortogonalidade, é visto no gráfico que ela possui um único pico e é **assimétrico a direita**[1], apresentando uma cauda mais longa estendendo-se para valores maiores (à direita). Ou seja:

$$\bar{x} > m_d > m_o;$$

A distribuição de  $X$  é limitada superiormente por 1 (colinearidade perfeita). A assimetria para a direita é observado devido à cauda alongada em direção a valores maiores, que representa a rara ocorrência de pares de vetores quase colineares/paralelos.

**Média da Distribuição do Máximo:** A média dos  $K = 1000$  valores máximos calculados é de aproximadamente 0.42.

- Isso significa que, para matrizes gaussianas de dimensão  $100 \times 300$ , o par de colunas *menos ortogonal* tipicamente apresenta um valor absoluto do cosseno do ângulo de cerca de 0.42. Tal valor de cosseno corresponde a um ângulo de aproximadamente  $\arccos(0.42) \approx 65.17^\circ$ . Este resultado para o *máximo* indica que sempre existe um par de colunas que se desvia consideravelmente da ortogonalidade perfeita ( $90^\circ$ ).

## 1.4 d) Complexidade computacional e estimativa de K

A função `distribuicao_do_maximo` calcula a maior "não-ortogonalidade" entre pares de colunas de uma matriz gaussiana. Dada uma matriz  $A \in \mathbb{R}^{m \times n}$ , o procedimento normaliza as colunas (custo  $\mathcal{O}(mn)$ ), e em seguida calcula o produto interno entre todas as colunas normalizadas, equivalente à multiplicação  $X^\top X$ , onde  $X$  é a matriz com colunas unitárias. Essa multiplicação envolve uma matriz de dimensões  $n \times m$  com uma  $m \times n$ , resultando em custo dominante de  $\mathcal{O}(mn^2)$ .

Como temos que repetir esse processo para  $K$  matrizes independentes, a função é chamada  $K$  vezes. Sendo assim, a complexidade total do processo é:

$$\mathcal{O}(Kmn^2)$$

O fator  $K$  aparece porque cada uma das  $K$  execuções realiza todas as operações de forma independente. Assim, o custo total cresce linearmente com o número de amostras utilizadas para estimar o valor esperado do máximo.

Seja  $X$  uma variável aleatória definida, em um único experimento, como:

$$X = \max_{i \neq j} \left\{ \frac{|\langle A_i, A_j \rangle|}{\|A_i\| \|A_j\|} \right\}$$

Estamos interessados em estimar:

$$\mu = \mathbb{E}[X]$$

através da execução de  $K$  réplicas independentes, considerando a média amostral:

$$\bar{X}_K = \frac{1}{K} \sum_{k=1}^K X_k.$$

Desejamos determinar  $K$  tal que  $\bar{X}_K$  aproxime-se suficientemente de  $\mu$  com alta probabilidade. Pela Lei Fraca dos Grandes Números, para todo  $\varepsilon > 0$ :

$$\lim_{K \rightarrow \infty} P(|\bar{X}_K - \mu| \geq \varepsilon) = 0,$$

o que garante convergência em probabilidade.

Supondo  $\text{Var}(X) = \sigma^2 < \infty$ , a desigualdade de Chebyshev fornece para todo  $\varepsilon > 0$ :

$$P(|\bar{X}_K - \mu| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X}_K)}{\varepsilon^2} = \frac{\sigma^2/K}{\varepsilon^2} = \frac{\sigma^2}{K\varepsilon^2}.$$

Para que esta probabilidade seja limitada por  $\delta$ , basta tomar:

$$\frac{\sigma^2}{K\varepsilon^2} \leq \delta \quad \implies \quad K \geq \frac{\sigma^2}{\delta\varepsilon^2}.$$

Uma abordagem alternativa utiliza o Teorema Central do Limite. Para  $K$  suficientemente grande:

$$\frac{\bar{X}_K - \mu}{\sigma/\sqrt{K}} \xrightarrow[K \rightarrow \infty]{d} Z, \quad \text{onde } Z \sim \mathcal{N}(0, 1).$$

Assim, para  $z_{\alpha/2}$  o quantil da normal padrão:

$$P(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1 - \alpha,$$

$$P\left(-z_{1-\alpha/2}\frac{\sigma}{\sqrt{K}} \leq \bar{X}_K - \mu \leq z_{1-\alpha/2}\frac{\sigma}{\sqrt{K}}\right) \approx 1 - \alpha.$$

$$P\left(|\bar{X}_K - \mu| \leq z_{1-\alpha/2}\frac{\sigma}{\sqrt{K}}\right) \approx 1 - \alpha.$$

Fixando a semi-largura do intervalo de confiança como  $\varepsilon$ :

$$z_{1-\alpha/2}\frac{\sigma}{\sqrt{K}} = \varepsilon \quad \implies \quad K = \left(\frac{z_{1-\alpha/2} \cdot \sigma}{\varepsilon}\right)^2.$$

Por exemplo, para um intervalo de confiança de 95% ( $\alpha = 0.05$ ), temos  $z_{0.975} \approx 1.96$ , resultando em:

$$K \approx \left(\frac{1.96 \cdot \sigma}{\varepsilon}\right)^2.$$

Em seguida, precisamos de um valor aproximado para  $\sigma^2$ . Para isso, utilizaremos um estimador não enviesado para  $\sigma^2$ [\[2\]](#):

$$E(S^2) = \sigma^2, \text{ onde } S^2 = \frac{1}{K-1} \sum_{i=1}^n (X_i - \bar{X}_K)^2.$$

Dessa forma, implementamos as seguintes funções em Python:

```

1  def compute_S2(dados):
2      S2 = 0
3      for i in range(len(dados)):
4          Xbar = np.mean(dados[:i+1])
5          S2 += (dados[i] - Xbar) ** 2
6      S2 /= len(dados) - 1
7      return S2

1  def encontrar_K_teorico(m, n, epsilon, alpha, K0=100, max_iter
   =1000):
2      z = stats.norm.ppf(1 - alpha/2)
3      K = K0
4      dados = np.array(
5          [distribuicao_do_maximo(A=None, create_matrix=True,
6           m=m, n=n) for _ in range(K)]
7      )
8
9      for it in range(max_iter):
10         S2 = compute_S2(dados)
11         S = np.sqrt(S2)
12         K_req = int(np.ceil((z * S / epsilon)**2))
13
14         if K_req <= K:
15             ME = z * S / np.sqrt(K)
16             print(ME)
17             if ME <= epsilon:
18                 # convergiu dentro da margem desejada
19                 K = K_req
20                 break

```

```

21
22         # se ainda nao convergiu, geramos ao menos uma
           r plica extra
23         K_req = K + 1
24         novas = np.array([
25             distribuicao_do_maximo(A=None, create_matrix=True,
26             m=m, n=n) for _ in range(K_req - K)
27         ])
28         dados = np.concatenate([dados, novas])
29         K = K_req
30
31     Xbar = dados.mean()
32     S      = np.std(dados, ddof=1)
33     ME     = z * S / np.sqrt(K)
34     # calculamos estatisticas finais
35     return {
36         'K_final': K,
37         'Xbar':     Xbar,
38         'S':        S,
39         'ME':       ME,
40         'alpha':    alpha,
41         'epsilon':  epsilon,
42         'dados':    dados
43     }

```

```

1     def analise_K(m, n, epsilon, alpha, K_max):
2         # Calcula K teorico e tolerancia de erro
3         resultado = encontrar_K_teorico(m, n, epsilon, alpha)
4         tol = resultado['ME']
5         K_teo = resultado['K_final']
6         S = resultado['S']
7         if K_teo > K_max:
8             raise ValueError(f"K teorico ({K_teo}) e maior que o K
              maximo permitido ({K_max}). Ajuste K_max ou os
              parametros de entrada.")
9
10        # Inicializa arrays para armazenar maximos e medias
11        maximo = np.empty(K_max, dtype=float)
12        medias = []
13
14        # Gera matrizes e computa maximos
15        for k in range(1, K_max + 1):
16            A = generate_gaussian_matrix(m, n)
17            maximo[k - 1] = distribuicao_do_maximo(A)
18            medias.append(np.mean(maximo[:k]))
19
20        # Calcula valor esperado e encontra K_real
21        valor_esperado = np.mean(maximo[:K_teo])
22
23        # Encontra todos os valores de K dentro da tolerancia
24        k_values = [k for k in range(1, K_max + 1)

```

```

25         if abs(medias[k - 1] - valor_esperado) < tol]
26
27     # Determina K_real encontrando a regioao estavel
28     K_real = None
29     for k in range(1, K_max + 1):
30         if all(abs(medias[i] - valor_esperado) < tol
31                 for i in range(k, K_max)):
32             K_real = k
33             break
34
35     # Calcula limites de confianca
36     limite_superior = valor_esperado + tol
37     limite_inferior = valor_esperado - tol
38
39     # Plota os resultados
40     (...)
41
42     return {
43         'k_values': k_values,
44         'valor_esperado': valor_esperado,
45         'K_teo': K_teo,
46         'K_real': K_real,
47         'limite_superior': limite_superior,
48         'limite_inferior': limite_inferior
49     }

```

A implementação proposta calcula, de um lado, o valor teórico de  $K$  a partir do Teorema Central do Limite,

$$K_{\text{terico}} = \left( \frac{z_{1-\alpha/2} \sigma}{\varepsilon} \right)^2,$$

e, de outro, obtém uma estimativa empírica de  $K$  por simulação de Monte Carlo, ajustando dinamicamente o número de réplicas até que a margem de erro

$$ME = z_{1-\alpha/2} \frac{\hat{\sigma}}{\sqrt{K}}$$

fique abaixo de  $\varepsilon$ . Ao sobrepor essas duas curvas em função de  $K$ , observa-se consistentemente que

$$K_{\text{real}} \ll K_{\text{teorico}},$$

conforme ilustrado nas 8 e 7. Essa discrepância decorre do caráter conservador das desigualdades de Chebyshev e da aproximação assintótica do TCL, que garantem cobertura de probabilidade mesmo no “pior caso”, mas não exploram propriedades específicas de  $X$  — em particular, seu suporte limitado a  $[0, 1]$  e sua concentração mais forte do que a exigida pelos limites genéricos.

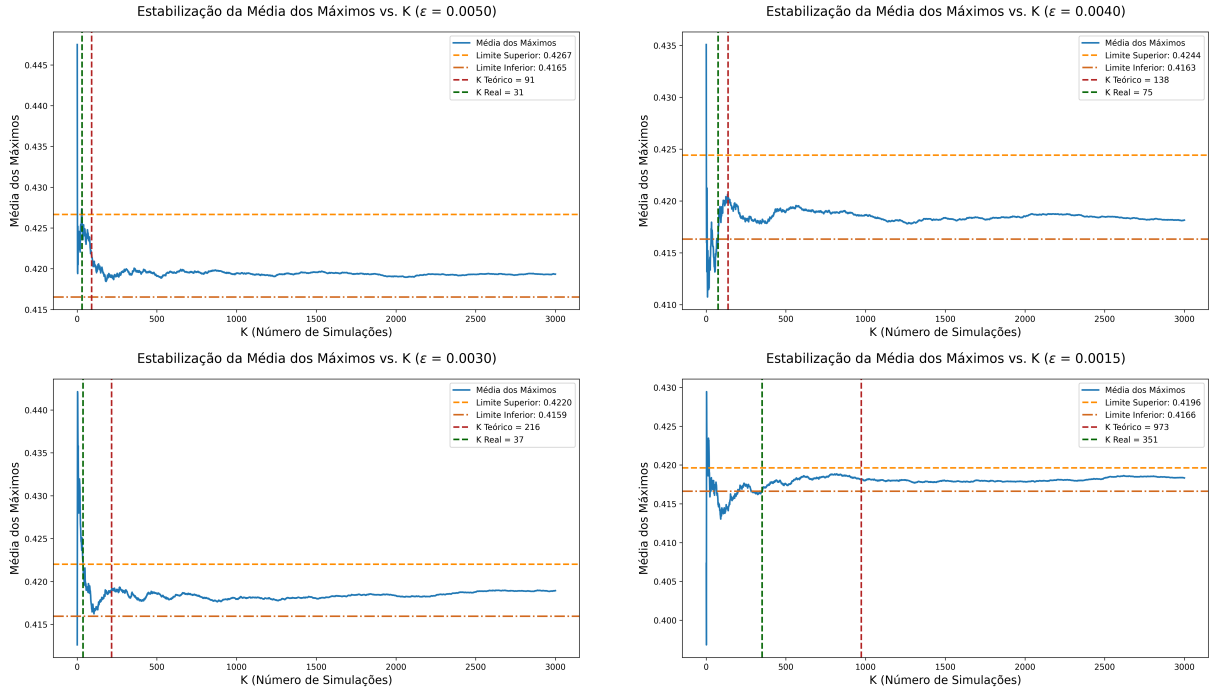


Figura 6: Comparação para diferentes valores de  $\epsilon$ .

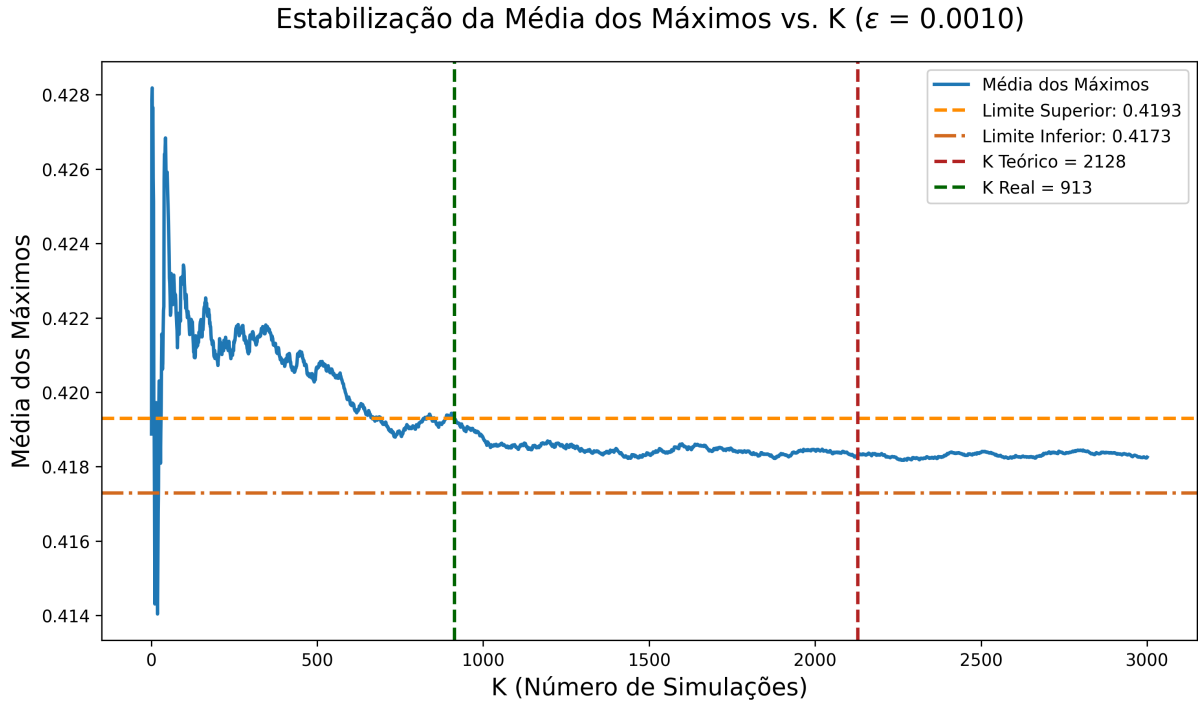


Figura 7: Comparação para um valor bem pequeno de  $\epsilon$ .

Nos experimentos realizados sobre matrizes gaussianas  $A \in \mathbb{R}^{100 \times 300}$ , verificou-se ainda que:

1. **Convergência da média:** à medida que  $K$  cresce, a trajetória de  $\bar{X}_K$  estabiliza-se dentro do intervalo  $[\mu \pm \epsilon]$ , confirmando a convergência em probabilidade garantida pela Lei Fraca dos Grandes Números.



2. **Dependência em  $\varepsilon^{-2}$ :** quanto menor o critério  $\varepsilon$ , maior o  $K$  necessário tanto no cálculo teórico quanto na prática — refletindo o fato de que  $\text{Var}(\bar{X}_K) = \sigma^2/K$  exige  $K \propto 1/\varepsilon^2$  para controlar o erro. Isso implica que reduzir  $\varepsilon$  pela metade leva a um aumento aproximado de  $K_{teorico}$  por um fator de 4, o que é compatível com os resultados observados.
3. **Eficiência do método empírico:** calibrando  $K$  com base em  $\hat{\sigma}$  e nas observações de  $\bar{X}_K$ , obtém-se estimativas muito mais enxutas e ainda assim confiáveis, pois o procedimento simula diretamente a variabilidade real de  $X$ .
4. **Valores empíricos de  $\bar{X}_K$ :** em seis repetições independentes foram obtidos os seguintes resultados:

Teste	$\bar{X}_K$
1	0,4186
2	0,4179
3	0,4189
4	0,4188
5	0,4174
6	0,4183
<b>Média</b>	<b>0,4183</b>

Tabela 2: Valores observados de  $\bar{X}_K$  em seis experimentos.

Assim, a comparação entre  $K_{teorico}$  e  $K_{real}$  evidencia não apenas o conservadorismo dos limites clássicos, mas também ressalta a eficiência prática de calibrar dinamicamente o número de réplicas por simulação, mantendo o nível de confiança desejado.

## 1.5 e) A distribuição do máximo, parte 2

A seguir, aplicando a função desenvolvida no item anterior com os valores de  $K_{teorico}$ , obtivemos o seguinte conjunto de resultados:

<b>m</b>	<b>n</b>	<b><math>K_{teorico}</math></b>
100	100	1357
100	300	973
200	200	589
200	600	484
500	500	219
500	1500	162
1000	1000	98
1000	3000	53

Tabela 3: Valores de  $K_{teorico}$  para diferentes pares  $(m, n)$ .

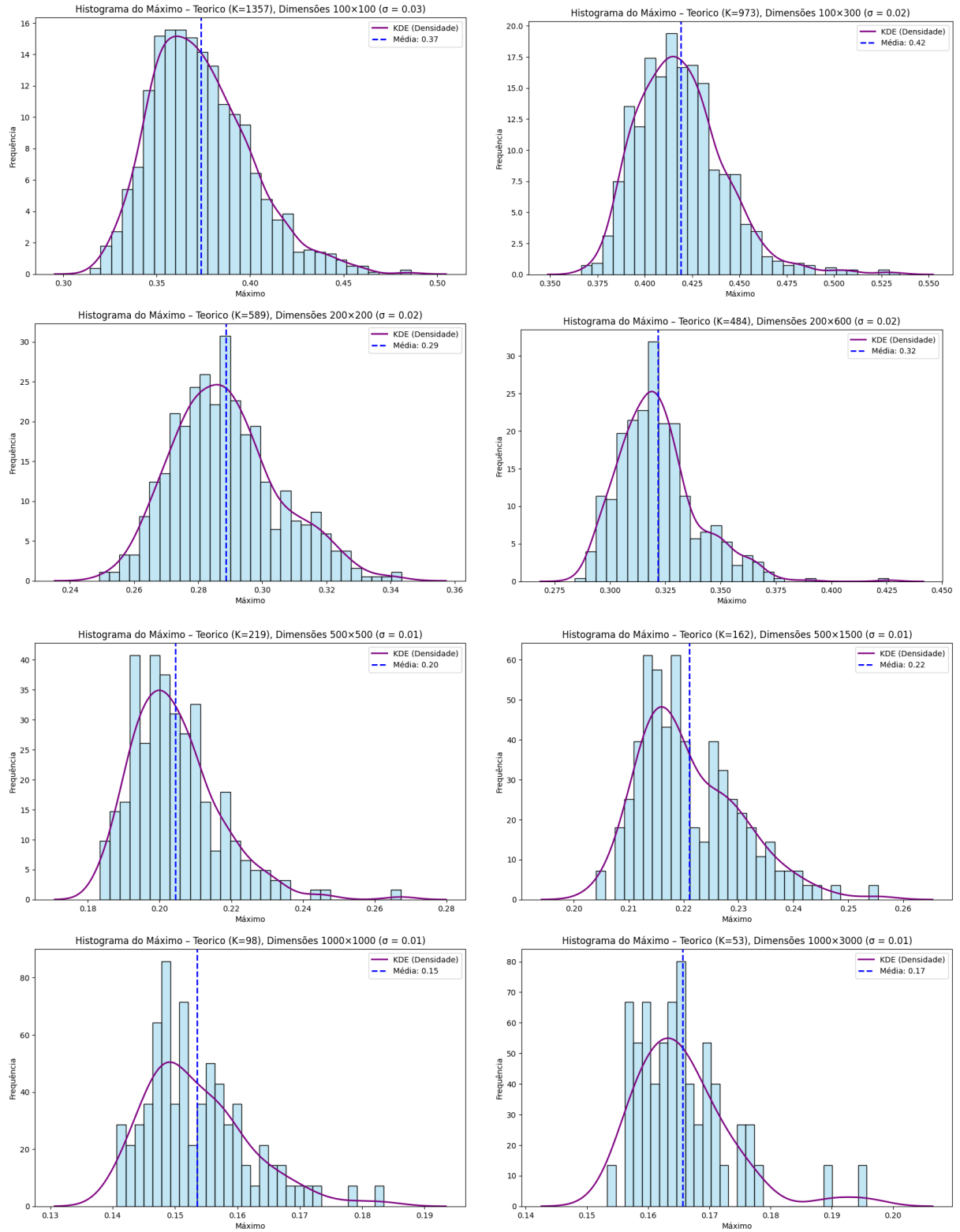


Figura 8: Comparação para diferentes valores de  $\varepsilon$ .

Analisando os gráficos, encontramos os seguintes resultados:

#### 1. Concentração da distribuição do máximo

- Para  $(m, n) = (100, 100)$ , o histograma do máximo de não-ortogonalidade é relativamente largo e pouco simétrico, indicando variabilidade significativa nos

valores de  $X$ . Isso exige um  $K$  alto ( $\approx 1357$ ) para reduzir o erro amostral abaixo de  $\varepsilon$ .

- Para  $(m, n) = (1000, 3000)$ , o histograma torna-se muito mais estreito e centrado em valores baixos de  $X$ , o histograma torna-se muito mais estreito e centrado em valores baixos de  $X$ , refletindo que, em altas dimensões, vetores gaussianos são quase ortogonais e o máximo de sobreposição cai. Assim, basta apenas  $\approx 53$  réplicas para obter a mesma precisão.

## 2. Efeito de aumentar $m$

- Comparando colunas fixas em  $n$ , elevar  $m$  reduz a variância intrínseca da não-ortogonalidade (mais “ortogonalidade” típica), pois  $\text{Var}(\langle A_i, A_j \rangle) = m$ . Como consequência, o desvio-padrão  $S$  do estimador amostral diminui e, pelo critério  $K_{teorico} \approx \left( \frac{zS}{\varepsilon} \right)$ , precisamos de muito menos réplicas para atingir a mesma margem de erro  $\epsilon$ . Assim, histogramas passam a ficar mais apertados em torno de valores pequenos, diminuindo  $\sigma$  e, consequentemente,  $K_{teorico}$ .

## 3. Efeito de aumentar $n$

- Com  $m$  fixo, crescer  $n$  (mais colunas) tende a elevar levemente o pico máximo — há mais pares para escolher o máximo — mas também torna o histograma mais “suave” (maior número de amostras de comparação), o que reduz a incerteza estimada. Por isso, mesmo aumentando  $n$  de 100 para 300 (mantendo  $m = 100$ ),  $K_{teorico}$  cai de 1357 para 973.

Com base nos experimentos realizados, foi possível observar que o número de réplicas  $K_{teorico}$  necessário para estimar com precisão o valor esperado do máximo de não-ortogonalidade entre colunas de uma matriz gaussiana depende fortemente das dimensões  $m$  e  $n$ . À medida que  $m$  aumenta, as colunas tendem a se tornar mais ortogonais entre si, reduzindo a variabilidade do máximo e, consequentemente, o número de amostras necessárias para garantir uma estimativa confiável. Da mesma forma, o aumento de  $n$  também contribui para a estabilização da estimativa, embora com efeito menos pronunciado. Esse comportamento evidencia como o crescimento da dimensionalidade influencia diretamente a concentração das distribuições envolvidas, permitindo reduzir significativamente o custo computacional sem comprometer a precisão estatística da estimativa.

## 2 Demonstrações

*Demonstração do Lema 1.1.*

$$\begin{aligned}\mathbb{E}[\sqrt{X}] &= \int_0^\infty \sqrt{x} f_X(x) dx \\ &= \int_0^\infty x^{\frac{1}{2}} \frac{1}{2^{m/2} \Gamma(\frac{m}{2})} x^{\frac{m}{2}-1} e^{-x/2} dx \\ &= \frac{1}{2^{m/2} \Gamma(\frac{m}{2})} \int_0^\infty x^{\frac{m+1}{2}-1} e^{-x/2} dx.\end{aligned}$$

Fazendo a mudança de variável:  $t = \frac{x}{2}$ ,  $dx = 2 dt$ , teremos que:

$$\int_0^\infty x^{\frac{m+1}{2}-1} e^{-x/2} dx = \int_0^\infty (2t)^{\frac{m+1}{2}-1} e^{-t} (2dt) = 2^{\frac{m+1}{2}} \int_0^\infty t^{\frac{m+1}{2}-1} e^{-t} dt.$$

Por definição,

$$\int_0^\infty t^{\frac{m+1}{2}-1} e^{-t} dt = \Gamma\left(\frac{m+1}{2}\right)$$

Por fim, juntamos tudo:

$$\mathbb{E}[\sqrt{X}] = \frac{1}{2^{m/2} \Gamma(\frac{m}{2})} \times 2^{\frac{m+1}{2}} \times \Gamma\left(\frac{m+1}{2}\right) = \sqrt{2} \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})}$$

Quando  $m \rightarrow \infty$ ,  $\Gamma\left(\frac{m+1}{2}\right) \sim \left(\frac{m}{2}\right)^{1/2} \Gamma\left(\frac{m}{2}\right)$ <sup>2</sup>, assim:

$$\mathbb{E}[\sqrt{X}] \sim \sqrt{2} \left(\frac{m}{2}\right)^{1/2} \sim \sqrt{m}$$

□

---

<sup>2</sup>Segue da aproximação assintótica de Stirling para a função Gamma,  $\Gamma(z) \sim \sqrt{2\pi} z^{z-1/2} e^{-z}$  e usando que  $\left(1 + \frac{a}{z}\right)^z \rightarrow e^a$ :

$$\Gamma(z+a) \sim \sqrt{2\pi} (z+a)^{z+a-1/2} e^{-(z+a)}, \text{ dividindo por } \Gamma(z), \frac{\Gamma(z+a)}{\Gamma(z)} \sim z^a.$$

*Demonstração do Lema 1.2.*

Considere dois vetores  $X = [X_1, \dots, X_m]^T$  e  $Y = [Y_1, \dots, Y_m]^T$ , onde todas as entradas  $X_k$  e  $Y_k$  são variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) de uma distribuição normal padrão  $\mathcal{N}(0, 1)$ . Além disso, as entradas de  $X$  são independentes das entradas de  $Y$ .

O produto interno entre  $X$  e  $Y$  é definido como  $\langle X, Y \rangle = \sum_{k=1}^m X_k Y_k$ .

Vamos calcular primeiramente a esperança:

1. A esperança de uma soma é a soma das esperanças:

$$E[\langle X, Y \rangle] = E\left[\sum_{k=1}^m X_k Y_k\right] = \sum_{k=1}^m E[X_k Y_k]$$

2. Como  $X_k$  e  $Y_k$  são independentes para cada  $k$ , temos que:

$$E[X_k Y_k] = E[X_k] E[Y_k]$$

3. Como  $X_k \sim \mathcal{N}(0, 1)$  e  $Y_k \sim \mathcal{N}(0, 1)$ , suas médias são  $E[X_k] = 0$  e  $E[Y_k] = 0$ .

4. Substituindo, obtemos:

$$E[\langle X, Y \rangle] = \sum_{k=1}^m (0 \times 0) = \sum_{k=1}^m 0 = 0$$

Com isso, chegamos que esperança do produto interno entre dois vetores gaussianos padrão independentes é zero. Vamos agora calcular a variância:

- (a) **Expansão do Quadrado do Produto Interno:**

$$(\langle X, Y \rangle)^2 = \left(\sum_{k=1}^m X_k Y_k\right)^2 = \sum_{k=1}^m (X_k Y_k)^2 + \sum_{k \neq j} (X_k Y_k)(X_j Y_j)$$

- (b) **Esperança dos Termos Quadráticos:** Para o primeiro termo da soma expandida:

$$E[(X_k Y_k)^2] = E[X_k^2 Y_k^2]$$

Como  $X_k$  e  $Y_k$  são independentes:

$$E[X_k^2 Y_k^2] = E[X_k^2] E[Y_k^2]$$

Sabemos que para uma variável  $Z \sim \mathcal{N}(0, 1)$ ,  $E[Z^2] = \text{Var}[Z] + (E[Z])^2 = 1 + 0^2 = 1$ . Portanto,  $E[X_k^2] = 1$  e  $E[Y_k^2] = 1$ .

$$E[(X_k Y_k)^2] = 1 \times 1 = 1$$

- (c) **Esperança dos Termos Cruzados:** Para o segundo termo (onde  $k \neq j$ ):

$$E[(X_k Y_k)(X_j Y_j)] = E[X_k X_j Y_k Y_j]$$

Como todas as entradas são independentes (já que as colunas  $X$  e  $Y$  são independentes e as entradas dentro de cada vetor também são i.i.d.), podemos fatorar:

$$E[X_k X_j Y_k Y_j] = E[X_k] E[X_j] E[Y_k] E[Y_j]$$

Como  $E[X_k] = 0$ ,  $E[X_j] = 0$ ,  $E[Y_k] = 0$ ,  $E[Y_j] = 0$ :

$$E[X_k X_j Y_k Y_j] = 0 \times 0 \times 0 \times 0 = 0$$

(d) **Resultado da Variância:** Combinando os termos esperados:

$$Var[\langle X, Y \rangle] = \sum_{k=1}^m E[(X_k Y_k)^2] + \sum_{k \neq j} E[(X_k Y_k)(X_j Y_j)]$$

$$Var[\langle X, Y \rangle] = \sum_{k=1}^m 1 + \sum_{k \neq j} 0 = m$$

*Conclusão:* A variância do produto interno entre dois vetores gaussianos padrão independentes, cada um de dimensão  $m$ , é  $m$ . Consequentemente, o desvio padrão é  $\sqrt{m}$ .

□

## Referências

- [1] Rinaldo Artes. *Coeficiente de Assimetria*. Notas de Aula. PDF disponibilizado pelo autor. Insper, 2014. URL: [https://www.ime.usp.br/~mbranco/MedidasdeAssimetria\\_2014.pdf](https://www.ime.usp.br/~mbranco/MedidasdeAssimetria_2014.pdf).
- [2] H. Pishro-Nik. *Introduction to Probability, Statistics, and Random Processes – Section 8.2.2: Point Estimators for Mean and Variance*. Webpage. Disponível em [https://www.probabilitycourse.com/chapter8/8\\_2\\_2\\_point\\_estimators\\_for\\_mean\\_and\\_var.php](https://www.probabilitycourse.com/chapter8/8_2_2_point_estimators_for_mean_and_var.php). Acesso em 20 de junho de 2025. Kappa Research LLC, 2014. URL: [https://www.probabilitycourse.com/chapter8/8\\_2\\_2\\_point\\_estimators\\_for\\_mean\\_and\\_var.php](https://www.probabilitycourse.com/chapter8/8_2_2_point_estimators_for_mean_and_var.php).
- [3] Gilbert Strang. *Introduction to Linear Algebra*. 5th. Wellesley-Cambridge Press, 2016. ISBN: 978-0980232776.
- [4] Gilbert Strang. *Linear Algebra and Its Applications*. 4th. Cengage Learning, 2016. ISBN: 978-0030105678.
- [5] Lloyd N. Trefethen e David Bau III. *Numerical Linear Algebra*. Philadelphia, PA: SIAM, 1997. ISBN: 978-0898713619.