

# ANÁLISE ASSINTÓTICA DE SISTEMAS DE PARTÍCULAS E MATRIZES ALEATÓRIAS

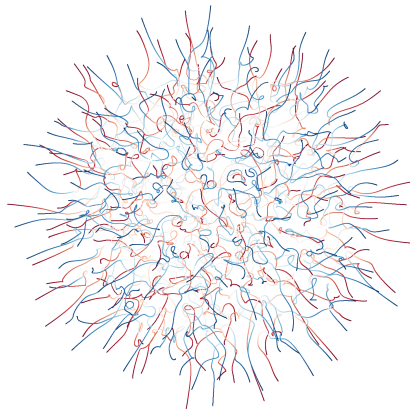
---

Relatório Científico Anual do projeto na modalidade Auxílio à Pesquisa Regular, fomentado pela Fundação de Amparo à Pesquisa do Estado de São Paulo.

---

Projeto FAPESP #2023/02674-0

Pesquisador Responsável: Guilherme L. F. Silva



São Carlos, 15 de outubro de 2023

# Informações Gerais do Projeto

- Título do projeto:

**Análise Assintótica de Sistemas de Partículas e Matrizes Aleatórias**

- Nome do pesquisador responsável:

**Guilherme L. F. Silva**

- Instituição sede do projeto:

**Instituto de Ciências Matemáticas e de Computação (ICMC) da  
Universidade de São Paulo**

- Equipe de pesquisa:

**Guilherme L. F. Silva**

**João Victor Alcantara Pimenta**

- Número do projeto de pesquisa:

**2023/02674-0**

- Período de vigência:

**01/06/2023 a 31/05/2024**

- Período coberto por este relatório científico:

**01/06/2023 a 10/11/2023**

## **Resumo**

O estudo de Matrizes Aleatórias demonstra aplicabilidade em uma gama diversa de áreas, com destaque no estudo de mecânica estatística, principalmente na simulação de gases. Estudando a densidade espectral de sistemas de matrizes Gaussianas pode-se desenvolver uma analogia que possibilita a simulação de sistemas de gases diversos, como o de Coulomb. Algumas dificuldades surgem na implementação de simulações baseadas nesta teoria, principalmente em escalabilidade do sistema e no tratamento de possíveis singularidades. Para resolver estes problemas, abordou-se na simulação na literatura, dentre outros, o Algoritmo Híbrido de Monte Carlo, de ótimo comportamento numérico. Nosso objetivo é explorar este assunto, as simulações de gases e o algoritmo citado acima além de expandir os potenciais em que foi-se bem documentado o comportamento destas simulações.

# 1 Introdução à Matrizes Aleatórias

Uma matriz aleatória é uma matriz cujas entradas são variáveis aleatórias, não necessariamente independentes tampouco de mesma distribuição. A princípio, de um ponto de vista puramente analítico, pode-se tratar uma matriz aleatória de tamanho  $N \times N$  como um vetor aleatório de tamanho  $N^2$ . No entanto, as estruturas algébrico-geométricas presentes a matrizes, como multiplicação natural, interpretação como operadores, ou decomposições espectrais, trazem à matrizes aleatórias aplicações múltiplas. Em particular, sua relevância estende um ponto comum que compartilham com variáveis aleatórias: permitir descrições estatísticas a sistemas e fenômenos.

É comum modelar com matrizes aleatórias, por exemplo, operadores com perturbações aleatórias. De um ponto de vista físico, autovalores de um dado operador descrevem espectros de energia do sistema descrito. Na quântica, por exemplo, autovalores são as medidas observadas. Surge uma pergunta naturalmente: dada uma matriz aleatória  $M$ , o que podemos dizer sobre estatísticas de seus autovalores? Essa resposta, claro, depende de maneira altamente não trivial das distribuições das entradas.

## 1.1 DISTRIBUIÇÃO DOS AUTOVALORES

Consideremos inicialmente um espaço de matrizes com  $N^2$  entradas independentes, sejam elas reais, complexas ou simpléticas. Se tivéssemos interesse de expressar a medida desse espaço poderíamos escrever

$$p(\hat{M})dM = p(M_{1,1}, \dots, M_{N,N}) \prod_{i,j=1}^N dM_{i,j} \quad (1.1)$$

Contido neste espaço temos um espaço de maior interesse correspondente ao espaço das matrizes *simétricas* ou *hermitianas*. A escolha do subespaço está relacionada com o fato de que essas matrizes são diagonalizáveis. Podemos escrever nossa matriz  $\hat{H}$  como

$$\hat{H} = \hat{U} \hat{\Lambda} \hat{U}^{-1}, \quad \hat{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_N), \quad \hat{U} \cdot \hat{U}^* = I$$

onde, claro,  $\hat{\Lambda}$  é matriz diagonal e  $\hat{U}$  é matriz unitária. Em geral, o conjunto de matrizes degeneradas tem medida nula e não é uma preocupação. Um cuidado deve ser tomado. A correspondência  $\hat{H} \implies (\hat{U} \in U(N), \hat{\Lambda})$  não é injetora, podemos tomar  $\hat{U}_1 \hat{\Lambda} \hat{U}_1^{-1} = \hat{U}_2 \hat{\Lambda} \hat{U}_2^{-1}$  se  $\hat{U}_1^{-1} \hat{U}_2 = \text{diag}(e^{i\phi_1}, \dots, e^{i\phi_N})$  para qualquer escolha de fases  $(\phi_1, \dots, \phi_N)$ . Para restringir nosso problema e tornar a função injetiva será necessário considerar as matrizes unitárias ao espaço de coset  $U(N)/U(1) \times \dots \times U(1)$ . Outra restrição necessária é ordenar os autovalores, ou seja,  $\lambda_1 < \dots < \lambda_n$ , isso deverá introduzir uma constante de normalização  $N!$  à expressão. Podemos assim reescrever a medida  $d\mu(\hat{H})$  em função dos autovalores. Nesse subespaço escreveríamos

$$p(M_{1,1}, \dots, M_{N,N}) \prod_{i < j} dM_{i,j} = p(\lambda_1, \dots, \lambda_N, \hat{U}) dU \prod_{i=1}^N d\lambda_i$$

Onde, é claro, sendo  $J(\hat{M} \rightarrow \lambda_i, U)$  o jacobiano da transformação

$$p(M_{1,1}, \dots, M_{N,N}) J(\hat{M} \rightarrow \lambda_i, U) = p(\lambda_1, \dots, \lambda_N, \hat{U})$$

Neste caso, podemos expressar o jacobiano como um determinante de Vandermonde

$$J(\hat{M} \rightarrow \{\lambda_i, U\}) = \prod_{j > k} (\lambda_j - \lambda_k)^\beta \quad (1.2)$$

Onde  $\beta > 0$  e depende das entradas da matriz. Se quisermos expressar a medida somente em termos dos autovalores poderíamos integrar no espaço dos autovetores. Isso nem sempre é simples ou possível. Por simplicidade temos ocultado a dependência das entradas que deveriam ser expressas  $M_{i,j}(\lambda, U)$ . Para efeitos deste trabalho tomaremos *ensembles* ortogonalmente invariantes, ou seja, tais que  $M_{i,j}(\lambda)$ . Basta então, definir o volume do espaço dos autovetores que nos dará uma constante na expressão.

$$p(\hat{M}) dM = \frac{1}{Z_N} p(\lambda_1, \dots, \lambda_N) \prod_{j > k} (\lambda_j - \lambda_k)^\beta \quad (1.3)$$

Notemos um ponto importante. Ao restringir o espaço das matrizes para o espaço das hermitianas introduzimos o determinante de Vandermonde. Este desempenha importante papel na caracterização da medida, note que agora, realizações com autovalores próximos são improváveis. Isso se expressa como uma repulsão de autovalores distintos quando introduzimos uma dinâmica.

## 1.2 ENSEMBLES GAUSSIANOS

Especial interesse é disposto em uma classe específica de matrizes aleatórias, denominada *Gaussian Ensemble* (abreviadamente, em qualquer de suas três formas tradicionais: *Gaussian Orthogonal Ensemble* (abreviadamente GOE), *Gaussian Unitary Ensemble* (GUE) e *Gaussian Symplectic Ensemble* (GSE)). As três formas se distinguem essencialmente pelo tipo de matrizes consideradas, a saber simétricas, unitárias ou unitárias auto-duais, com seus grupos de simetria associados, matrizes ortogonais, unitárias ou unitárias-simpléticas, respectivamente. Estes tipos de matrizes são especialmente interessantes pois possuem uma propriedade única: a distribuição conjunta das entradas é invariante pela ação do seu grupo de simetria e, simultaneamente, possuem entradas independentes, neste caso Gaussianas no corpo real, complexo ou quaterniônico para o GOE, GUE e GSE, respectivamente. Para seus autovalores autovalores, a distribuição assume a forma

$$p(\lambda_1, \dots, \lambda_N) \prod_i d\lambda_i = \frac{1}{Z_N} e^{-\beta \sum_k \lambda_k^2} \prod_{j < k} |\lambda_k - \lambda_j|^\beta \prod_i d\lambda_i, \quad \lambda_1, \dots, \lambda_N \in R,$$

onde  $\beta$  tem relação com o tipo de matriz Gaussiana utilizada, tendo valor 1 para o GOE, 2 para o GUE e 4 para o GSE,  $Z_N$  é a constante de normalização, também chamada de função de partição, e  $d\lambda_j$  é a medida de Lebesgue unidimensional. Essa expressão pode ser reescrita de forma mais interessante como uma medida de Gibbs, a saber

$$P(\lambda_1, \dots, \lambda_N) = \frac{1}{Z_N} e^{-\beta H(\lambda_1, \dots, \lambda_N)}, \quad H(\lambda_1, \dots, \lambda_N) := \sum_{j < k} \log \frac{1}{|\lambda_k - \lambda_j|^\beta} + \sum_k \lambda_k^2. \quad (1.4)$$

Desta forma, fator  $\beta$  pode - e deve - ser interpretado como temperatura inversa. Nesta representação, ele é o peso de Boltzmann e nossa distribuição é uma analogia à distribuição do sistema canônico da mecânica estatística. Nesta analogia,  $H$  seria o Hamiltoniano do sistema que determina o potencial, energia cinética e da interação entre partículas.

A medida de Gibbs (1.4) admite uma extensão natural

$$P_V(\lambda_1, \dots, \lambda_N) := \frac{1}{Z_N^V} e^{-\beta H_V(\lambda_1, \dots, \lambda_N)}, \quad H_V(\lambda_1, \dots, \lambda_N) := \sum_{j < k} \log \frac{1}{|\lambda_k - \lambda_j|^\beta} + \sum_k V(\lambda_k), \quad (1.5)$$

onde  $V : \mathbb{R} \rightarrow \mathbb{R}$  é uma função suficientemente regular. A escolha  $V(x) = x^2$ , obviamente, recupera (1.4). A distribuição (1.5) também descreve autovalores de modelos de matrizes aleatórias apropriados, mas agora as entradas já não são mais independentes.

### 1.3 A MEDIDA DE EQUILÍBRIO

Assim como na física, para resolver nossa distribuição, vamos precisar minimizar a energia livre do nosso ensemble. De alguma forma recuperamos a noção da energia livre de Helmholtz,

$$F = -\frac{1}{k_b} \ln(Z_N)$$

Teremos que os estados mais prováveis serão aqueles em que for maximizada a expressão

$$\exp \left[ -N^2 \left( \frac{1}{N^2} \sum_{i \neq j} \log \frac{1}{|\lambda_i - \lambda_j|} + \frac{1}{N^2} \sum_{i=1}^N \tilde{V}(\lambda_i) \right) \right] \quad (1.6)$$

Onde identificamos o Hamiltoniano e escrevemos  $\exp(-N^2 \tilde{\mathcal{H}}_N(\lambda))$ . Precisamos minimizar o Hamiltoniano do sistema  $\tilde{\mathcal{H}}_N(\lambda)$ . Vamos introduzir uma função contagem para facilitar o tratamento do conjunto de pontos em  $\mathbb{R}$ . Definimos

$$v_\lambda = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i} \quad (1.7)$$

De forma que

$$\mathcal{H}_N(v_\lambda) = \int \int_{x \neq y} \log \frac{1}{|\lambda_i - \lambda_j|} v_\lambda(x) v_\lambda(y) dx dy + \frac{1}{N} \int \tilde{V}(x) v_\lambda(x) dx \quad (1.8)$$

O que acontece quando tratamos do limite termodinâmico? Ou seja, quando  $N \rightarrow \infty$ . Estaremos transicionando da nossa função  $v_\lambda$  para uma densidade  $\mu_V(x)dx$ , ou seja,

$$\int f(x) v_\lambda(x) dx = \int f(x) \mu_V(x) dx$$

Precisamos garantir ainda um potencial  $V(x) = N\tilde{V}(x)$  para trabalharmos a assintótica e mantermos a integrabilidade. Escreve-se

$$\frac{1}{\mathcal{Z}_N} \prod_{i < j} (\lambda_i - \lambda_j)^\beta \prod_{i=1}^N e^{-NV(\lambda_i)} d\lambda = \frac{1}{\mathcal{Z}_N} e^{-N^2 \mathcal{H}_N(\lambda)}$$

E finalmente poderemos expressar de forma correta

$$\mathcal{H}_N(v_\lambda) \rightarrow \int \int_{x \neq y} \log \frac{1}{|\lambda_i - \lambda_j|} d\mu_V(dx) d\mu_V(dy) + \int V(x) \mu_V(dx) \equiv \epsilon^V(\mu_V) \quad (1.9)$$

Onde  $\mu_V(x)$  será medida de probabilidade não aleatória tal que na assintótica,

$$\mu_V^* = \arg \inf \epsilon$$

sendo o minimizante único da função 'energia'  $\epsilon$  convexa e semi continua.

## 1.4 EXEMPLOS

### 1.4.1 Potencial Gaussiano

Em geral, para os ensembles gaussianos estaremos interessados no potencial quadrado, salvo uma escala

$$V(x) = x^2$$

Para estes ensembles vale o clássico resultado da medida de equilíbrio dada pela Lei do Semi-Círculo de Wigner. Se expressa,

$$\text{supp } \mu_V = [-\sqrt{2}, \sqrt{2}], \frac{d\mu_V}{dx}(x) = \frac{1}{\pi} \sqrt{2 - x^2}$$

e teremos das simulações para os três ensembles e suas distribuições esperadas em 1.1.

Podemos considerar agora exemplos de potenciais que podem ser aplicados. Lembre que isso implica que nossa entradas são correlacionadas, os ensembles de matrizes para este caso não são tão diretos quanto para entradas independentes. Consideraremos para os exemplos  $\beta = 2$ . As explicações para os resultados gráficos são retomados na seção sobre as simulações implementadas.

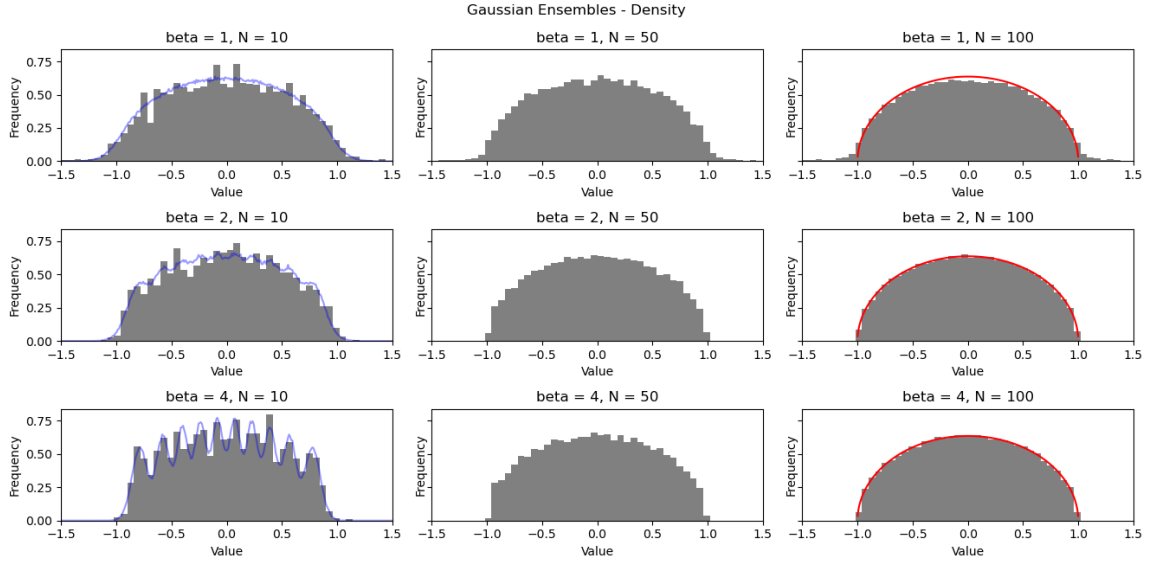


Figura 1.1: Validação para ensembles clássicos, utilizamos 200000 passos registrando a cada 500 a partir da metade dos passos.  $\Delta t = 0.1$ ,  $\gamma = 1$ ,  $\alpha = 1.0$ . Para replicar a semente foi 987991650.

### 1.4.2 Potencial Mônico

Considere o potencial

$$V(x) = \frac{t}{2\alpha} x^{2\alpha}$$

Onde  $t > 0$  é escala e  $\alpha \in \mathbb{Z}$ . A medida de equilíbrio para  $\alpha = 1$  é o semi-círculo de podemos validar na figura com a distribuição em vermelho. De qualquer forma, podemos observar os resultados das simulações para este potencial e notar o comportamento esperado e a distribuição teórica em vermelho para o semicírculo 1.2.

### 1.4.3 Potencial Quártico

Considere o potencial

$$V(x) = \frac{x^4}{4} + t \frac{x^2}{2}$$

Aqui observaremos pela primeira vez uma transição de estado. Teremos um ponto crítico em  $t = -2$  onde a medida se para em dois intervalos  $[-b_t, -a_t]$  e  $[a_t, b_t]$ . Ou seja:

- $t > -2$

$$\text{supp } \mu_V = [-b_t, b_t], \frac{d\mu_V}{dx}(x) = \frac{1}{2\pi}(x^2 - c_t^2)\sqrt{b_t^2 - x^2}$$

com



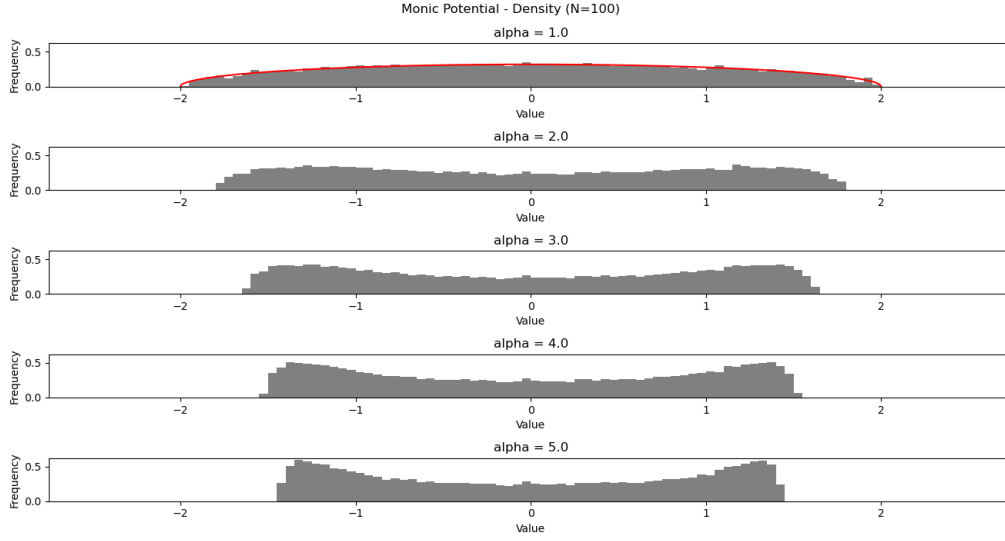


Figura 1.2: Validação para potencial mônico,  $V(x) = \frac{t}{2\alpha}x^{2\alpha}$ . Utilizamos 1000000 passos registrando a cada 500 a partir da metade dos passos.  $t = 1$ ,  $\Delta t = 0.1$ ,  $\gamma = 10$ ,  $\alpha = 0.1$ . Para replicar a semente foi 987991650.

$$c_t^2 := \frac{1}{2}b_t^2 + t := \frac{1}{3}(-2t + 2\sqrt{t^2 + 12})$$

- $t < -2$

$$\text{supp } \mu_V = [-b_t, -a_t] \cup [a_t, b_t], \quad \frac{d\mu_V}{dx}(x) = \frac{1}{2\pi}|x|\sqrt{(x^2 - a_t^2)(b_t^2 - x^2)}$$

com

$$a_t := \sqrt{-2 - t}, b_t := \sqrt{2 - t}$$

Observamos o comportamento esperado e a distribuição teórica em vermelho em 1.3.

## 1.5 SIMULAÇÕES

Para alguns modelos de Gases, sabemos que é conveniente usar modelos de matrizes aleatórias quando estas estiverem disponíveis. Processos determinantis também podem ser de ajuda quando  $\beta = 2$ . Fora esses casos, existem ainda métodos alternativos como o *Overdamped Langevin Difusion Algorithm*, *Metropolis-Hastings algorithm*, *Metropolis adjusted Langevin algorithm* e versões cinéticas chamadas *Hybrid or Hamiltonian Monte Carlo* baseada em uma versão cinética (*underdamped*) da difusão de Langevin.

Em geral amostrar a medida resulta em dificuldades. A computação de forças e de energias escala com  $N^2$  pelo Hamiltoniano tratar de interações par a par. Outra dificuldade são as singularidades em  $W$ , que resultam em instabilidades numéricas.

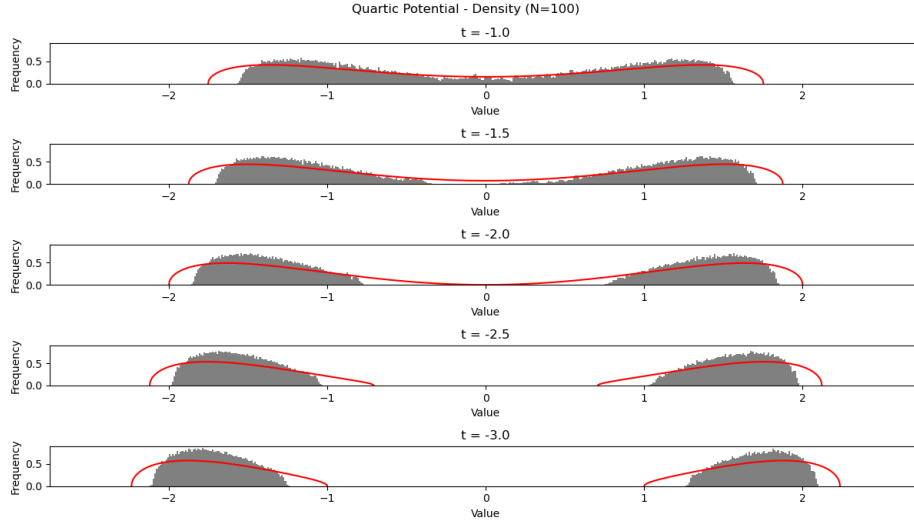


Figura 1.3: Validação para potencial quártico,  $V(x) = \frac{1}{4}x^4 + \frac{1}{2}x^2$ . Utilizamos 1000000 passos registrando a cada 500 a partir da metade dos passos.  $\Delta t = 0.1$ ,  $\gamma = 10$ ,  $\alpha = 0.1$ . Para replicar a semente usada é 987991650.

### 1.5.1 Os típicos

A ideia explorada é que  $P_N$  é medida de probabilidade invariante reversível do processo de difusão de Markov  $(X_t)_{t>0}$  solução de

$$dX_t = -\alpha_N \nabla H_N(X_t) dt + \sqrt{2 \frac{\alpha_N}{\beta_N}} dB_t$$

Sob alguma condição em  $\beta_N$  e  $V$  podemos afirmar que

$$X_t \xrightarrow[t \rightarrow \infty]{Law} P_N$$

discretizado, tomaríamos o processo

$$x_{k+1} = x_k - \nabla H_N(x_k) \alpha_N \Delta t + \sqrt{2 \frac{\alpha_N}{\beta_N}} \Delta t G_k$$

onde  $G_k$  é a família de variáveis gaussianas usuais. Uma forma de contornar o viés embutido é amenizar a dinâmica com a forma

$$x_{k+1} = x_k - \frac{\nabla H_N(x_k) \alpha_N \Delta t}{1 + |\nabla H_N(x_k)| \alpha_N \Delta t} + \sqrt{2 \frac{\alpha_N}{\beta_N}} \Delta t G_k$$

Ainda assim, precisamos otimizar o processo. A ideia do algoritmo de Metropolis é adicionar um processo de seleção para evitar passos irrelevantes, algo do tipo:

- defina  $\tilde{x}_{k+1}$  de acordo com o kernel  $K(x_k, \cdot)$  gaussiano;

- defina  $p_k$

$$p_k = 1 \wedge \frac{K(\tilde{x}_{k+1}, x_k) e^{\beta_N H_N(\tilde{x}_{k+1})}}{K(x_k, \tilde{x}_{k+1}) e^{\beta_N H_N(\tilde{x}_k)}}$$

- defina

$$x_{k+1} = \begin{cases} \tilde{x}_{k+1} & w/p_k \\ x_k & w/1 - p_k \end{cases}$$

### 1.5.2 O Híbrido de Monte Carlo

O algoritmo híbrido de Monte Carlo é baseado o algoritmo anterior mas adicionando uma variável de momento para melhor explorar o espaço. Defina  $E = \mathbb{R}^{dN}$  e deixe  $U_N : E \rightarrow \mathbb{R}$  ser suave para que  $e^{-\beta_N U_N}$  seja Lebesgue integrável. Seja ainda  $(X_t, Y_t)_{t>0}$  ser o processo de difusão em  $E \times E$  solução de

$$\begin{cases} dX_t = \alpha_N \nabla U_N(Y_t) dt \\ dY_t = \alpha_N \nabla H_N(X_t) dt - \gamma_N \alpha_N \nabla U_N(Y_t) dt + \sqrt{2 \frac{\gamma_N \alpha_N}{\beta_N}} dB_t \end{cases}$$

Onde  $(B_t)_{t>0}$  é o movimento browniano em  $E$  e  $\gamma_N > 0$  parâmetro representando atrito.

Quando  $U_N(y) = \frac{1}{2}|y|^2$  temos  $Y_t = dX_t/dt$  e teremos que  $X_t$  e  $Y_t$  poderão ser interpretados como posição e velocidade do sistema de  $N$  pontos em  $S$  no tempo  $t$ . Nesse caso,  $U_n$  é energia cinética.

#### 1.5.2.1 O algoritmo discreto

Descrevemos agora o algoritmo discretizado. Inicie de uma configuração  $(x_0, y_0)$  e para todo  $k \geq 0$  faça

1. atualize as velocidades com

$$\tilde{y}_k = \eta y_k + \sqrt{\frac{1 - \eta^2}{\beta_N}} G_k, \quad \eta = e^{-\gamma_N \alpha_N \Delta t}$$

2. calcule os termos

$$\begin{cases} \tilde{y}_{k+\frac{1}{2}} = \tilde{y}_k - \nabla H_N(x_k) \alpha_N \frac{\Delta t}{2} \\ \tilde{x}_{k+1} = x_k + \tilde{y}_{k+\frac{1}{2}} \alpha_N \Delta t \\ \tilde{y}_{k+1} = \tilde{y}_{k+\frac{1}{2}} - \nabla H_N(x_{k+1}) \alpha_N \frac{\Delta t}{2} \end{cases}$$

3. definir  $p_k$

$$p_k = 1 \wedge \exp \left[ -\beta_N \left( H_N(\tilde{x}_{k+1}) + \frac{\tilde{y}_{k+1}^2}{2} - H_N(x_k) - \frac{\tilde{y}_k^2}{2} \right) \right]$$

4. defina

$$(x_{k+1}, y_{k+1}) = \begin{cases} (\tilde{x}_{k+1}, \tilde{y}_{k+1}) & w / p_k \\ (x_k, -\tilde{y}_k) & w / 1 - p_k \end{cases}$$

### 1.5.3 Validação

Já apresentamos ao longo da apresentação alguns dos resultados das simulações nas figuras 1.1, 1.2 e 1.3. Para 1.1 temos concordância do comportamento teórico assintótico (semi-círculo). Note que as simulações estão normalizadas para que o suporte seja  $[-1, 1]$ . Em especial plotamos também em azul para  $N = 10$ , resultados da simulação equivalente utilizando matrizes aleatórias dos ensembles adequados. De novo, temos concordância muito boa.

Para 1.2, podemos validar qualitativamente a forma do potencial e o efeito que causa na distribuição. O movimento da densidade para as bordas é justificado nesse sentido.

Finalmente, em 1.3, temos uma comparação direta do que se espera da distribuição e do resultado da simulação. Apesar do resultado parecer discordante, alguma ponderação pode esclarecer as coisas. Para a quantidade de partículas consideradas, a densidade deve, claro, ser diferente da assintótica. Por exemplo, encontrar partículas em torno de  $x = 0$  é razoavelmente raro para os casos próximos da transição. De forma que a quantidade de passos e partículas influencia fortemente a observação de realizações na região. AS bordas externas também podem ser explicadas pela quantidade limitada de partículas, uma vez que elas se repelem.

## 2 Realizações do período

### 2.1 GRADUAÇÃO

Durante o período referente ao relatório o aluno completou as matérias do primeiro semestre e iniciou as matérias do segundo semestre listadas na tabela abaixo.

Disciplina	Sigla	Nota	Semestre
Mecânica Estatística Avançada	7600041	10.0	1 - 2023
Introdução aos Sistemas de Computação	7600056	9.2	1 - 2023
Física Estatística Computacional	7600073	9.7	1 - 2023
Teoria Espectral de Matrizes	SME0243	10.0	1 - 2023
Mecânica Quântica	7600022	-	2 - 2023
Física Matemática Avançada	7600034	-	2 - 2023
Noções Básicas de Fabricação Mecânica	7600134	-	2 - 2023
Espaços Métricos	SMA0343	-	2 - 2023
Trabalho de Conclusão de Curso	7600039	-	2 - 2023

## 2.2 PESQUISA

Durando os meses passados no período referente a esse relatório grande parte do esforço foi no estudo da bibliografia e conteúdo de interesse dentro da teoria de matrizes aleatórias. Para isso, permitiu-se exploração ampla de conceitos relacionados e implementação de algoritmos especiais em interesse ao aluno. Algumas das atividades mais visuais podem ser visualizadas na figura e todos os resultados podem ser encontrados em GitHub - Repositório Geral. A outra atividade principal realizada foi a implementação do algoritmo descrito em [1], onde podemos simular algumas medidas de probabilidade relacionadas à ensembles clássicos para validar seu funcionamento. Aqui, como é possível ver em , simulamos a distribuição para a GOE, GUE e GSE. Podemos ver nas imagens a concordância das simulações com o método clássico utilizando autovalores de ensembles de matrizes aleatórias.

## 3 Plano de atividades

O projeto tem como base um planejamento de 12 meses, que se deram início em junho de 2023, até o momento foram realizados 5 meses de projeto.

### 3.1 ATIVIDADES DESENVOLVIDAS

A execução do projeto foi dividida nas seguintes etapas:

1. **Revisão da Literatura em RMT, e estudo de teoria básica do GUE**, é necessário fazer vasta revisão de literatura no tema para que o aluno tenha domínio das ferramentas e métodos utilizados para o tratamento de matrizes aleatórias e suas implicações em mecânica estatística. Para isso, durante esse período será realizado o estudo da bibliografia adequada;
2. **Estudo dos métodos de Simulação**, como mencionado, uma das aplicações importantes da teoria de matrizes aleatórias reside em sua conexão com gases de

Coulomb. Em 2018 publicou-se o [1], artigo que é base para o estudo de métodos de simulação desses gases;

3. **Implementação dos algoritmos**, implementa-se os métodos descritos no artigo e tenta-se estender seu uso em condições diferentes das utilizadas no artigo, como por exemplo em outros potenciais;
4. **Redação dos Relatórios Científicos**, quando serão escritos os relatórios exigidos pelas normas da *FAPESP*.

## 3.2 CRONOGRAMA

Com base nas tarefas enumeradas na Seção 3.1, é mostrado na Tabela 3.1 o cronograma atual de desenvolvimento do projeto. Em especial, os métodos de simulação puderam ser adiantados no desenvolvimento para o mês 4, previamente previsto para o mês 5 e consequentemente as implementações também puderam ser iniciadas.

Fases	Meses											
	1	2	3	4	5	6	7	8	9	10	11	12
<b>1. Revisão Literatura RMT</b>	✓	✓	✓	✓	✓							
<b>2. Métodos de Simulação</b>				✓	✓	x	x	x				
<b>3. Implementação algoritmos</b>					✓	x	x	x	x	x	x	
<b>4. Redação Relatórios</b>				✓	✓						x	x

Tabela 3.1: Cronograma das atividades.

## 4 Participação em eventos científicos

O bolsista apresentou em dois eventos no período em que se refere o presente relatório. O Colóquio Brasileiro de Matemática (CBM) e a Semana Integrada da Física de São Carlos (SIFSC). Apenas para o primeiro, realizado no Rio de Janeiro, foi necessário o uso da reserva técnica. Por isso, segue o pôster apresentado na página que se segue. O trabalho é complementar aos estudos assintóticos e de probabilidade realizados nos meses cobertos por este relatório.

Evento	Sede	Data	Modalidade	Apresentação	Reserva Técnica
CBM	IMPA	24-28/07/23	Presencial	Pôster - Oral	Sim
SIFSC	IFSC-USP	21-25/08/23	Presencial	Pôster - Oral	Não

O trabalho apresentado no CBM foi apresentado oralmente por meio de pôsteres no evento científico Colóquio Brasileiro de Matemática ocorrido de 24-28 de Julho no IMPA, Rio de Janeiro. Foram utilizadas duas diárias da reserva para a participação do colóquio.

# Dados Intransitivos

Joao V. A. Pimenta & Joao P. C. de Paula & Lael V. Lima & Luis G. C. Bueno

Guilherme L. F. Silva & Daniel Ungaretti & Tertuliano Franco

Universidade De Sao Paulo

joaovictor@usp.br



## Resumo

Intransitividade em um jogo de dados é um conceito muitas vezes pouco intuitivo. Afinal, se existem dados  $A, B, C$  tais que  $A \triangleright B$  e  $B \triangleright C$ , é pouco natural imaginar que seja possível  $C \triangleright A$ . Fato é, escolhendo devidamente seus dados, esse fenômeno é possível e alguns resultados seguem desde que um bom modelo seja definido. No primeiro momento, exploraremos uma representação dos dados como *palavras* que nos possibilita definir a existência de conjuntos intransitivos para todas configurações de número de dados ( $m$ ) e faces ( $n$ ). Em um segundo modelo, no qual as faces dos dados são dadas por uma distribuição uniforme  $[0, 1]$ , note, sem repetição e por isso, ligado as palavras. Por fim, exploramos um pouco a razão de conjuntos de dados ordenados intransitivos ( $\mathcal{I}$ ) sobre o conjunto total de dados possível ( $\mathcal{D}$ ). Analisando a assintótica em relação a  $n$  simulamos o modelo e conjecturamos, levados por resultados computacionais e uma análise algébrica que  $L = 3 \log 3$  em  $|\mathcal{I}_n| = e^{nL(1+o(1))}$ .

## Palavras

Construiremos um modelo sem repetição baseado em palavras onde cada letra é associada a um dado. Na tabela representamos  $abccbbcaabcb$ .

A	12				8			4	3		
B		11				7	6			2	
C			10	9				5			1

$A \triangleright B$  se, na *palavra*, a contagem dos  $b$ 's na direita de todos  $a$ 's for menor que a recíproca. Seja  $S_m^n$  uma palavra de  $m$  tipos de letras com  $n$  repetições, notamos:

**Lema 1.** Se existe uma palavra  $S_{m,n}$  intransitiva, então existe uma palavra  $S_{m+1,n}$  que também é intransitiva.

**Lema 2.** Se existe  $S_{m,n}$  intransitiva, então existe uma palavra  $S_{m,n+2}$  que também é intransitiva.

**Lema 3.** Uma palavra  $S_{3,2}$  não pode ser intransitiva.

**Teorema 4.** Para todo  $n \geq 3$  e  $m \geq 3$  existem palavras intransitivas de característica  $m$  e ordem  $n$ .

5	9	4	8	6	7
1		3			2
6		9			8
5	12	4	11	7	10
3			2		1

É possível mostrar para um modelo enviesado de dados que para todo  $m \geq 3$  e  $n \geq 4$ , existem  $m$  dados enviesados de  $n$  faces que são intransitivos.

## Dados Aleatórios

$A$  e  $B$  possuem  $n$  faces ( $A_1, A_2, \dots, A_n$ ), variáveis aleatórias i.i.d assumindo valores em  $[0, 1]$  com distribuição uniforme. O número de vitórias de  $A$  sobre  $B$  é dado por

$$N_{A>B} = \sum_{i=1}^n \sum_{j=1}^n \chi_{A_i > B_j}$$

e podemos explorar a probabilidade como a razão relacionado com as triplas de dados com  $n$  faces ( $A_n, B_n, C_n$ )

$$\mathbb{P}((A_n, B_n, C_n) \in \mathcal{I}_n) = \frac{|\mathcal{I}_n|}{|\mathcal{D}_n|},$$

$\mathcal{I}_n$  é o conjunto de trios ordenados de dados intransitivos e  $\mathcal{D}_n$  todos trios ordenados possíveis. Pela expansão de Stirling

$$|\mathcal{D}_n| = \frac{(3n)!}{(n!)^3} \sim \frac{\sqrt{3}}{2\pi n} 3^{3n}.$$

**Proposição 5.** Se  $r \in \mathcal{I}_n$  e  $s \in \mathcal{I}_m$ , então a concatenação  $rs \in \mathcal{I}_{n+m}$

**Corolário 6.** Sejam  $m$  e  $n$  inteiros positivos, então  $|\mathcal{I}_{m+n}| \geq |\mathcal{I}_m| |\mathcal{I}_n|$

**Teorema 7.**  $|\mathcal{I}_n| = e^{nL(1+o(1))}$ , para alguma constante  $L \in (2.5, 3 \log 3]$

**Prova.** Por ser subaditivo, pelo lema de Fekete, existe  $L$  tal que

$$\lim_{n \rightarrow \infty} \frac{\log |\mathcal{I}_n|}{n} = \sup_n \frac{\log |\mathcal{I}_n|}{n} = L.$$

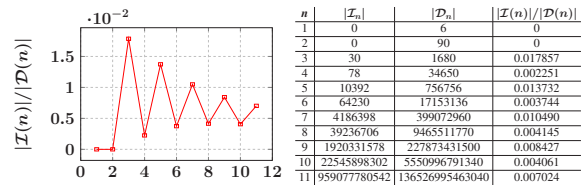
$$\implies |\mathcal{I}_n| = e^{nL(1+o(1))}$$

Podemos afirmar pelos resultados computacionais

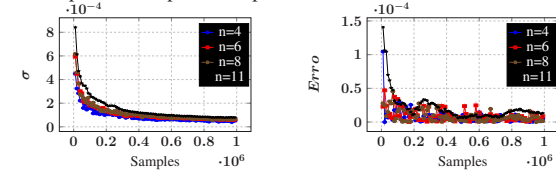
$$L \geq \frac{\log |\mathcal{I}_{11}|}{11} \approx 2.5.$$

E, como  $|\mathcal{I}_n| \leq |\mathcal{D}_n|$ ,

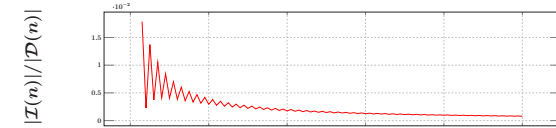
$$L = \lim_{n \rightarrow \infty} \frac{\log |\mathcal{I}_n|}{n} \leq \lim_{n \rightarrow \infty} \frac{\log |\mathcal{D}_n|}{n} = 3 \log 3.$$



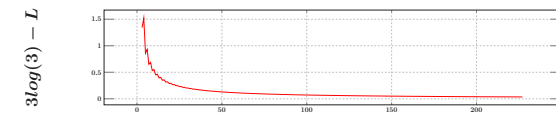
Validando para o espaço já conhecido de palavras podemos escolher os parâmetros de uma busca aleatória. Tomamos a média de 100 experimentos para cada ponto.



Estenderemos nossa busca realizando  $10^2$  experimentos para  $5 \cdot 10^5$  amostras. Sabemos que nosso erro é da ordem de  $\pm 5 \cdot 10^{-5}$ .



Para  $L$ ,



Até o momento, com  $n = 1000$ ,  $3 \log 3 - 0.01 \leq L \leq 3 \log 3$ .

**Conjectura 8.**  $L = 3 \log 3$ .

## Conclusão

Dados Intransitivos instigam a intuição e fornecem uma plataforma interessante de desenvolvimento matemático. Neste trabalho, demonstrou-se para os modelos considerados a existência de conjuntos de dados intransitivos para todas configurações para os modelos. A discussão da razão para o limite de número de faces culminou em uma exploração numérica e algébrica que culmina na Conjectura 8. Também mostramos posteriormente que  $|\mathcal{I}_n|/|\mathcal{D}_n| \rightarrow 0$ , e provamos Teorema do Limite Central para  $N_{a>b}$

## Referências

- [1] D. H. J. Polymath. The probability that a random triple of dice is transitive, 2022. arXiv:2211.16156.
- [2] Calyampudi Radhakrishna Rao. *Linear Statistical Inference and its Applications*. John Wiley & Sons, Inc, 2 edition, 1973.

## Agradecimentos

Pesquisa desenvolvida com utilização dos recursos computacionais do Centro de Ciências Matemáticas Aplicadas à Indústria (CE-MEAI) financiados pela FAPESP.

## Referências Bibliográficas

- [1] CHAFAÏ, D.; FERRÉ, G. Simulating coulomb and log-gases with hybrid monte carlo algorithms. *Journal of Statistical Physics*, Springer Science and Business Media LLC, v. 174, n. 3, p. 692–714, nov 2018. Disponível em: <<https://doi.org/10.1007>