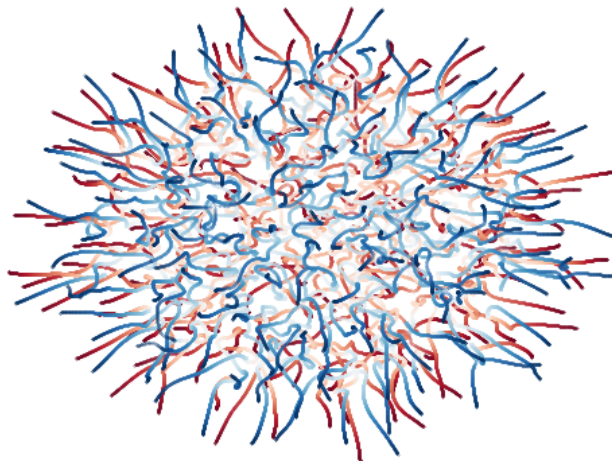


ANÁLISE ASSINTÓTICA DE SISTEMAS DE PARTÍCULAS E MATRIZES ALEATÓRIAS

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, 2 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 10/11/2023

- Período coberto por este relatório científico:

01/06/2023 a 31/05/2024

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 Resumo do projeto proposto

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.

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.

Em particular, a propriedade de invariância mencionada nos permite calcular a distribuição induzida nos autovalores. A diagonalização de uma matriz M dos Gaussian ensembles, a saber $M = U\Lambda U^*$, onde U é matriz ortogonal, unitária ou unitária-simplética respectivamente para o GOE, GUE e GSE, e onde $\text{diag}(\lambda_1, \dots, \lambda_n)$ é a matriz diagonal dos autovalores (reais), induz uma distribuição nos autovalores Λ e autovetores U . Um cálculo que remonta aos trabalhos de Dyson nos ensina que autovalores e autovetores são estatisticamente independentes e, ademais, suas distribuições induzidas são explícitas. No caso dos autovetores, a distribuição induzida é a uniforme, isto é, coincide com a medida normalizada de Haar no grupo. Já no caso dos autovalores, a distribuição assume a forma

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

onde β 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

$$P(\lambda_1, \dots, \lambda_N) = \frac{1}{Z_N} e^{-\beta H(\lambda_1, \dots, \lambda_N)}, \quad H(\lambda_1, \dots, \lambda_N) := \sum_{j < k} \log \frac{1}{|\lambda_k - \lambda_j|} + \sum_k \lambda_k^2. \quad (1.1)$$

Desta forma, fator β 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 (1.1) admite uma extensão natural

$$P_V(\lambda_1, \dots, \lambda_N) := \frac{1}{Z_N^V} e^{-\beta H_V(\lambda_1, \dots, \lambda_N)}, \quad H_V(\lambda_1, \dots, \lambda_N) := \sum_{j < k} \log \frac{1}{|\lambda_k - \lambda_j|} + \sum_k V(\lambda_k), \quad (1.2)$$

onde $V : \mathbb{R} \rightarrow \mathbb{R}$ é uma função suficientemente regular. A escolha $V(x) = x^2$, obviamente, recupera (1.1). A distribuição (1.2) também descreve autovalores de modelos de matrizes aleatórias apropriados, mas agora as entradas já não são mais independentes. Alternativamente, (1.2) é interpretado como um gás de Coulomb sobre a reta real, confinado pelo potencial V e com interações singulares entre partículas, descritas pelo núcleo logarítmico $\log|x - y|^{-1}$ que gera efeito de repulsão interna. Essa distribuição P_V é o objeto central deste projeto.

2 Realizações do período

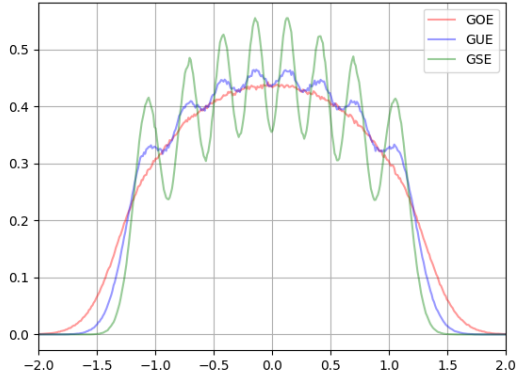
2.1 GRADUAÇÃO

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.

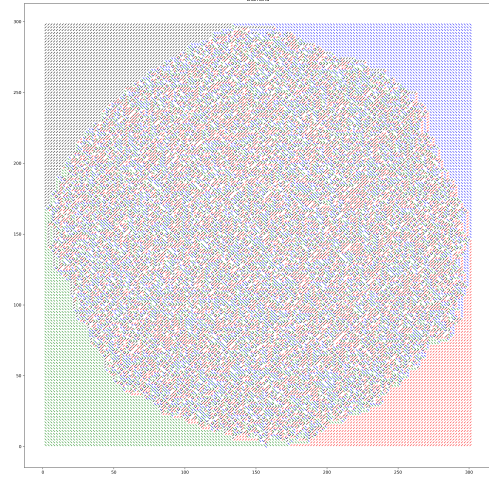
Disciplina	Sigla	Nota	Semestre
Mecânica Estatística Avançada	7600041	10.0	1 - 2023
Introdução aos Sistemas de Computação	7600056	9.2	1 - 2023
Física Estatística Computacional	7600073	9.7	1 - 2023
Teoria Espectral de Matrizes	SME0243	10.0	1 - 2023
Mecânica Quântica	7600022	-	2 - 2023
Física Matemática 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

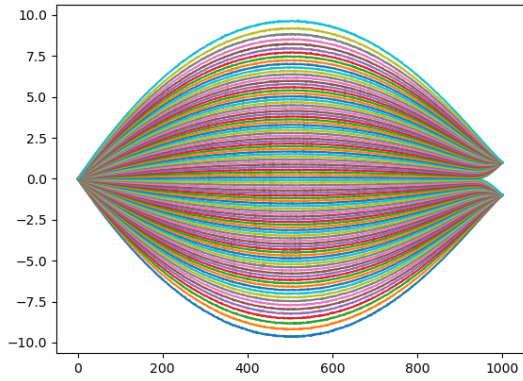
Durando os meses passados no período referente a esse relatório grande parte do esforço foi no estudo da bibliografia e conteúdo de interesse dentro da teoria de matrizes aleatórias. Para isso, permitiu-se exploração ampla de conceitos relacionados e implementação de algoritmos especiais em interesse ao aluno. Algumas das atividades mais visuais podem ser visualizadas na figura 2.1 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 2.2, 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.



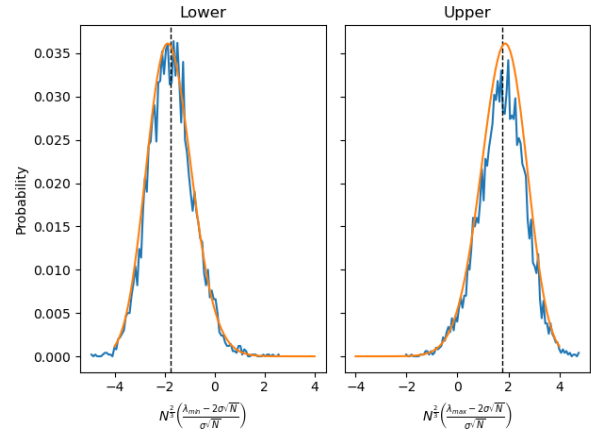
(a) GitHub - Gaussian Density



(b) GitHub - Tiling do Círculo Ártico

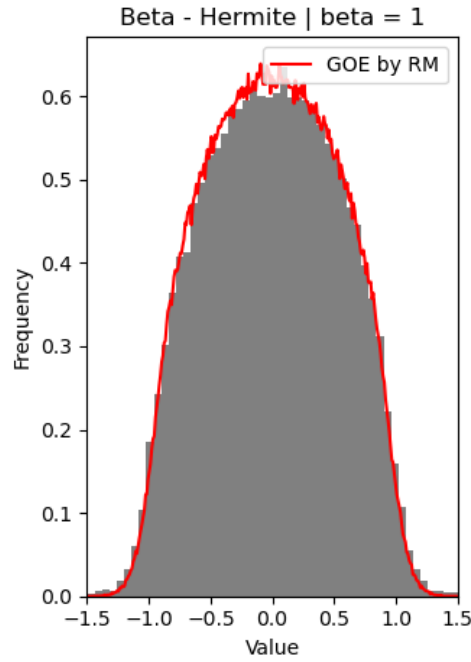


(c) GitHub - N Caminhos Brownianos

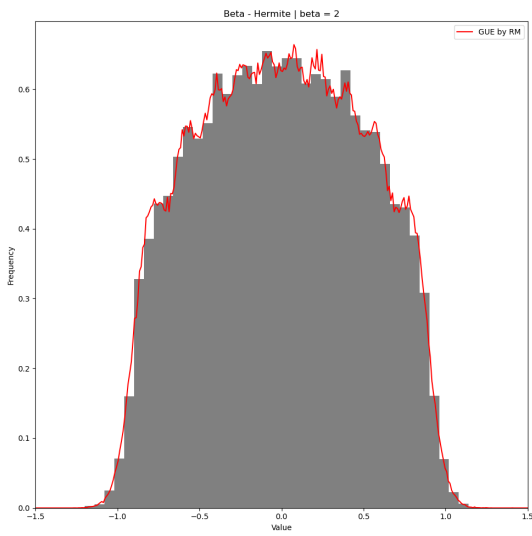


(d) GitHub - Determinação da Tracy Widow

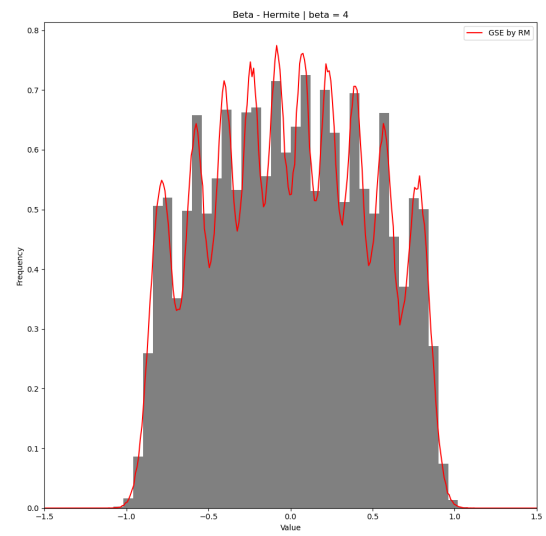
Figura 2.1: Figura 2.1a mostra a densidade de autovalores normalizada geradas pelos ensembles de matrizes GOE, GUE, GSE para N pequeno. Figura 2.1b é um possível tiling por dominós (peças 2×1) formando o famoso Diamante Asteca, o círculo formado é denominado Círculo Ártico e a distribuição da borda está relacionada com distribuição de autovalores máximos de ensembles especiais de matrizes. Figura 2.1c é a representação do caminho browniano de partículas com ponto inicial determinado e m pontos finais. A distribuição das partículas em cada momento é também a mesma da medida de probabilidade de autovalores do GUE. Figura 2.1d é a validação da distribuição de borda dos caminhos brownianos que deve ser a distribuição de Tracy Widow, demonstrada na figura.



(a)



(b)



(c)

Figura 2.2: As figuras representam a validação das distribuições geradas pelo algoritmo descrito no artigo [1]. Respectivamente, nas figuras 2.2a, 2.2b e 2.2c está descrito a medida obtida pela dinâmica dos pontos descrita pelo algoritmo no histograma. Em vermelho é a distribuição equivalente gerada pela mesma quantidade de observações usando modelos de matrizes aleatórios dos ensembles GOE, GUE e GSE. O programa usado está disponível em GitHub - Implementação do Algoritmo do Artigo

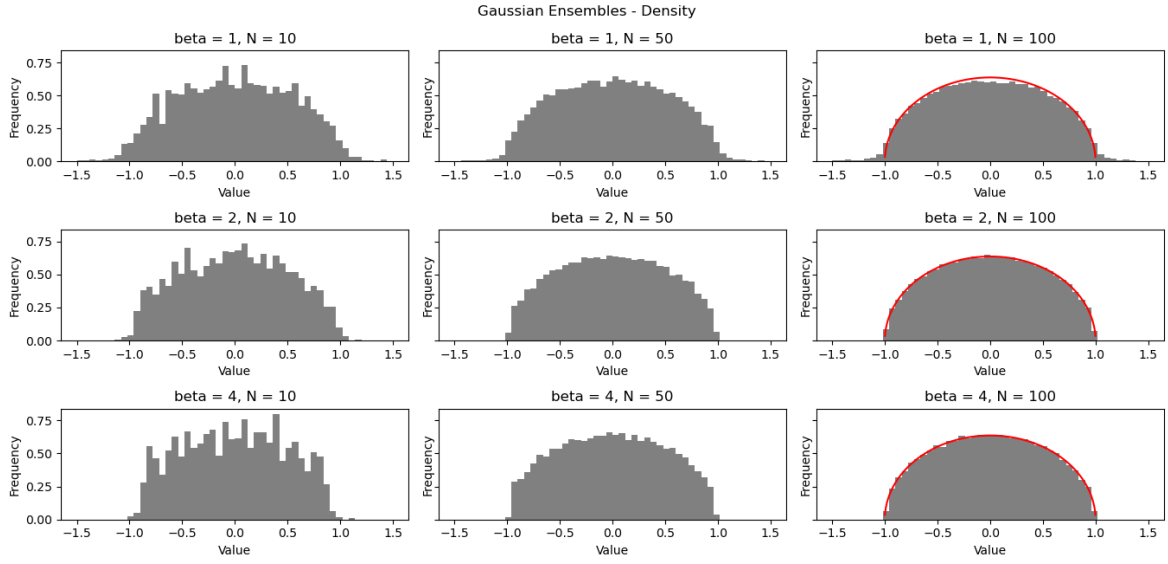


Figura 2.3: Validação para ensembles clássicos, utilizamos 200000 passos registrando a cada 500 a partir da metade dos passos. $\Delta t = 0.1$, $\gamma = 1$, $\alpha = 1.0$. Para replicar a semente foi 987991650.

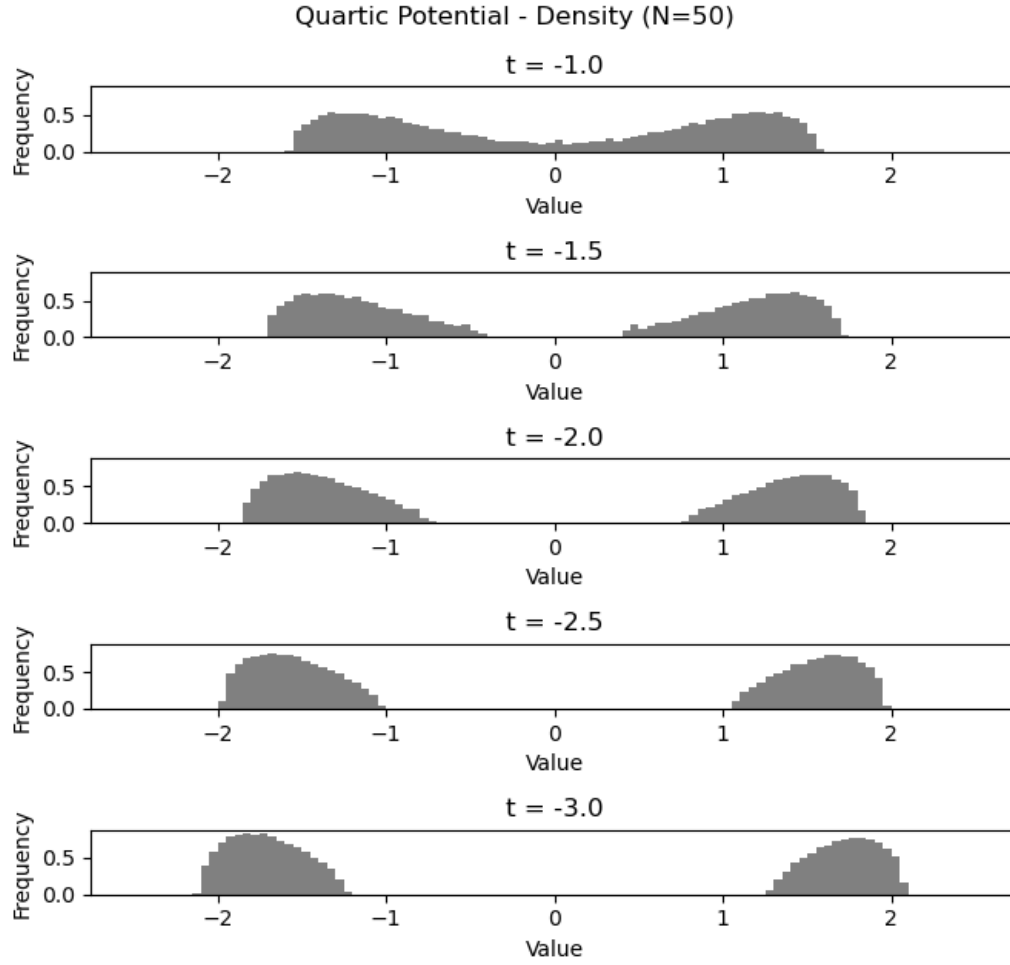


Figura 2.4: Validação para potencial quártico, $V(x) = \frac{1}{4}x^4 + \frac{1}{2}x^2$. Utilizamos 1000000 passos registrando a cada 500 a partir da metade dos passos. $\Delta t = 0.1$, $\gamma = 10$, $\alpha = 0.1$. Para replicar a semente foi 987991650.

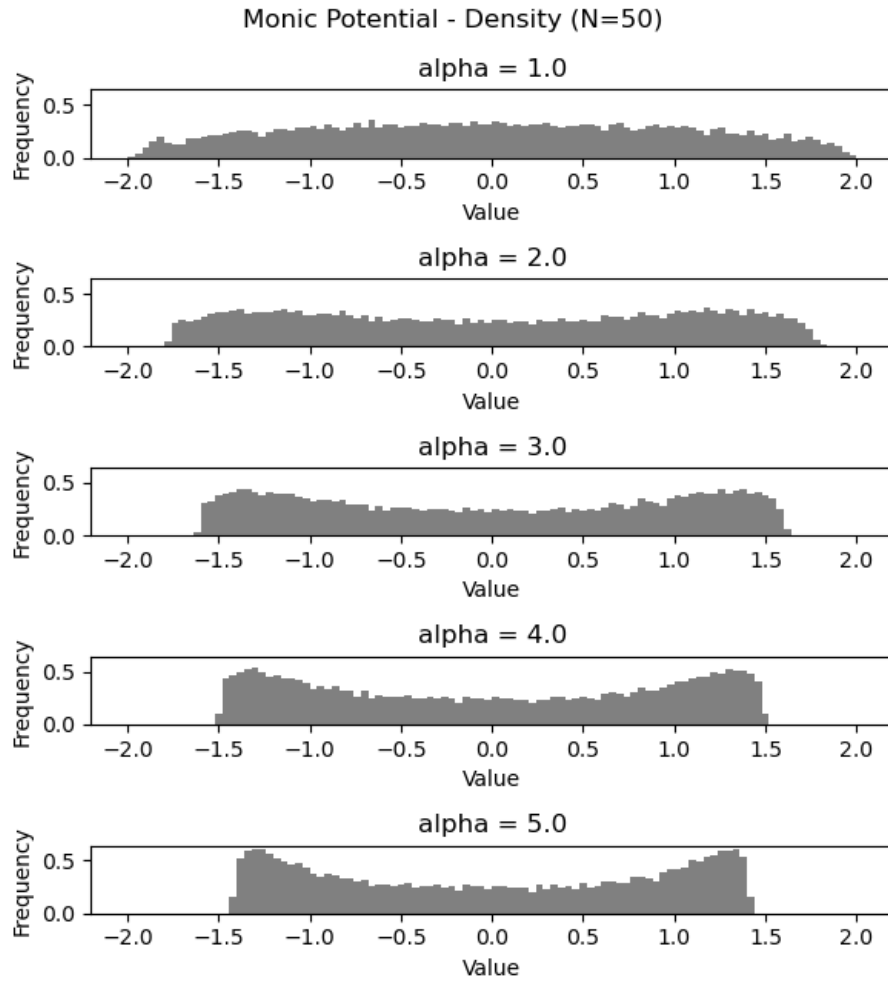


Figura 2.5: Validação para potencial mônico, $V(x) = \frac{t}{2\alpha}x^{2\alpha}$. Utilizamos 1000000 passos registrando a cada 500 a partir da metade dos passos. $t = 1$, $\Delta t = 0.1$, $\gamma = 10$, $\alpha = 0.1$. Para replicar a semente foi 987991650.

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
1. Revisão Literatura RMT	✓	✓	✓	✓	✓							
2. Métodos de Simulação				✓	✓	x	x	x				
3. Implementação algoritmos					✓	x	x	x	x	x	x	
4. Redação Relatórios				✓	✓						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.

Nome do Evento	Cidade - Sede	Data	Modalidade	Apresentação	Reserva Técnica
CBM	Rio de Janeiro / IMPA	24-28 Julho 2023	Presencial	Pôster - Oral	Sim
SIFSC	São Carlos / IFSC-USP	21-25 Agosto 2023	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.pimenta@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

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

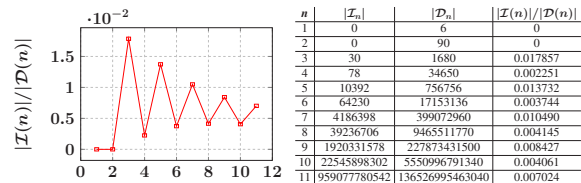
$$\implies |\mathcal{I}_n| = e^{nL(1+o(1))}$$

Podemos afirmar pelos resultados computacionais

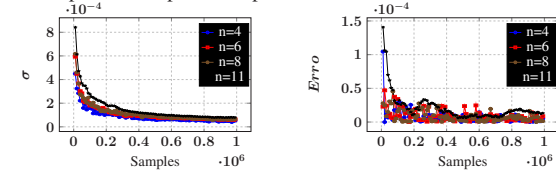
$$L \geq \frac{\log |\mathcal{I}_{11}|}{11} \approx 2.5.$$

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

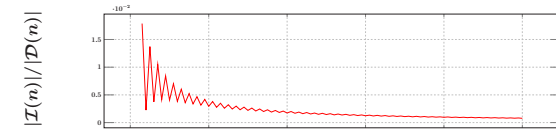
$$L = \lim_{n \rightarrow \infty} \frac{\log |\mathcal{I}_n|}{n} \leq \lim_{n \rightarrow \infty} \frac{\log |\mathcal{D}_n|}{n} = 3 \log 3.$$



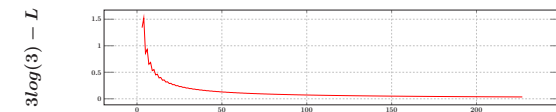
Validando para o espaço já conhecido de palavras podemos escolher os parâmetros de uma busca aleatória. Tomamos a média de 100 experimentos para cada ponto.



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>