PROJETO DE INICIAÇÃO CIENTÍFICA

“Simulando urnas de Pólya”.

Aluno: Felipe Augusto Ferreira de Castro

Orientador: Rodrigo Lambert

1 Introdução

Estudar a evolução de um modelo dinâmico em que retiram-se e colocam-se bolas de uma urna é um problema clássico em Probabilidade. O estudo de teoremas-limite (como lei dos grandes números e teorema central do limite, por exemplo) para a proporção de bolas de uma determinada cor na urna é, por si só, um assunto de grande importância científica. Além disso, tais resultados encontram inúmeras aplicações em áreas como modelagem estocástica, física, estudo comportamental, entre outras.

A proposta inicial presente projeto é desenvolver no aluno algumas ferramentas necessárias para atacar problemas relacionados ao assunto supracitado e introduzi-lo no estudo de um modelo probabilístico clássico: o modelo de urnas de Pólya. Finalmente, esperamos que o aluno desenvolva algoritmos capazes de simular computacionalmente alguns casos particulares para esse modelo, e confronte os resultados obtidos com os teoremas clássicos da literatura relacionada.

2 Justificativa e relevância do tema

Um processo estocástico é uma sequência de variáveis aleatórias indexadas por um conjunto de índices, que representa, por exemplo, a evolução temporal de um sistema. Exemplos de importância prática são: a flutuação da corrente em um circuito elétrico na presença do assim chamado ruído térmico; as mudanças aleatórias no nível de sinais de rádio sintonizados na presença de distúrbios meteorológicos; e o fluxo turbulento de um líquido ou gás. A estes pode-se adicionar muitos processos industriais acompanhados por flutuações aleatórias e também certos processos encontrados em geofísica (por exemplo, variações no campo magnético da Terra), biofísica (por exemplo, variações nos potenciais elétricos do cérebro registrados em um eletroencefalograma), economia, e Teoria da Informação (por exemplo, a codificação de sequências - o que inclui a compressão de dados). A área de Probabilidade e Processos Estocásticos constitui, portanto, uma ferramenta essencial para estudo de modelos que possuam algum tipo de comportamento aleatório.

Nesse contexto, o modelo de urnas se torna uma notável porta de entrada para o universo dos modelos probabilísticos discretos. Entender um modelo de urnas como um processo estocástico a tempo discreto, e a convergência das proporções de cores de bolas como o estudo de seus teoremas de convergência constituem uma imersão bastante intuitiva no mundo da probabilidade. No presente projeto, consideraremos modelos de urnas com q cores de bolas. Mais formalmente, estudaremos o processo estocástico q-dimensional $(X_n^{(1)}, X_n^{(2)}, \cdots, X_n^{(q)})_{n\in\mathbb{N}}$, onde $X_n^{(i)}$ representa o número de bolas da cor i na urna no instante n. O ponto central desse trabalho é entender a evolução do processo de proporção de cores, definido por

$$\left(\frac{X_n^{(1)}}{T_n}, \frac{X_n^{(2)}}{T_n}, \cdots, \frac{X_n^{(q)}}{T_n}\right)_{n\in\mathbb{N}},$$

onde $T_n = \sum_{i=1}^{q} X_n^{(i)}$ é o total de bolas na urna no instante n.

Aqui nos vemos diantes da necessidade de uma formulação que explique como evolui tal processo. Nesse sentido, deveremos definir como funciona a dinâmica de reposição de bolas na urna, assim como condições necessárias para o funcionamento dessa dinâmica. A seguir, explicaremos brevemente o conceito de matriz de reposição para urnas, e seu papel na dinâmica de reposição de bolas na urna.

Definição 2.1. Para cada $i \in \{1, 2, \cdots, q\}$, defina o vetor-coluna $\vec{V}(i) = (V_{1i}, V_{2i}, \cdots, V_{qi})$. A matriz de reposição de uma urna de Pólya de q cores é uma matriz $q \times q$ dada por

$$A = \left(\vec{V}(1), \vec{V}(2), \cdots, \vec{V}(q)\right).$$

A dinâmica da urna funciona da seguinte maneira: a urna começa com uma quantidade inicial de $(X_0^{(1)}, X_0^{(2)}, \cdots, X_0^{(q)})$ bolas de cada cor. A cada passo, escolhemos uma bola ao acaso da urna, observamos sua cor (digamos, a cor i), e a recolocamos na urna. Junto com ela, adicionamos mais V_{1i} bolas da cor 1, V_{2i} bolas da cor 2, seguindo até recolocarmos V_{qi} bolas da cor q na urna. Em palavras, o passo-a-passo da dinâmica da urna pode ser resumido a seguir.

- A urna inicia com uma quantidade de bolas de cada cor. O vetor inicial de bolas é dado por $(X_0^{(1)}, X_0^{(2)}, \cdots, X_0^{(q)})$;
- A cada passo, retiramos uma bola ao acaso, observamos sua cor, e a recolocamos;
- A cor da bola retirada indicará qual vetor-coluna da matriz de reposição A será usado para adicionar mais bolas na urna;

3 Objetivos

O objetivo geral do presente projeto é estudar algumas condições no número inicial de bolas na urna $(X_0^{(1)}, X_0^{(2)}, \cdots, X_0^{(q)})$ assim como condições específicas na matriz de reposição A sob as quais consigamos entender a evolução da urna. Tal estudo deverá abordar tópicos elementares e desenvolver no aluno ferramentas específicas que pertencem às seguintes áreas da Matemática, Estatística e Computação:

(a) Álgebra linear e teoria espectral;

(b) Teoremas-limite em espaços de probabilidade;

(c) Simulação estocástica;

(c) Programação em linguagem JavaScript.

Como objetivos específicos, podemos destacar

- Desenvolver no aluno conhecimentos e habilidades nas áreas de Probabilidade e Computação;
- Aprender técnicas e adquirir maturidade para entender problemas relacionados com essas áreas, assim como algumas de suas potenciais aplicações;
- Compreender do ponto de vista prático alguns tipos básicos de convergência em espaços de probabilidade, através do estudo do comportamento assintótico dos modelos de urnas;

- Conseguir aplicar tais conhecimentos aos modelos de difusão;
- Ser capaz de fazer simulação computacional de alguns modelos clássicos de urnas;
- Desenvolver uma plataforma com interface intuitiva para permitir simulação desses modelos por outras pessoas.

Vale aqui ressaltar um ponto interessante desse projeto. O aluno será inevitavelmente confrontado com conteúdos novos e desafiadores, além de ter a oportunidade de abordar alguns assuntos que não fazem parte da grade curricular do curso de bacharelado em computação, o qual ele se encontra matriculado.

4 Metodologia

O projeto será desenvolvido por aluno e professor através de estudos individuais e apresentação de seminários baseados na bibliografia do projeto. Reuniões semanais serão feitas, para falar sobre a evolução do trabalho e discutir eventuais dúvidas.

Além disso, o aluno deverá confeccionar um relatório final de atividades, e apresentar os resultados do seu trabalho em um ou mais eventos acadêmicos.

5 Cronograma de Atividades do Projeto

Antes de apresentar o cronograma, vale ressaltar que os tópicos abordados em uma etapa serão revisitados nas etapas seguintes. Também deverão acontecer "avanços" para tópicos de etapas seguintes durante etapas anteriores. Isso se deve principalmente a dois fatos. O primeiro é a forte interação entre esses conceitos, com uma grande interseção de assuntos. O segundo é que abordaremos conceitos teóricos com ajuda dos resultados obtidos pelas simulações, que têm um apelo intuitivo muito bom para ser explorado nesse sentido.

Além disso, a abordagem de todos os tópicos será feita com o desenvolvimento de simulações em JavaScript dos assuntos estudados.

Os trabalhos deverão ter como ponto de orientação o seguinte cronograma:

De Junho a Dezembro de 2020

a) Definição probabilística e exemplos de Processos Estocásticos e Cadeias de Markov;

b) Urnas de Pólya. Primeiras idéias;

c) Urnas de Friedman;

d) Urnas de Bagchi-Pal;

e) Álgebra linear e alguns elementos de teoria espectral. Aplicação aos itens (a), (b), (c) e (d).

De Janeiro a Julho de 2021

f) Definição de variável aleatória como uma função mensurável num espaço de probabilidade;

g) Tipos de convergência. Lei dos grandes números e teorema central do limite;

h) Dos pseudo-códigos aos algoritmos em linguagem JavaScript. Intuindo teoremas de convergência.

i) Abordando computacionalmente modelos não-triviais: difusão em meios determinísticos, e urnas com a propriedade dos lapsos de memória;

j) Confecção do relatório final.

6 Algumas referências bibliográficas

1. M. González-Navarrete & R. Lambert. *Non-Markovian random walks with memory lapses.* J. Math. Phys. 59 (11), 113301. (2018)

2. M. González-Navarrete & R. Lambert. *The diffusion of opposite opinions in a randomly biased environment.* J. Math. Phys. 60 (11), 113301. (2019)

3. P. G. Hoel, S. C. Port, & C. J. Stone. *Introduction to Stochastic Processes*, Houghton Mifhin Series in Statistics. (1972)

4. S. Janson. *Functional limit theorems for multitype branching processes and generalized Pólya urns.* Sotch. Proc. Appl. 110 (2004) 177 – 245.

5. V.V. Júnior. *Notas de aula de processos estocásticos.* (2018)

6. H.M. Mahmoud, Pólya Urn Models. Chapman and Hall (2008)

7. B. Ripley. *Sotchastic Simulation.* John Wiley & Sons (1987)

8. S. M. Ross. *Probabilidade: um curso moderno com aplicações*, oitava edição. ed. Bookman (2010)

9. S. M. Ross. *Introduction to Probability Models*, 10th. ed. Elsevier. (2010)

10. S. M. Ross. *Simulation*, 5th. ed. Elsevier. (2013)