PROFESSOR DE MATEMÁTICA ONLINE
Revista eletrônica da Sociedade Brasileira de Matemática

PMO v.12, n.3, 2024
ISSN: 2319-023X
https://doi.org/10.21711/2319023x2024/pmo1222

Uma simulação em *Python* do paradigma bayesiano da estatística

Saulo Cavalcante dos Reis[1]   Renato de Sá Teles   Anderson Augusto Ferreira

Resumo

Este trabalho ilustra um caso de inferência estatística no contexto do experimento mental de Thomas Bayes, isto é, o experimento originário do paradigma bayesiano da estatística. Partindo de uma formalização da teoria da probabilidade, realizou-se uma simulação computacional em Python desse experimento mental, cujos resultados tornaram possível a inferência estatística bayesiana.

Palavras-chave: Teorema de Bayes; Probabilidade; Simulação Computacional.

Abstract

This paper shows a case of statistical inference in the context of Thomas Bayes' mental experiment, which is the origin of the bayesian paradigm of statistics. Beginning with a formalization of probability theory, a computer simulation of this mental experiment in Python was realized, whose results made possible the bayesian inference.

Keywords: Bayes' Theorem; Probability; Computer Simulation.

1. Motivação

A Base Nacional Comum Curricular (BNCC), quanto à Matemática a ser trabalhada no Ensino Médio, propõe "a consolidação, a ampliação e o aprofundamento das aprendizagens essenciais (...), a fim de possibilitar que os estudantes construam uma visão mais integrada da Matemática, ainda na perspectiva de sua aplicação à realidade" [1, p. 527]. A expressão "aplicação à realidade" sugere um aspecto por vezes esquecido da Matemática, que é sua capacidade de descrever realidades objetivas através de linguagem simbólica, linguagem essa que exige bons níveis de abstração por parte dos estudantes para sua compreensão apropriada. Também, essa "aplicação à realidade" é aparentemente posta em prática na descrição da primeira das Competências Específicas da Matemática e Suas Tecnologias para o Ensino Médio, do mesmo documento: "utilizar estratégias, conceitos e procedimentos matemáticos para interpretar situações em diversos contextos, sejam atividades cotidianas, sejam fatos das Ciências da Natureza e Humanas, das questões socioeconômicas ou tecnológicas, divulgados por diferentes meios, de modo a contribuir para uma formação geral" [1, p. 531].

Essa competência, segundo a BNCC, prevê que os estudantes sejam capazes, por exemplo, "de analisar criticamente o que é produzido e divulgado nos meios de comunicação" [1, p. 532]. Por isso, é muito

[1]Egresso do Profmat da Unifesp *Campus* Diadema/SP.

importante que os estudantes de hoje estejam mais equipados com ferramentas matemáticas que permitam com maior facilidade distinguir informações verdadeiras das falsas, para que o desenvolvimento desses estudantes até a plena cidadania não fique comprometido pela falta dessas ferramentas. Tendo em vista essa motivação, no contexto da Matemática, é difícil não pensar no uso da Probabilidade e da Estatística como ferramentas de análise crítica, dada sua aplicabilidade quase universal nas ciências. Entendendo a necessidade adicional de formar os estudantes para serem não apenas cidadãos, mas futuros cientistas, torna-se ainda mais relevante a importância da obtenção apropriada de dados sobre a realidade — sob inevitáveis limitações de natureza social ou econômica — e também da análise crítica rigorosa sobre esses dados.

A inferência estatística é, assim, a ferramenta matemática por excelência em favor da competência descrita. Chama a atenção um dos itens listados como "Habilidades" a serem desenvolvidas pelos estudantes segundo a organização curricular proposta na BNCC, no quadro da Probabilidade e Estatística: "Identificar situações da vida cotidiana nas quais seja necessário fazer escolhas levando-se em conta os riscos probabilísticos (usar este ou aquele método contraceptivo, optar por um tratamento médico em detrimento de outro etc)" [1, p. 546].

Esta habilidade parece ser a decorrência imediata do aprendizado e da aplicação da inferência estatística. Dentro dela, considera-se que existem dois grandes paradigmas: o paradigma frequencista e o paradigma bayesiano. Neste trabalho, pretende-se ilustrar um caso de inferência estatística segundo o paradigma bayesiano, orientado pelo arcabouço teórico da probabilidade. Os autores deste trabalho entendem que o ensino da probabilidade e estatística no Ensino Médio segundo o paradigma bayesiano – e claro, do Teorema de Bayes – tem muito a ganhar, levando-se em conta as considerações feitas sobre as conexões quase explícitas deste teorema com as propostas educacionais e curriculares da BNCC.

É comum dividir a estatística em dois ramos: a descritiva e a inferencial. O ramo da estatística descritiva já é introduzido no currículo da Matemática do Ensino Médio através de conceitos como média, desvio-padrão, mediana e moda, por exemplo, que servem como propriedades de um conjunto de dados. Outro modo pelo qual a estatística descritiva é introduzida no Ensino Médio é pela descrição gráfica de dados com uso de histogramas, por exemplo. Essas propriedades e ferramentas servem para suscitar no estudante a busca de tendências e/ou significado do conjunto de dados estudado, mas tal preocupação já faz parte do outro ramo da estatística, a inferencial. Nesse sentido, o paradigma bayesiano acrescenta uma ferramenta matemática importante, especialmente para explicitar o nível de incerteza sobre eventuais inferências feitas no estudo do conjunto de dados.

O Teorema de Bayes representa um marco na história do raciocínio lógico e o primeiro grande triunfo da inferência estatística [3]. Contudo, assim como toda fórmula matemática, requer-se um cuidado especial: dada a possibilidade de aplicação quase universal nas ciências, o conhecimento prévio das probabilidades envolvidas precisa ser garantido para haver menor margem de erro no cálculo da probabilidade desejada. Isto é significativo, quando o conhecimento envolvido se relaciona com alguma decisão de impacto público.

2. O experimento mental de Bayes e seu teorema

Assim como é comum na História da Matemática dar nomes a fórmulas, teoremas e outros desenvolvimentos homenageando seus inventores ou primeiros descobridores, não foi diferente com o teorema de Bayes, cujo nome vem do reverendo e matemático inglês Thomas Bayes (ilustrado na Figura 1), que se baseou em definições e teoremas da probabilidade condicional no seu ensaio entitulado *An Essay towards solving a Problem in the Doctrine of Chances*, de 10 de novembro de 1763. Esse ensaio foi editado por Richard Price dois anos após a morte de Bayes, e continha alguns anexos ao original, autorados pelo editor. Prize escreveu a introdução ao ensaio, descrevendo as bases filosóficas sob as quais se assentavam o trabalho de Bayes. O

Figura 1: Thomas Bayes (1701—1761).
Fonte: https://upload.wikimedia.org/wikipedia/commons/d/d4/Thomas_Bayes.gif

ensaio editado foi direcionado a John Canton, à época membro da Royal Society e receptor dos escritos de Bayes.

Numa tradução literal para o português, o título do ensaio vem como "Um ensaio para resolver um Problema da Doutrina das Chances"[2]. No século XVIII, "Doutrina das Chances" era o nome que se dava ao que hoje chamamos de "Teoria da Probabilidade". Neste trabalho, Bayes buscou estabelecer um método matemático com probabilidades preestabelecidas, e que fizesse uso de evidências presentes, para descobrir a probabilidade de causas, ou a "probabilidade inversa". A pergunta motivadora de Bayes poderia ser formulada do seguinte modo: como podemos descobrir a probabilidade de um evento no futuro, dado que no passado o mesmo evento tenha ocorrido ou não uma quantidade de vezes sob certas condições? O teorema pode ser descrito numa expressão simplista: *crenças iniciais* + *evidências objetivas recentes* = *crenças novas e aprimoradas* [5]. Essa expressão sugere uma profunda aplicabilidade desse método para as ciências naturais, sociais e econômicas, e parece ir de acordo com as próprias palavras de Prize logo no início da sua introdução, se direcionando a Canton: "A filosofia experimental, você perceberá, está intimamente interessada no assunto"[2].

O teorema de Bayes surgiu a partir de um experimento mental ([2]), que pode ser facilmente reproduzido tanto na prática como por simulação computacional. O experimento mental de Bayes começa com uma mesa retangular, disposta paralela ao plano do chão, e uma pequena bola, a ser lançada na mesa de modo que possa parar em qualquer posição dela com igual probabilidade, e essa posição deve ser facilmente verificada quando necessário. Sem enxergar a mesa, uma pessoa – chama-la-emos de "Bayes" – lança a bola pela primeira vez, e pede a uma segunda pessoa – chama-la-emos de "ajudante" – que marque na mesa o local onde a bola parou, para que Bayes descubra mais tarde essa localização sem olhar. Depois de a marcação ser feita, Bayes inicia o processo de descoberta do local da marcação: o ajudante toma a bola e a lança na mesa, reportando a Bayes se essa bola ficou à esquerda ou à direita da marcação inicial. Com essa informação, Bayes pensa nas possíveis posições onde a marcação inicial teria sido feita. Esse processo continua, recursivamente, com o ajudante jogando a bola na mesa e informando a Bayes sobre a posição da bola à esquerda ou à direita da marcação. A cada informação nova, as possíveis posições da marcação tornam-se mais claras para Bayes. Em algum momento, Bayes afirma, com certo grau de confiança, em qual região da mesa a marcação inicial foi feita. A Figura 2 ilustra um esquema simplificado do experimento, exibindo a direção de lançamento da bola na mesa pelo ajudante, e Bayes numa posição em que ele não possa enxergar a mesa.

Conceitualmente falando, o raciocínio de Bayes era simples. A crença ou opinião anterior, e que parece arbitrária (no experimento de Bayes, representa o chute sobre o local possível da marcação), começa a ser refinada com a introdução de dados objetivos (representados no experimento pelas informações dos lançamentos da bola pararem à esquerda ou à direita da marcação). O resultado disso é uma crença ou opinião

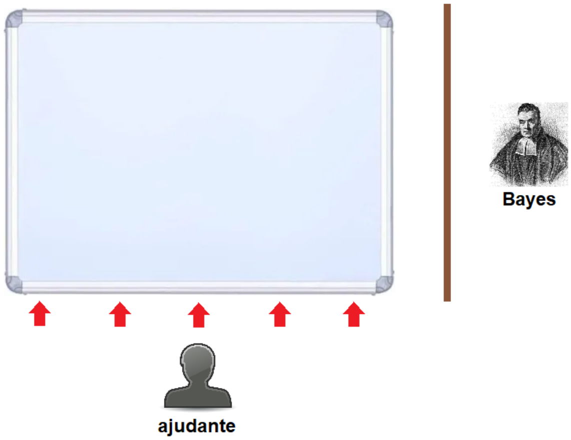

Figura 2: Ilustração simplificada do esquema do experimento mental de Bayes.
Fonte: autoria própria.

posterior mais apurada, mais precisa, sobre o paradeiro da marcação. Com mais iterações desse processo, a crença posterior obtida num experimento torna-se a opinião anterior para o próximo experimento. Quanto mais iterações desse experimento, mais certeira e coincidente com a realidade vai ficando a crença posterior. Daí, justifica-se a expressão simplista descrita anteriormente, de que "crenças iniciais + evidências objetivas recentes = crenças novas e aprimoradas". Tecnicamente falando, a fórmula de Bayes pode ser descrita numa frase: *a probabilidade* a priori *multiplicada pela verossimilhança é proporcional à probabilidade* a posteriori.

$$\mathrm{p}(\mathrm{A}|\mathrm{B}) = \frac{\mathrm{p}(\mathrm{A}) \times \mathrm{p}(\mathrm{B}|\mathrm{A})}{\mathrm{p}(\mathrm{B})}. \tag{1}$$

Nessa fórmula:

- $\mathrm{p}(\mathrm{A}|\mathrm{B})$ representa a probabilidade *a posteriori* da proposição A com respeito à evidência B.
- $\mathrm{p}(\mathrm{A})$ representa a probabilidade *a priori* da proposição A.
- $\frac{\mathrm{p}(\mathrm{B}|\mathrm{A})}{\mathrm{p}(\mathrm{B})}$ representa a verossimilhança da evidência B para a proposição A.

Essa relação matemática entre probabilidades, por mais simples que possa parecer, tem implicações muito interessantes para o próprio desenvolvimento do pensamento científico, e de como se torna possível sistematizar informações que tornem plausível uma hipótese científica através da avaliação do grau de confiança nela. O Teorema de Bayes, bem compreendido, permitiria avaliar, por exemplo, a probabilidade de uma hipótese ser verdadeira a partir de certas evidências e de reavaliar essa mesma probabilidade quando novas evidências relevantes surgem. O grau de confiança na hipótese avaliada pode mudar de acordo com as evidências novas que eventualmente apareçam, e o teorema torna possível sistematizar as contribuições de cada evidência à medida que elas surjam.

3. Axiomatização da Teoria da Probabilidade

Faremos uma axiomatização da teoria da probabilidade, expondo definições e conceitos, até a dedução do Teorema de Bayes na próxima seção. Essa axiomatização é baseada em Kolmogorov ([4]).

Definição 1 (Número de elementos de um conjunto)**.** Para qualquer conjunto C com uma quantidade finita de elementos, a função $\#(C)$ retorna seu número de elementos.

Definição 2 (Espaço Amostral)**.** Considerando-se $e_1, e_2, \ldots$ como elementos que representem de modo único os possíveis resultados de um experimento, chama-se de espaço amostral ao conjunto de todos esses elementos, a ser denotado por $\Omega = \{e_1, e_2, \ldots\}$.

Definição 3 (Evento)**.** Um evento E é um subconjunto do espaço amostral Ω, isto é, $E \subset \Omega$.

Definição 4 (Evento Elementar)**.** Classifica-se um evento E como sendo elementar se ele for conjunto unitário, isto é, se $\#(E) = 1$.

Definição 5 (Espaço de Eventos)**.** O espaço de eventos S é o conjunto de todos os possíveis eventos de Ω; isto é, $S = \{s \mid s \subset \Omega\}$.

Definição 6 (Função de Probabilidade)**.** A função de probabilidade $P(E)$ é uma função com domínio S, contradomínio $\mathbb{R}$, e uma lei que satisfaça a três condições:

1. A probabilidade de qualquer evento elementar é sempre maior ou igual a zero, isto é,
$$P(\{e_i\}) \geq 0, \forall i \in \mathbb{N}.$$
2. Para $\forall E \in S$ em que $\#(E) > 1$, é necessário que a probabilidade do evento E seja igual à soma das probabilidades dos eventos elementares que, em união, resultariam no próprio evento E, ou seja,
$$P(E) = \sum_{e \in E} P(\{e\}).$$
3. A probabilidade de Ω, isto é, a soma das probabilidades de todos os eventos elementares, é igual a um, isto é,
$$P(\Omega) = \sum_{e \in \Omega} P(\{e\}) = 1.$$

Explicação. A função de probabilidade é a função que associa um evento a um valor, e esse valor é costumeiramente chamado de *probabilidade*. Essa função – mais especificamente, a lei da função – precisa ser concebida para satisfazer às três condições. Nesse sentido, as probabilidades dos eventos são arbitrárias, mas devem obedecer a critérios que satisfaçam algum propósito experimental. Um critério muito usado é o do *Princípio da Indiferença* (ou Princípio da Razão Insuficiente). □

Definição 7 (Espaço Probabilístico)**.** O espaço probabilístico (Ω, P) é a dupla constituída por um espaço amostral Ω e uma função probabilística $P(E)$, sendo que E é um evento qualquer do espaço de eventos S de Ω.

Definição 8 (Partição do Espaço Amostral)**.** Os eventos A_1, A_2, …, A_n constituem uma partição $\mathcal{P}$ do espaço amostral Ω quando as seguintes condições foram satisfeitas:

1. $A_i \in S, \forall i \in \mathbb{N}$.

2. $A_i \cap A_j = \emptyset, \forall i, j \in \mathbb{N}$ e $i \neq j$, isto é, eventos distintos são mutuamente exclusivos em Ω.

3. $\bigcup_{i=1}^{n} A_i = \Omega$, isto é, a união de todos os eventos resulta no espaço amostral.

Teorema 1 (Lei da Probabilidade Total)**.** *Considerando eventos* $E_1, E_2, \dots, E_n$ *que constituem uma partição de* Ω *(sendo por isso mutuamente exclusivos), e um outro evento qualquer* A*, os eventos* $A \cap E_1$, $A \cap E_2, \dots, A \cap E_n$ *também serão mutuamente exclusivos. Sendo assim, a probabilidade do evento* A *poderá ser obtida:*

$$\begin{aligned} P(A) &= P(A \cap E_1) + P(A \cap E_2) + \cdots + P(A \cap E_n) = \\ &= \sum_{i=1}^{n} P(A \cap E_i). \end{aligned} \tag{2}$$

que é a equação da Lei da Probabilidade Total.

Teorema 2 (Probabilidade do Evento Complementar)**.** *Considerando um espaço amostral* $\Omega = \{e_{k_1}, e_{k_2}, \dots, e$ *e um possível evento dele* $A = \{e_{k_1}, e_{k_2}, \dots\}$*, o evento complementar de* A *seria* $\overline{A} = \{e_{w_1}, e_{w_2}, \dots\}$*, com* $k_i \neq w_j$ *para* $\forall i, j \in \mathbb{N}$.

De acordo com os Itens 2 e 3 da Definição 6:

$$\begin{aligned} 1 = P(\Omega) = \sum_{e \in \Omega} P(\{e\}) &= \sum_{i=1}^{\infty} P(\{e_{k_i}\}) + \sum_{j=1}^{\infty} P(\{e_{w_j}\}) = \\ &= P(A) + P(\overline{A}) \Longrightarrow P(\overline{A}) = 1 - P(A). \end{aligned} \tag{3}$$

Definição 9 (Probabilidade Condicional)**.** Sejam dois eventos E e Ω' do espaço amostral Ω, em que $P(\Omega') \neq 0$. Então a probabilidade condicional de sucesso de E em Ω' é dada por

$$P(E|\Omega') = \frac{P(E \cap \Omega')}{P(\Omega')}. \tag{4}$$

A partir da Definição 9 e usando H no lugar de Ω', tem-se $P(E \cap H) = P(E) \times P(H|E)$. Como não há diferença entre $E \cap H$ e $H \cap E$, então é possível escrever

$$P(E \cap H) = P(H \cap E) \Longrightarrow P(E) \times P(H|E) = P(H) \times P(E|H)$$

de onde deduzimos o

Teorema 3 (Teorema de Bayes)**.** *Sejam dois eventos* E *e* H *do espaço amostral* Ω*, em que* $P(H) \neq 0$ *e* $P(E) \neq 0$*. Então a probabilidade condicional de sucesso de* H *em* E *é dada por*

$$P(H|E) = \frac{P(H) \times P(E|H)}{P(E)}. \tag{5}$$

Explicação. Existem, de cada lado da igualdade, probabilidades condicionais "trocadas"; isto é, de um lado, a probabilidade de um evento H em E, e, de outro, a probabilidade do evento E em H. Isto é o que foi chamado historicamente de *probabilidade inversa*. □

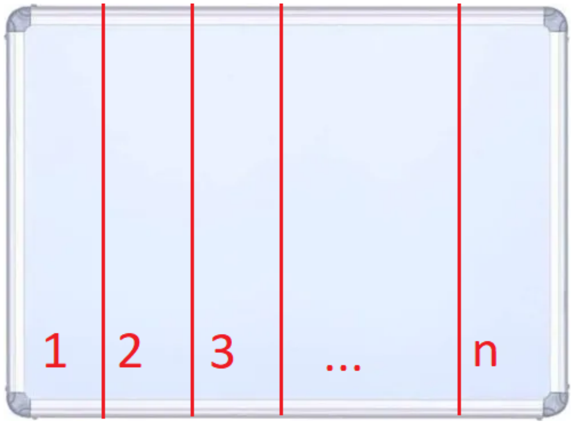

Figura 3: Ilustração da mesa dividida em **n** regiões verticais.
Fonte: autoria própria.

Definição 10 (Razão de Chances)**.** Seja um evento E em um espaço amostral Ω, com $0 < P(E) < 1$. Chama-se de razão de chances de E ao valor $O(E)$ que representa a razão entre a probabilidade do seu sucesso e a probabilidade da sua falha, isto é,

$$O(E) = \frac{P(E)}{P(\overline{E})} = \frac{P(E)}{1 - P(E)}. \tag{6}$$

Uma interpretação prática bastante comum para a noção de razão de chances é a de que, se $O(E)$ puder ser expresso por uma fração p/q com $p, q \in \mathbb{N}$, então para cada $p + q$ repetições do experimento, há expectativa de p sucessos e q falhas do evento E.

4. Descrição da simulação computacional

Para realizar a simulação, dividimos o espaço da mesa em um número finito de regiões verticais **n** para que possamos localizar as bolas lançadas. Essas bolas serão lançadas na mesa na direção vertical, não importando o sentido delas (de baixo para cima ou de cima para baixo).

Antes de irmos à didática dessa simulação, é preciso mostrar que há um modelo probabilístico nesse experimento. Por simplicidade, será usada a notação $P(e)$ em vez de $P(\{e\})$ quando o conjunto for elementar. Sendo assim, o modelo probabilístico desse experimento segundo a Definição 7 será:

- Os possíveis resultados de um lançamento de bola são as regiões numeradas de 1 à **n**, por isso, $\Omega = \{1, 2, \ldots, n\}$, e $\#(\Omega) = n$.
- Assumindo o Princípio da Indiferença – a posição final da bola pode ser qualquer uma das **n** regiões –, a função de probabilidade $P(E)$ é tal que $P(1) = P(2) = \cdots = P(n) = \dfrac{1}{\#(\Omega)} = \dfrac{1}{n}$.

No entanto, o experimento de Bayes não se resume apenas em lançar bolas na mesa e localizá-las em alguma das **n** regiões. O objetivo do experimento é descobrir a posição da marcação a partir de informações colhidas a partir dos lançamentos da bola. O processo para fazer essa descoberta é o que se chama de inferência bayesiana.

Suponhamos, portanto, que Bayes lançou a bola na mesa e o ajudante já tenha feito a marcação na região **k**, sendo $k \in \{1, 2, \ldots, n\}$. Nesse momento, só o ajudante sabe em qual região da mesa há marcação, enquanto Bayes não tem essa informação. Sabendo que os próximos lançamentos da bola pelo ajudante não dependem do resultado de lançamentos anteriores, o modelo probabilístico de cada um desses lançamentos é o que foi descrito acima, com $P(e) = \frac{1}{n}$, sendo $\{e\}$ um evento elementar qualquer.

Com esse modelo, o objetivo de Bayes é descobrir o valor de **k**. O ajudante fará novos lançamentos da bola, e reportará a Bayes uma informação sobre a posição do lançamento da bola. No modelo aqui desenvolvido, a referência da informação será sempre a marcação inicial: o ajudante reportará se *a marcação ficou à esquerda ou à direita dos lançamentos* – trocando-se, portanto, o ponto de referência em relação à descrição feita na Seção 2. Conforme as informações que o ajudante fornece, Bayes poderá, gradualmente, afirmar com maior confiança em qual região da mesa a marcação inicial tinha sido feita.

O processo de inferência que Bayes fará começa pela avaliação das probabilidades de a marcação estar à esquerda ou à direita dos novos lançamentos. O evento "marcação está à esquerda de um lançamento" será representado por $Q = \{k + 1, k + 2, \ldots, n\}$ (que representam os possíveis lugares onde o lançamento teria caído) e, por isso, $\#(Q) = n - k$. De modo análogo, o evento "marcação está à direita de um lançamento" será representado por $D = \{1, 2, \ldots, k - 1\}$, com $\#(D) = k - 1$. Desse modo, tem-se as probabilidades de a marcação estar à esquerda ou à direita de cada lançamento novo, condicionados à marcação ter sido feita na região k:

$$P(Q|k) = \frac{\#(Q)}{\#(\Omega)} = \frac{n-k}{n}, \tag{7}$$

$$P(D|k) = \frac{\#(D)}{\#(\Omega)} = \frac{k-1}{n}. \tag{8}$$

Interessante notar que os eventos $\{k\}$, Q e D representam uma partição do espaço amostral Ω (ver Definição 8) e que, por isso, $P(k)+P(Q|k)+P(D|k) = 1$. Na prática, o que as probabilidades $P(Q|k)$ e $P(D|k)$ sugerem é que, usando uma ideia semelhante à da razão de chances (ver Equação 6), para cada **n** lançamentos da bola, espera-se que em **k-1** lançamentos a marcação esteja à direita de um lançamento, e em **n-k** lançamentos a marcação esteja à esquerda do lançamento. Essas seriam as *frequências esperadas* dos lançamentos. Contudo, as *frequências reais* dos lançamentos podem ser diferentes, mas conforme a quantidade de lançamentos aumente, *a tendência é que as frequências reais dos resultados dos lançamentos se aproxime das frequências esperadas* segundo o espaço amostral escolhido inicialmente.

A inferência bayesiana dar-se-á, portanto, à medida que mais novos lançamentos forem feitos. O ajudante informa as frequências reais dos resultados, de tal modo que Bayes possa melhorar gradualmente a sua "opinião inicial" sobre a posição da marcação. Essa é uma maneira de proceder na inferência. Outra maneira que Bayes poderia proceder seria simplesmente aguardar os resultados de uma quantidade arbitrária de lançamentos, para que, ao tomar conhecimento das frequências reais e compará-las com as frequências esperadas, possa ter maior certeza sobre qual seria a posição da marcação.

O código-fonte do algoritmo de simulação do experimento de Bayes, escrito em linguagem *Python*, está disponível em https://github.com/saulocreis/experimento_mental_bayes. Façamos agora exemplificações da inferência fazendo uso do algoritmo de simulação do experimento.

5. Execução da Simulação Computacional e Resultados

Nesta seção, apresentamos uma exemplificação do processo de inferência que será feito pelo programa simulador usando uma mesa dividida em 6 regiões de igual área. Esse experimento numa mesa de 6 regiões é interessante pela sua didática, pois poderá ser facilmente traduzido como se fosse também um lançamento de um dado de 6 faces, objeto que um estudante de Ensino Médio tem fácil acesso, além da grande quantidade de exemplos que já existem em livros didáticos com esse objeto. O raciocínio empregado para este experimento é basicamente o mesmo para o lançamento com dado de 6 faces.

Como mencionado anteriormente, há duas maneiras de realizar a inferência. Uma ocorre gradualmente, e a outra após uma quantidade arbitária de lançamentos. Neste trabalho, fez-se a simulação de modo que a inferência seja feita gradualmente, ilustrando o raciocínio bayesiano.

Suponhamos que a marcação tenha sido feita na região 4, isto é, $\mathrm{k} = 4$. Então, há 3 regiões em que a marcação fique à direita do lançamento, de modo que $\#(\mathrm{D}) = 3$ e, há 2 regiões em que a marcação fique à esquerda do lançamento, de modo que $\#(\mathrm{Q}) = 2$. Mais exatamente, essas probabilidades serão $\mathrm{P}(\mathrm{D}|4) = \frac{3}{6}$ e $\mathrm{P}(\mathrm{Q}|4) = \frac{2}{6}$ respectivamente.

Ao executar o programa, o usuário faz o papel de Bayes, ou seja, o de descobrir onde a marcação foi feita. Já o programa faz o papel do ajudante, exibindo mensagens para ajudar o usuário durante a realização do experimento para que ele descubra o local da marcação. Para fins de esclarecimento, as menções futuras a "programa" ou a "usuário", em termos da descrição do experimento mental (na página 330), referir-se-ão, respectivamente, ao ajudante e a Bayes.

Inicialmente, a primeira informação que o usuário precisa é de saber em quantas regiões a mesa foi dividida. A Figura 4 ilustra a tela com as instruções iniciais do programa até o momento em que é solicitado o número de regiões em que a mesa será dividida:

Figura 4: Tela contendo a informação sobre a quantidade de regiões na mesa

```
=== ALGORITMO BAYESIANO ===
Para sair do programa, basta apertar ENTER sem fornecer informacao util ao programa
     em qualquer momento.
Numero de regioes da mesa: 6
```

Depois que o usuário informa a quantidade de regiões – neste exemplo, 6 –, o programa faz um lançamento da bola na mesa e faz a marcação na mesa. Essa informação não é revelada ao usuário, pois o objetivo desse usuário é justamente descobrir em qual região da mesa a marcação foi feita. A partir daí, o programa passará a fazer novos lançamentos da bola, sempre solicitando ao usuário quantos novos lançamentos deseja que sejam feitos. A Figura 5 ilustra a tela em que são exibidas as mensagens do programa até esse ponto:

Figura 5: Tela que exibe a informação de um lançamento na mesa

```
A mesa foi dividida igualmente em regioes numeradas de 1 a 6, da esquerda para a
    direita.
(quanto mais proximo de 1, a regiao estah mais a esquerda)
(quanto mais proximo de 6, a regiao estah mais a direita)
A bola foi lancada na mesa. O ajudante marcou a posicao dela em uma dessas regioes.

SEU OBJETIVO: descobrir em qual regiao a marcacao foi feita.
Agora o ajudante vai lancar a bola quantas vezes voce desejar para obter informacao
    .
Depois desses lancamentos, ele lhe dira em quantos deles a marcacao ficou a
    ESQUERDA ou a DIREITA da bola.
```

```
AJUDANTE: Quantas vezes devo lancar a bola? 1
```

Procura-se simular o experimento mental de Bayes de um modo mais "realista" possível, imaginando duas pessoas (Bayes e seu ajudante) diante de uma mesa realizando o experimento de fato, e buscar com isso entender como inferências bayesianas poderão ser feitas ao longo do processo para descobrir onde a marcação foi feita. Por isso, será informado ao programa que apenas um lançamento será feito por vez (como exibido na Figura 5).

Após o programa "fazer o lançamento", ele informa ao usuário quantos deles ocorreram à esquerda e à direita da marcação original, conforme a Figura 6 ilustra.

Figura 6: Tela que exibe o resultado do lançamento

```
AJUDANTE: A bola foi lancada 1 vez(es) agora. Em relacao a bola, a marcacao ficou
1 vez(es) a direita (100.0 % desses lancamentos)
0 vez(es) a esquerda (0.0 % desses lancamentos)

AJUDANTE: Total de 1 lancamentos, sendo:
1 a direita (100.0 % do total)
0 a esquerda (0.0 % do total)
```

A partir de agora, a busca pela resposta dar-se-á com as informações exibidas pelo programa. O que é possível inferir sobre o resultado acima?

Ainda não há certeza do local de a marcação, mas já é possível ter certeza de que é impossível a marcação ter sido feita na região 1. Lembrando que as regiões são marcadas na mesa da esquerda para a direita (sendo 1 aquela mais à esquerda da mesa e 6 aquela mais à direita), então pode-se inferir com esse primeiro lançamento que, se a marcação tivesse sido feita na região 1, é impossível que essa região esteja à direita de quaisquer lançamentos.. Essa impossibilidade é matematicamente traduzida pelo valor obtido com as equações da página 335, com $\mathrm{n} = 6$ e $\mathrm{k} = 1$: $\mathrm{P(D|1)} = \frac{1-1}{6} = 0$.

Neste momento, percebe-se que cada novo lançamento significa uma nova informação a respeito da marcação. Por causa do resultado deste lançamento, já se entende que é necessário *condicionar* a probabilidade da marcação estar em alguma das regiões da mesa, ou seja, obter uma função de probabilidade que leve em consideração as condições ou restrições impostas. No início, foi assumido o Princípio da Indiferença para essa função. Agora, esse mesmo princípio não pode ser o único critério para a função de probabilidade. A inferência bayesiana, portanto, corresponde a um método de incorporação gradual de novas informações, sejam elas na forma de condições ou de restrições. No final das contas, o objetivo do usuário neste experimento é diminuir a incerteza que havia sobre onde a marcação inicial poderia estar.

Depois de realizar o primeiro lançamento, sabendo que a marcação está à direita de um lançamento, é preciso uma função de probabilidade que esteja atualizada com as informações disponíveis até então. Simbolizando essa informação por D – isto é, o evento "marcação está à direita de um lançamento" –, e lembrando que $\mathrm{P(D|1)} = 0$, passar-se-á a trabalhar com uma função de probabilidade atualizada, indicada por $\mathrm{P_1(E)}$, que será entendida como $\mathrm{P(E)}$ condicionada a D, ou seja, $\mathrm{P_1(E)} = \mathrm{P(E|D)}$. Essa função terá seus valores avaliados para cada elemento de Ω, fazendo uso do Teorema de Bayes (Equação 5). Avaliando primeiramente $\mathrm{P_1(1)} = \mathrm{P(1|D)}$, tem-se:

$$\mathrm{P_1(1)} = \mathrm{P(1|D)} = \mathrm{P(D|1)}\frac{\mathrm{P(1)}}{\mathrm{P(D)}} = 0.$$

Isso já era esperado pois, se a marcação estivesse na região 1, era impossível que essa marcação estivesse à direita de qualquer lançamento posterior. Então, por isso, se fosse observado que a marcação está à direita de um lançamento, então é impossível que a marcação estivesse na região 1, não importando o valor de $\mathrm{P(D)}$. Contudo, o valor de $\mathrm{P(D)}$ será importante na avaliação de $\mathrm{P_1(E)}$ para os outros elementos de Ω. O modo de obter tal valor sem qualquer presunção sobre **k** vem da Lei da Probabilidade Total (Equação 2):

$$\begin{aligned}\mathrm{P(D)} &= \mathrm{P(D\cap\{2\}) + P(D\cap\{3\}) + P(D\cap\{4\}) + P(D\cap\{5\}) + P(D\cap\{6\})} = \\ &= \mathrm{P(2)P(D|2) + P(3)P(D|3) + P(4)P(D|4) + P(5)P(D|5) + P(6)P(D|6)} = \\ &= \frac{1}{6}\left(\frac{1}{6}+\frac{2}{6}+\frac{3}{6}+\frac{4}{6}+\frac{5}{6}\right) = \frac{15}{36}.\end{aligned}$$

O valor de $\mathrm{P(Q)}$ seria obtido de modo similar. Pela simetria do problema, o resultado também seria $\mathrm{P(Q)} = \frac{15}{36}$.

O restante dos valores de probabilidade de $\mathrm{P_1(E)}$ são assim encontrados:

- $\mathrm{P_1(2) = P(2|D) = P(D|2)\frac{P(2)}{P(D)}} = \frac{1}{6}\times\frac{\frac{1}{6}}{\frac{15}{36}} = \frac{1}{6}\times\frac{6}{15} = \frac{1}{15}$.
- $\mathrm{P_1(3) = P(3|D) = P(D|3)\frac{P(3)}{P(D)}} = \frac{2}{6}\times\frac{6}{15} = \frac{2}{15}$.
- $\mathrm{P_1(4) = P(4|D) = P(D|4)\frac{P(4)}{P(D)}} = \frac{3}{6}\times\frac{6}{15} = \frac{3}{15}$.
- $\mathrm{P_1(5) = P(5|D) = P(D|5)\frac{P(5)}{P(D)}} = \frac{4}{6}\times\frac{6}{15} = \frac{4}{15}$.
- $\mathrm{P_1(6) = P(6|D) = P(D|6)\frac{P(6)}{P(D)}} = \frac{5}{6}\times\frac{6}{15} = \frac{5}{15}$.

Em suma, o que essa nova função de probabilidade $\mathrm{P_1(E)}$ faz é oferecer valores de probabilidade dadas todas as informações disponíveis. Até agora, a única informação disponível é a condicional D. Sem essa informação, as probabilidades de cada região eram as mesmas: $\frac{1}{6}$ (ou seja, havia a máxima incerteza possível sobre o local da marcação). Porém, com a informação, sabe-se pela função que há probabilidades diferenciadas, o que faz uma região da mesa mais provável ou menos provável de ter a marcação em comparação a outra (isto é, já não haveria máxima incerteza). É dessa forma que o usuário poderá ajustar gradualmente sua função de probabilidade de acordo com as informações obtidas em novos lançamentos e, com isso, ter "opiniões" mais objetivas sobre o local da marcação na mesa.

Retornando à simulação. Depois de o programa ter informado ao usuário o resultado do lançamento, ele pergunta ao usuário onde este acha que a marcação foi feita. Dada a função de probabilidade $\mathrm{P_1(E)}$, ainda que exista incerteza, já é possível fazer tentativas mais objetivas, não baseadas na completa ignorância, pois a função indica haver regiões mais prováveis que outras. Suponhamos que essa tentativa seja feita (ver Figura 7) para eliminar o outro extremo da mesa, a região 6, já que a sua probabilidade $\mathrm{P_1(6)} = \frac{5}{15}$ é a maior encontrada.

Figura 7: Tela que ilustra a primeira tentativa

```
AJUDANTE: Em qual regiao voce acha marcacao foi feita? 6
AJUDANTE: VOCE ERROU. A marcacao nao foi feita na regiao 6. Vamos tentar de novo?
```

Essa tentativa na região 6 foi feita pois era a região mais provável de a marcação ter sido feita segundo as informações que se tem até agora. Caso não haja acerto – segundo a Figura 7, foi o que realmente aconteceu –, o programa informará que não houve sucesso, e oferecerá oportunidade de lançar outra vez a bola para que o usuário obtenha mais informações. Aqui se inicia um ciclo que se repetirá até que o usuário acerte sua tentativa de qual região a marcação foi feita.

A marcação também não está na região 6. Isso serve como informação nova e, portanto, ela pode ser usada como restrição em uma nova função de probabilidade. Simbolizando essa restrição por $\overline{\{6\}}$ – isto é, a marcação não está na região 6 –, monta-se uma segunda função de probabilidade $\mathrm{F}_2(\mathrm{E}) = \mathrm{P}_1(\mathrm{E}|\overline{\{6\}})$, que assume os seguintes valores para cada evento elementar:

- $\mathrm{F}_2(1) = 0$.
- $\mathrm{F}_2(2) = \mathrm{P}_1(2|\overline{\{6\}}) = \mathrm{P}_1(\overline{\{6\}}|2)\frac{\mathrm{P}_1(2)}{\mathrm{P}_1(\overline{\{6\}})} = 1 \times \frac{\mathrm{P}_1(2)}{1-\mathrm{P}_1(6)} = \frac{\frac{1}{15}}{1-\frac{5}{15}} = \frac{\frac{1}{15}}{\frac{10}{15}} = \frac{1}{10}$.
- $\mathrm{F}_2(3) = \mathrm{P}_1(3|\overline{\{6\}}) = \mathrm{P}_1(\overline{\{6\}}|3)\frac{\mathrm{P}_1(3)}{\mathrm{P}_1(\overline{\{6\}})} = \frac{\frac{2}{15}}{\frac{10}{15}} = \frac{2}{10}$.
- $\mathrm{F}_2(4) = \mathrm{P}_1(4|\overline{\{6\}}) = \mathrm{P}_1(\overline{\{6\}}|4)\frac{\mathrm{P}_1(4)}{\mathrm{P}_1(\overline{\{6\}})} = \frac{\frac{3}{15}}{\frac{10}{15}} = \frac{3}{10}$.
- $\mathrm{F}_2(5) = \mathrm{P}_1(5|\overline{\{6\}}) = \mathrm{P}_1(\overline{\{6\}}|5)\frac{\mathrm{P}_1(5)}{\mathrm{P}_1(\overline{\{6\}})} = \frac{\frac{4}{15}}{\frac{10}{15}} = \frac{4}{10}$.
- $\mathrm{F}_2(6) = \mathrm{P}_1(6|\overline{\{6\}}) = 0$.

Dando prosseguimento ao ciclo, após o usuário atualizar sua função de probabilidade, ele poderá solicitar um novo lançamento. Esse lançamento foi feito conforme a Figura 8.

Figura 8: Tela que ilustra um lançamento sem informação nova

```
AJUDANTE: Quantas vezes devo lancar a bola? 1
AJUDANTE: A bola foi lancada 1 vez(es) agora. Em relacao a bola, a marcacao ficou
0 vez(es) a direita (0.0 % desses lancamentos)
0 vez(es) a esquerda (0.0 % desses lancamentos)

AJUDANTE: Total de 2 lancamentos, sendo:
1 a direita (50.0 % do total)
0 a esquerda (0.0 % do total)
```

Dessa vez, o programa informou que a marcação não está nem à direita nem à esquerda da bola lançada. Isto é, a bola caiu na mesma região da marcação. Por isso, neste lançamento, não há informação nova, e o usuário permanece com a mesma função de probabilidade e, com ela, precisa fazer uma nova tentativa. Considerando que ainda são possíveis 4 regiões onde a marcação foi feita – aquelas em que a probabilidade não é zero –, a melhor tentativa será aquela com a maior probabilidade, ou seja, a região 5. Essa tentativa aparece na Figura 9:

Figura 9: Tela que ilustra a segunda tentativa

```
AJUDANTE: Em qual regiao voce acha marcacao foi feita? 5
AJUDANTE: VOCE ERROU. A marcacao nao foi feita na regiao 5. Vamos tentar de novo?
```

O programa exibiu as mensagens da Figura 9. Não houve acerto, mas uma nova informação foi obtida: a impossibilidade de a marcação estar na região 5. A nova informação na forma de restrição $\overline{\{5\}}$ será incorporada em uma terceira função de probabilidade $\mathrm{F}_3(\mathrm{E}) = \mathrm{F}_2(\mathrm{E}|\overline{\{5\}})$, mais atualizada, cujos valores nos eventos elementares serão:

- $F_3(1) = F_3(5) = F_3(6) = 0$.
- $F_3(2) = F_2(2|\overline{\{5\}}) = F_2(\overline{\{5\}}|2)\frac{F_2(2)}{F_2(\overline{\{5\}})} = 1 \times \frac{F_2(2)}{1-F_2(5)} = \frac{\frac{1}{10}}{1-\frac{4}{10}} = \frac{\frac{1}{10}}{\frac{6}{10}} = \frac{1}{6}$.
- $F_3(3) = F_2(3|\overline{\{5\}}) = F_2(\overline{\{5\}}|3)\frac{F_2(3)}{F_2(\overline{\{5\}})} = \frac{F_2(3)}{F_2(\overline{\{5\}})} = \frac{\frac{2}{10}}{\frac{6}{10}} = \frac{2}{6}$.
- $F_3(4) = F_2(4|\overline{\{5\}}) = F_2(\overline{\{5\}}|4)\frac{F_2(4)}{F_2(\overline{\{5\}})} = \frac{F_2(4)}{F_2(\overline{\{5\}})} = \frac{\frac{3}{10}}{\frac{6}{10}} = \frac{3}{6}$.

Nesse momento, o programa oferece ao usuário uma nova oportunidade de lançamento, conforme a Figura 10.

Figura 10: Tela que ilustra o último lançamento

```
AJUDANTE: Quantas vezes devo lancar a bola? 1
AJUDANTE: A bola foi lancada 1 vez(es) agora. Em relacao a bola, a marcacao ficou
1 vez(es) a direita (100.0 % desses lancamentos)
0 vez(es) a esquerda (0.0 % desses lancamentos)

AJUDANTE: Total de 6 lancamentos, sendo:
2 a direita (66.7 % do total)
0 a esquerda (0.0 % do total)
```

Esse último lançamento indicou que a marcação ficou à sua direita. Assim, a informação total obtida até o momento – a marcação ficou à direita em 2 lançamentos – será simbolizada pela condicional D^2. Assim, monta-se uma nova função de probabilidade $P_2(E) = P(E|D^2)$ e, com ela, monta-se outra função $F_4(E)$ que incorpore as restrições que já temos conhecimento, de que a marcação não está na região 5 e nem na 6, isto é, $\overline{\{5,6\}}$.

Essa função F_4 incorporará a informação total até agora. Para conhecê-la, será necessário obter $P_2(E)$, de modo muito análogo a $P_1(E)$. Primeiro, fazemos $P(D^2)$:

$$
\begin{aligned}
P(D^2) &= P(D^2 \cap \{2\}) + P(D^2 \cap \{3\}) + P(D^2 \cap \{4\}) + P(D^2 \cap \{5\}) + P(D^2 \cap \{6\}) = \\
&= P(2)P(D^2|2) + P(3)P(D^2|3) + P(4)P(D^2|4) + P(5)P(D^2|5) + P(6)P(D^2|6) = \\
&= \frac{1}{6}\left(\frac{1}{36} + \frac{4}{36} + \frac{9}{36} + \frac{16}{36} + \frac{25}{36}\right) = \frac{55}{216}.
\end{aligned}
$$

Agora, fazemos $P_2(E) = P(E|D^2)$:

- $P_2(1) = P(1|D^2) = 0$.
- $P_2(2) = P(2|D^2) = P(D^2|2)\frac{P(2)}{P(D^2)} = \frac{1}{36} \times \frac{\frac{1}{6}}{\frac{55}{216}} = \frac{1}{36} \times \frac{36}{55} = \frac{1}{55}$.
- $P_2(3) = P(3|D^2) = P(D^2|3)\frac{P(3)}{P(D^2)} = \frac{4}{36} \times \frac{36}{55} = \frac{4}{55}$.
- $P_2(4) = P(4|D^2) = P(D^2|4)\frac{P(4)}{P(D^2)} = \frac{9}{36} \times \frac{36}{55} = \frac{9}{55}$.
- $P_2(5) = P(5|D^2) = P(D^2|5)\frac{P(5)}{P(D^2)} = \frac{16}{36} \times \frac{36}{55} = \frac{16}{55}$.

• $P_2(6) = P(6|D^2) = P(D^2|6)\frac{P(6)}{P(D^2)} = \frac{25}{36} \times \frac{36}{55} = \frac{25}{55}$.

Depois de tudo isso, é possível conhecer $F_4(E)$, que incorpora a restrição $\overline{\{5,6\}}$ a partir de $P_2(E)$. Isso significa que $F_4(E) = P_2(E|\overline{\{5,6\}})$:

• $F_4(1) = F_4(5) = F_4(6) = 0$.

• $F_4(2) = P_2(2|\overline{\{5,6\}}) = P_2(\overline{\{5,6\}}|2)\frac{P_2(2)}{P_2(\overline{\{5,6\}})} = 1 \times \frac{P_2(2)}{1-P_2(5)-P_2(6)} = \frac{\frac{1}{55}}{\frac{14}{55}} = \frac{1}{14}$.

• $F_4(3) = P_2(3|\overline{\{5,6\}}) = P_2(\overline{\{5,6\}}|3)\frac{P_2(3)}{P_2(\overline{\{5,6\}})} = \frac{\frac{4}{55}}{\frac{14}{55}} = \frac{4}{14}$.

• $F_4(4) = P_2(4|\overline{\{5,6\}}) = P_2(\overline{\{5,6\}}|4)\frac{P_2(4)}{P_2(\overline{\{5,6\}})} = \frac{\frac{9}{55}}{\frac{14}{55}} = \frac{9}{14}$.

Agora sim tem-se uma função de probabilidade completamente atualizada com a informação total disponível até então. Assim, o usuário está pronto para fazer outra tentativa objetiva, conforme a função indica. A região com maior probabilidade de ter a marcação é a 4, e é essa sua próxima tentativa, conforme a Figura 11:

Figura 11: Tela que ilustra a tentativa correta

```
AJUDANTE: Em qual regiao voce acha marcacao foi feita? 4
AJUDANTE: VOCE ACERTOU! A marcacao foi feita na regiao 4.
```

Neste momento, o programa encerra a rodada, já que o usuário acertou a tentativa, e volta ao ponto inicial, como que iniciando um novo experimento.

Usar a função de probabilidade como critério de inferência bayesiana pode servir para tomada de decisão, por exemplo. Um observador, que esteja em busca de aperfeiçoar sua opinião inicial sobre um determinado experimento, poderá usufrir do arcabouço teórico apresentado – em especial, da noção de probabilidade condicional (Equação 4), da Lei da Probabilidade Total (Equação 2) e do Teorema de Bayes (Equação 5) – para estabelecer um modelo probabilístico de o problema, obter uma função de probabilidade inicial e acrescentar informações gradualmente ao modelo para adaptar a função de probabilidade. Desse modo, sua opinião inicial sobre o problema evoluirá ao ponto de ele ser capaz de tomar decisões com confiança, sempre se baseando na função de probabilidade mais atualizada possível com informações relevantes ao problema.

6. Discussões

Qual foi a grande lição tirada dessa simulação? É que o uso do Teorema de Bayes foi essencial para atualizar a função de probabilidade e, com ela, aprender sobre eventos passados baseando-se em informações obtidas no presente. A opinião inicial foi formada na ausência de quaisquer informações sobre o problema, isto é, em completa ignorância sobre o que aconteceu no passado. À medida que informações novas e relevantes para o experimento passam a ser incorporadas – via Teorema de Bayes –, a opinião sobre o que tenha ocorrido no passado deixa de ser formulada por ignorância, mas sim com base em condições e restrições impostas por essas informações. A função de probabilidade resultante tende a oferecer valores de probabilidade cada vez mais objetivos.

Agradecimentos

Este trabalho representa uma parte da dissertação do autor principal defendida no PROFMAT da UNIFESP Campus Diadema/SP.

Referências

[1] Ministério da Educação do Brasil. *Base Nacional Comum Curricular*, 2018. Acessível em http://basenacionalcomum.mec.gov.br/images/BNCC_EI_EF_110518_versaofinal_site.pdf.

[2] Thomas Bayes. "An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S". Em: *Philosophical Transactions of the Royal Society of London* 53 (0) (1763), pp. 370–418. DOI: 10.1098/rstl.1763.0053.

[3] Bradley Efron. Bayes' Theorem in the 21st Century. *Science* 340, 1177-1178 (2013). DOI: 10.1126/science.1236536.

[4] A. N. Kolmogorov. *Foundations of the theory of probability*. 2ª ed. Chelsea Publishing Co., 1956.

[5] Sharon Bertsch McGrayne. *The theory that would not die: how Bayes' rule cracked the enigma code, hunted down Russian submarines, and emerged triumphant from two centuries of controversy*. Yale University Press, 2011.

Saulo Cavalcante dos Reis
Universidade Presbiteriana Mackenzie
<saulo.reis@mackenzie.br>

Renato de Sá Teles
Universidade Federal de São Paulo
<renato.teles@unifesp.br>

Anderson Augusto Ferreira
Universidade Federal de São Paulo
<ferreira07@unifesp.br>

Recebido: 09/11/2023
Publicado: 01/11/2024