
Universidade Federal de São Paulo

Instituto De Ciências Ambientais,

Químicas E Farmacêuticas

Campus Diadema



**Mestrado Profissional em Matemática
em Rede Nacional - PROFMAT**

**Teorema de Bayes: Uma Proposta para o
Ensino Médio**

Saulo Cavalcante dos Reis

Orientador: Prof. Dr. Renato de Sá Teles

Co-orientador: Prof. Dr. Anderson Augusto Ferreira

Diadema

Novembro, 2022



PROFMAT

Título: *Teorema de Bayes: Uma Proposta para o Ensino Médio*

Dissertação apresentada ao Instituto De Ciências Ambientais, Químicas E Farmacêuticas da UNIFESP, campus Diadema/SP, como parte dos requisitos exigidos para a obtenção do título de Mestre pelo Programa de Mestrado Profissional em Matemática em Rede Nacional – PROFMAT.

Diadema
Novembro, 2022

Dados Internacionais da Catalogação na Publicação (CIP)

Reis, Saulo C. dos

Teorema de Bayes: Uma Proposta para o Ensino Médio / Saulo C. dos Reis. -- Diadema, 2022.
63 f.

Dissertação (Mestrado Profissional em Matemática em Rede Nacional) - Universidade Federal de São Paulo - Campus Diadema, 2022.

Orientador: Renato de Sá Teles

Coorientador: Anderson Augusto Ferreira

1. Teorema de Bayes. 2. Ensino da Matemática. 3. História da Matemática. 4. Probabilidade. 5. Simulação Computacional. I. Título.

UNIVERSIDADE FEDERAL DE SÃO PAULO
Instituto De Ciências Ambientais, Químicas E Farmacêuticas
Mestrado Profissional em Matemática em Rede Nacional
PROFMAT

Chefe de departamento:

Prof. Dr. Renato Marcone José de Souza

Coordenador do Programa de Pós-Graduação:

Prof. Dr. Renato de Sá Teles

SAULO CAVALCANTE DOS REIS

TEOREMA DE BAYES: UMA PROPOSTA PARA O ENSINO MÉDIO

Presidente da banca: Prof. Dr. Renato de Sá Teles

Banca examinadora:

Prof. Dr. Evaldo Araujo De Oliveira Filho (UNIFESP)

Prof^ª Dr^ª Roseli Kunzel (UNIFESP)

Prof. Dr. Delmárcio Gomes da Silva (MACKENZIE)

Data da Defesa: 18 de novembro de 2022

“A sabedoria é melhor que a insensatez, assim como a luz é melhor que as trevas. O sábio vê para onde está indo, mas o tolo anda na escuridão.” (Ec 2:13–14, NVT)

AGRADECIMENTOS

Agradeço a Deus pela força, persistência e sabedoria concedidas a mim, pó em forma de homem. Agradeço à minha noiva, pelo carinho e dedicação constantes, em meio aos altos e baixos durante este tempo da pandemia da COVID-19. Agradeço à minha família, que mesmo de longe, sempre deram apoio aos meus estudos, desde criança. Agradeço à Universidade Presbiteriana Mackenzie e, em especial, à Escola de Engenharia, dentro da qual tenho crescido e conhecido pessoas espetaculares, seja pelo seu perfil acadêmico, seja pela sua postura diante dos desafios.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

RESUMO

Esta dissertação de mestrado objetiva apresentar uma proposta de ensino do Teorema de Bayes para alunos do ensino médio, tendo em vista o cumprimento de algumas competências a serem desenvolvidas nesses alunos segundo a BNCC. Antes de expor o teorema, uma história contemplando a origem e o desenvolvimento do teorema é apresentada de modo não-exaustivo, focando em personagens sem os quais o teorema não seria concebido como método sistemático de busca de conhecimento, ou sem os quais também não se poderia enxergá-lo como parte da matemática aplicada. Segue-se daí uma formalização teórica da probabilidade, fornecendo bases para os últimos dois capítulos. Primeiro, para que o experimento mental originário de Thomas Bayes seja corretamente compreendido e, segundo, mostrar duas propostas de ensino do teorema: facilitar a descoberta de informações faltantes ou desconhecidas na resolução de problemas da probabilidade, e também enquanto método de evitar vieses de julgamento probabilístico em situações da vida real.

Palavras-chave: 1. Teorema de Bayes. 2. Ensino da Matemática. 3. História da Matemática. 4. Probabilidade. 5. Simulação Computacional.

ABSTRACT

This masters' dissertation seeks to propose a way of teaching Bayes' Theorem for High School students, aiming to accomplish certain goals in the development in students previously suggested in BNCC. Before presenting the theorem, a history concerning the origin and development of the subject is presented in a non-exhaustive fashion, focusing on characters without whom the theorem would not be thought of as a systematic way of searching for knowledge, or without whom it would not be possible to see it as part of applied mathematics. Then, it is presented a formal theory of probability, thus providing the ground on which the last two chapters are founded. First, to make the mental experiment originating Bayes' theorem to be correctly grasped and, second, to show two proposals of teaching the theorem: to ease the discovery of missing or unknown information in problem solving, and also as a method to avoid biased probability judgements in real life situations.

Keywords: 1. Bayes' Theorem. 2. Teaching Mathematics. 3. History of Mathematics. 4. Probability Theory. 5. Computer Simulation.

SUMÁRIO

INTRODUÇÃO	3
1 ORIGEM E HISTÓRIA DO TEOREMA DE BAYES	6
1.1 O início de tudo: Thomas Bayes	6
1.2 A redescoberta: Pierre-Simon Laplace	8
1.3 Rejeição de Bayes na Europa	10
1.4 Uso de Bayes nos Estados Unidos no início do séc. XX	11
1.5 As oposições de Karl Pearson, Ronald Fisher e Jerzy Neyman	12
1.6 Ressurgimento discreto e a contribuição de Harold Jeffreys	14
1.7 Alan Turing, Claude Shannon e início da era da informação	15
1.8 Fundamentação moderna do bayesianismo	17
2 INTRODUÇÃO À TEORIA DA PROBABILIDADE	20
2.1 Axiomatização de Espaços Probabilísticos	20
2.2 Partição do Espaço Amostral e Probabilidade Condicional	23
2.3 Teorema de Bayes	27
3 SIMULAÇÃO COMPUTACIONAL DO EXPERIMENTO DE BAYES	31
3.1 Um lançamento por vez	33
3.2 Vários lançamentos por vez	40
3.3 Comparando paradigmas	42
4 PROPOSTAS PARA O ENSINO MÉDIO	44
4.1 Uso de Tabelas de Contingência	44
4.1.1 Observância da probabilidade do evento complementar	45
4.1.2 Observância da Lei da Probabilidade Total	46
4.1.3 Observância da Probabilidade Condicional	47
4.2 Diminuir vieses de julgamento probabilístico com fator de Bayes	50
4.2.1 Fator de Bayes em Tversky & Kahneman	51
4.2.2 Fator de Bayes em Gigerenzer et al.	53
4.3 Considerações finais	55
A CÓDIGO-FONTE DA SIMULAÇÃO	57
B SUGESTÃO DE PLANO DE AULA COM DADO DE 6 FACES	59
B.1 OBJETIVOS	59
B.2 CONTEÚDOS	59
B.3 RECURSOS	59

B.4 PROCEDIMENTOS	59
B.5 AVALIAÇÃO	61
BIBLIOGRAFIA	63

INTRODUÇÃO

Dentro da Base Nacional Comum Curricular (BNCC), a respeito da área da Matemática a ser trabalhada no Ensino Médio, propõe-se “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 mesma 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]. Levando-se em conta todo o debate sobre *fake news* dos últimos anos, torna-se ainda mais importante que os estudantes de hoje estejam mais equipados com ferramentas matemáticas que permitam com maior facilidade distinguir informações verdadeiras das falsas, e 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, então 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, que é entendida como a realização de afirmações sobre um conjunto de elementos considerados representativos de um universo, torna-se 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 maiores paradigmas: o paradigma frequentista e o paradigma bayesiano. Neste trabalho, propõe-se desenvolver algumas propostas de ensino da probabilidade e estatística no currículo do Ensino Médio segundo o paradigma bayesiano (e claro, do Teorema de Bayes), levando-se em conta todas as considerações previamente 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 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. Nesta monografia, há o entendimento que a inferência estatística – dentro do ramo da estatística inferencial – acrescenta uma ferramenta importante ao conteúdo do Ensino Médio ao propor o paradigma bayesiano com o Teorema de Bayes.

Ao mesmo tempo que o Teorema de Bayes representa um marco na história do raciocínio lógico e o primeiro grande triunfo da inferência estatística [2], ele pode causar problemas quando houver aplicação indevida, assim como toda fórmula aprendida na Matemática, mas com uma motivação especial: dada a possibilidade da sua aplicação quase universal nas ciências, o conhecimento prévio das probabilidades envolvidas precisa ser garantido para que haja menor margem de erro no cálculo final da probabilidade desejada. Isto é significativo, quando o conhecimento envolvido esteja relacionado a alguma decisão de impacto público.

Afinal de contas, o que é o Teorema de Bayes? Considerando que $p(A)$ e $p(B)$ como as probabilidades de eventos A e B, ambos realizados no espaço amostral (que é o universo de possibilidades), então denota-se por $p(A|B)$ a probabilidade do evento A dado que o evento B já tenha acontecido, e também denota-se por $p(B|A)$ a probabilidade do evento B dado que o evento A já tenha acontecido. Sendo assim, o teorema de Bayes traz a relação entre todas essas probabilidades:

$$p(A|B) = \frac{p(A) \times p(B|A)}{p(B)}$$

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.

ORIGEM E HISTÓRIA DO TEOREMA DE BAYES

Contemplar-se-á, no primeiro capítulo desta monografia, uma história do Teorema de Bayes, com a sua origem e seu desenvolvimento. A principal fonte para essa história é a obra de McGrayne [3]. Considerando o propósito deste trabalho de exibir a aplicabilidade quase universal do teorema nas ciências, o foco será dado a personagens sem os quais o teorema não seria concebido como método sistemático de busca de conhecimento, ou sem os quais também não se poderia enxergá-lo como parte da matemática aplicada. Também, seguindo outro propósito deste trabalho, este capítulo se encerrará com uma breve fundamentação do bayesianismo, estabelecida por pensadores mais modernos cuja resposta à problemática dos paradigmas da estatística coincide com a abordagem feita pelo autor desta monografia enquanto proposta de ensino do teorema para o ensino médio.

1.1 O INÍCIO DE TUDO: THOMAS BAYES

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 (1701—1761), que se baseou em definições e teoremas da probabilidade condicional no seu ensaio intitulado *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. Price escreveu a introdução ao ensaio, descrevendo as bases filosóficas sob as quais se assentavam o trabalho de Bayes. O 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”. 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 pré-estabelecidas, 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* [4]. Essa expressão sugere uma profunda aplicabilidade desse método para as ciências, 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” [5, p. 1, tradução livre].

O problema das probabilidades inversas era claro na mente de Bayes, e com isso, ele procurou aprender o máximo possível sobre eventos que ocorreram no passado para, com eles, buscar a chance de ocorrerem novamente no futuro. De início, ele concebeu que precisaria de um número, uma “crença inicial”, ou um “chute”, a partir do qual novas informações sobre o evento tornariam-no mais refinado e preciso. O processo que Bayes usou para refinar esse número consistiu num experimento mental, de enorme utilidade e que ilustra bem os usos análogos e mais modernos do seu teorema.

O experimento mental de Bayes começava com uma mesa plana e uma bola. A bola que fosse lançada na mesa teria a mesma chance de cair e parar em qualquer região da mesa, seja à esquerda ou à direita. Estando de costas para a mesa, ele lança a bola pela primeira vez, e pede a um ajudante que marque na mesa o local onde essa bola parou, para que ele descubra posteriormente sua provável localização. Depois de a marcação ter sido feita, então se inicia o processo de descoberta: o ajudante toma a bola e a lança novamente na mesa, reportando a Bayes se essa bola ficou à esquerda ou à direita da marcação inicial. A partir dessa resposta, Bayes reformula seu chute, imaginando a faixa de posições à esquerda ou à direita da mesa onde aquela marcação inicial teria sido feita. E o processo continua, com o ajudante jogando a bola na mesa e reportando a posição da bola como sendo à esquerda ou à direita da marcação, e a cada informação nova, a faixa de possíveis posições da marcação vai se estreitando, até que em algum momento, torna-se possível afirmar, com algum grau de confiança, em qual região da mesa a marcação inicial foi feita.

Conceitualmente falando, esse sistema 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 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 certa 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*.

$$p(A|B) = \frac{p(A) \times p(B|A)}{p(B)} \quad (1)$$

Nessa fórmula:

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

A história intelectual de Bayes oscilou entre a teologia e a matemática. A contar pelos anos em que o reverendo Bayes viveu, entre 1701 e 1761, pode-se afirmar que sua vida foi marcada por um período de tentativas da Inglaterra de se recompor, após dois séculos de contendas religiosas e guerra civil. Bayes era membro da Igreja Presbiteriana, a qual naquela época estava sob perseguição em função da sua falta de apoio à Igreja Anglicana, igreja oficial da Inglaterra, e como tal, Bayes foi considerado como não-conformista. Muitos matemáticos por conta disso eram barrados em universidades inglesas, e assim continuavam apenas como “amadores”, categoria na qual Bayes se encontrou por boa parte da sua vida, apesar de ter entrado na universidade. Em 1742 ele foi convidado a fazer parte da Royal Society, mas à época ela ainda não tinha o prestígio que hoje possui. Era uma organização privada que publicava e exibia trabalhos de intelectuais amadores — e que acabaram sendo responsáveis por boa parcela dos maiores avanços científicos da época. Mas, antes de tudo isso, Bayes estudou teologia e matemática na Universidade de Edimburgo, na Escócia, que era presbiteriana. Em 1711 ele foi para Londres onde seu próprio pai, que era sacerdote, o ordenou e o empregou como assistente eclesiástico.

Mais próximo aos seus trinta anos de idade, ele se envolveu numa discussão teológica que questionava se era possível conciliar a presença do mal do mundo e a benevolência de Deus. No seu panfleto publicado em 1731, Bayes declarou que “Deus dava às pessoas a maior felicidade que elas fossem capazes de ter”. Já aos seus quarenta anos, ele se envolveu na discussão da validade do cálculo inventado por Newton. Havia na época uma forte discussão sobre a validade do cálculo, embora isso possa parecer hoje aos nossos olhos um absurdo. Essa discussão talvez tenha sido inflamada por George Berkeley, um bispo anglicano irlandês, que havia publicado um panfleto atacando o cálculo newtoniano e o próprio Isaac Newton, os matemáticos não-conformistas, e todos os outros “livres pensadores” que acreditavam que a razão conseguiria iluminar qualquer assunto. Nesse sentido, Bayes entrou novamente na discussão, defendendo e explicando o cálculo de Newton. Essa veio a ser a sua única publicação em matemática durante sua vida [3, p. 20].

1.2 A REDESCOBERTA: PIERRE-SIMON LAPLACE

No entanto, Bayes não foi o único a desenvolver o método probabilístico. Uma forma modificada do método foi desenvolvida em 1774 independentemente por Pierre-Simon Laplace (1749—1827), que usou continuamente sua fórmula para distinguir hipóteses mais plausíveis das outras à medida que novos dados relevantes eram incorporados às suas ideias.

A interpretação bayesiana da probabilidade foi ampliada por Laplace, e é o seu modelo o mais usado na matemática e na ciência modernas. A motivação de Laplace teria vindo de uma ciência bastante distinta da probabilidade: a astronomia.

Naquele século XVIII, a astronomia era a ciência natural mais quantificada, e era com base nela que as recentes teorias da gravitação de Isaac Newton seriam testadas. A preocupação que havia a respeito da estabilidade dos corpos celestes, o que estava diretamente relacionado à interação que havia entre eles, somado aos movimentos das marés, tudo isso motivou os cientistas da época a buscarem explicações mais precisas sobre esses fenômenos. Não era por acaso que a astronomia era o assunto do dia, já que muito do que se acreditava a respeito do nosso planeta e dos corpos celestes havia mudado desde que Nicolau Copérnico revelou que a Terra não estava no centro do universo. Após ele, Kepler conseguiu uma explicação relativamente simples para o movimento e o período de corpos celestes e, não muito depois, Newton trouxe a ideia de “gravidade”. Embora houvesse poder persuasivo nos escritos de Newton, ainda havia dúvida sobre o que era a gravidade, se era uma mera hipótese ou alguma lei fundamental. Essa dúvida, associada à quantidade enorme de dados astronômicos que foram produzidos nessa época — graças ao desenvolvimento de telescópios mais precisos — parecia exigir uma forma diferente de avaliação das teorias.

Pierre-Simon Laplace, ao ter contato com a obra *Doctrine of Chances* de Abraham de Moivre, passou a se interessar muito por probabilidade. De algum modo, ele enxergava que a probabilidade poderia ajudar a descobrir as causas dos mais variados fenômenos astronômicos. A quantidade de dados existente até então revelava por vezes observações discrepantes ou até errôneas, e não parecia haver um bom método para descobrir quais dados eram mais prováveis de estarem corretos. Nesse sentido, Laplace começou a buscar uma teoria que o permitisse, matematicamente falando, usar dados sobre fenômenos conhecidos para achar as causas mais prováveis desses fenômenos.

Foi no ensaio chamado *Mémoire sur la Probabilité des Causes par les Évènements* [6] em que ele apresentou uma primeira versão do que hoje se chama de inferência bayesiana. Laplace entendia que o problema dos corpos celestes era muito complexo e que não seria possível esperar soluções precisas para ele; logo, a probabilidade ao menos poderia mostrar quais dos inúmeros dados astronômicos existentes até então era mais prováveis de estarem corretos ou relevantes para explicar algum fenômeno. Essa percepção de ignorância quanto a quais dados eram mais corretos ou não ainda assim era passível de estudo matemático, já que, para ele, a probabilidade era uma expressão matemática da nossa ignorância: “devemos à fragilidade da mente humana uma das mais delicadas e engenhosas teorias matemáticas, a saber, a ciência do acaso ou das probabilidades” [7] p. 114, tradução livre]. Mas a probabilidade para Laplace não parecia ser só quantificação de ignorância, mas também um tipo de bom senso instintivo justificado matematicamente: “a teoria das probabilidades é, no fundo, apenas o senso comum traduzido em cálculo; faz-nos apreciar

com exatidão o que mentes sadias sentem como uma espécie de instinto, muitas vezes sem poder dar uma razão disso” [8, p. 196, tradução livre].

Desde 1781, Laplace era dono do Teorema de Bayes em tudo: a fórmula, o método e sua utilização, menos no nome. Ele tornou comuns as estatísticas baseadas em probabilidades. E no meio das revoluções políticas na França do final do século XVIII e início do século XIX, Laplace ainda continuou seu trabalho científico. Ele anunciou o Teorema do Limite Central, que é provavelmente a maior descoberta da estatística em todos os tempos e que também causou um grande impacto no uso do Teorema de Bayes por várias décadas depois. A facilidade que o Teorema do Limite Central oferecia — de trabalhar com grandes quantidades de dados segundo o padrão da distribuição normal em forma de sino — tornou possível uma abordagem diferente aos problemas científicos. Daquele ano em diante, até sua morte em 1827, Laplace passou a usar uma abordagem que hoje se conhece como “frequentista”, não apenas por conta da descoberta do Teorema do Limite Central, mas também por que a enorme quantidade de dados sendo produzidos eram mais confiáveis do que até então. Acreditava-se que as duas abordagens — da probabilidade de causas pelo Teorema de Bayes e a frequentista — geravam os mesmos resultados com grandes volumes de dados. Ainda assim, em 1813, Laplace ainda descrevia sua abordagem de probabilidade de causas como seu principal método de pesquisa de causas desconhecidas ou complexas de fenômenos naturais [3, p. 56].

Apesar do endosso de Laplace à abordagem bayesiana, parecia-lhe mais fácil lidar com o grande volume de dados com sua nova abordagem frequentista. Por causa disso, muitos teóricos no futuro iriam justificar que a abordagem frequentista da probabilidade era melhor e, nesse sentido, a probabilidade de causas pelo Teorema de Bayes foi sendo esquecida. Esse esquecimento foi não intencional por uns, mas por outros foi bastante consciente e decisivo, por se supor que a abordagem bayesiana fosse subjetiva demais para a ciência nascente.

1.3 REJEIÇÃO DE BAYES NA EUROPA

Após a morte de Laplace, o Teorema de Bayes entrou em um período tumultuado. Mesmo em meio às lutas, o teorema prosseguiu, ajudando a resolver problemas práticos para o meio militar, para as comunicações, para a assistência social e a medicina nos Estados Unidos e na Europa.

Pode-se dizer que esse período de lutas começou com as frequentes tentativas de assassinato de reputação de Laplace. A probabilidade enquanto “uma expressão matemática da nossa ignorância” foi sendo fortemente rejeitada, e dando lugar àquela baseada em frequência, ou seja, no número de vezes em que um fenômeno foi observado. Os teóricos críticos argumentavam que as duas abordagens eram opostas, ainda que o próprio Laplace as considerasse equivalentes. Não demorou muito para que o equilíbrio sugerido por Laplace entre as duas abordagens fosse quebrado. Nas palavras de um dos mais famosos

críticos, John Stuart Mill, essa probabilidade de causas era “uma aberração do intelecto” e “ignorância [...] em forma de ciência” [9], pp. 34–35, tradução livre].

A ideia de quantificar a falta de conhecimento humano através da probabilidade de causas, originalmente entendida como um aspecto importante da nascente teoria da probabilidade com Thomas Bayes, Richard Price e Pierre-Simon Laplace foi esquecida. No entanto, muitas das críticas à abordagem bayesiana não foram feitas somente com critérios acadêmicos, mas também sob a forma de insultos pessoais. Por pelo menos um século, até a década de 20 do século XX, parecia haver uma condenação universal da pessoa de Laplace nas publicações acadêmicas. Um grande estatístico deste período, Karl Pearson, escreveu em 1929: “Tais declarações publicadas [...] sobre um dos homens mais ilustres [...] e totalmente infundadas por referências, são deploráveis em todos os sentidos” [10], p. 208, tradução livre].

Longe dos círculos acadêmicos, contudo, a probabilidade de causas parecia ser muito útil. Na França, um matemático chamado Joseph Louis François Bertrand (1822 – 1900) tornou-se uma referência no meio militar para cálculos de artilharia que precisavam lidar com incertezas. Em seus livros, Bertrand afirmava que a probabilidade de causas de Laplace era o único método válido para verificar uma hipótese com novas observações. Tanto as forças armadas francesas e russas fizeram uso dos livros de Bertrand para atingir seus objetivos em campo de batalha. Nesta mesma época, início do século XX, em que as tensões na Europa iam se acumulando e a Primeira Guerra Mundial se aproximava, um militar francês de nome Jean Baptiste Eugène Estienne (1860 – 1936) desenvolvia tabelas de probabilidade bayesianas para artilharia, além de métodos para testar munição. Tais métodos foram muito importantes para economizar materiais bélicos, já que era necessário muito desperdício sob os tradicionais métodos frequentistas [3], p. 66].

1.4 USO DE BAYES NOS ESTADOS UNIDOS NO INÍCIO DO SÉC. XX

Nos EUA, o que ocorria era uma industrialização rápida e, nela, também crescia a demanda por dados e por conhecimento para tomada de decisões. Bayes parecia ser uma boa ferramenta para responder a essa demanda.

Chama a atenção o trabalho de Edward Charles Dixon Molina (1877 – 1964), um funcionário do departamento de engenharia e pesquisa da AT&T (mais tarde chamado de Laboratórios Bell). O problema do sistema de telefonia Bell, que era propriedade da AT&T, era o de ficar regularmente sobrecarregado num determinado período de tempo do dia (mais especificamente, por volta do horário do almoço) e subutilizado no restante do dia. Trabalhando nesse problema, Molina percebeu que “na literatura sobre teoria da probabilidade, existe uma grande confusão porque muitas autoridades falharam em distinguir claramente entre o teorema original de Bayes da probabilidade inversa e a sua generalização subsequente feita por Laplace. O teorema generalizado abrange, ou reúne,

tanto os dados obtidos de uma série de observações quanto qualquer outra informação colateral existente em relação aos resultados observados” [11], pp. 95-96, tradução livre].

Tal declaração de Molina parece tornar claro seu conhecimento do desenvolvimento da probabilidade de causas e da confusão que existia na literatura da época sobre o assunto. Foi precisamente pelo uso dessa abordagem da probabilidade que ele conseguiu estabelecer um maior nível de automação dos sistemas telefônicos, levando em consideração conhecimentos anteriores sobre o uso do telefone. O sistema de telefonia Bell passou a depender bastante da probabilidade bayesiana e, apesar do sucesso de Molina em solucionar o problema do sistema, ele ainda teve dificuldades em publicar suas pesquisas. A AT&T considerou que muitos dos seus trabalhos eram segredos proprietários e, por isso, suas publicações circulariam apenas internamente na companhia.

Ainda nessa mesma época, nos EUA, a rápida industrialização fez aparecer problemas como acidentes, doenças ou invalidez em maior proporção entre a classe trabalhadora. Não havia qualquer sistema de seguro destinado a proteger os trabalhadores que, em função disso, começaram a se sindicalizar. Legislações exigindo que os empregadores assegurassem seus trabalhadores contra tais tipos de problemas ocupacionais foram surgindo em vários estados. Contudo, não havia ainda uma ciência que pudesse orientar as seguradoras no cálculo dos benefícios dos segurados. Essa ciência precisaria se basear em dados como taxas de acidentes nos mais diversos campos de trabalho, políticas trabalhistas detalhadas das companhias, e etc. No entanto, esses dados não existiam até então. Os benefícios dos assegurados precisaram ser inventados num período muito curto de tempo.

Graças ao trabalho de dois homens, Isaac Max Rubinow (1875 – 1936) e Albert Wurts Whitney (1870 – 1943), houve grandes avanços no desenvolvimento dessa ciência. Rubinow, economista e também médico da American Medical Association, conseguiu reunir 11 profissionais com inclinações científicas para fundar a Casualty Actuarial Society em 1914. Whitney, estatístico de Berkeley, vindo após Rubinow, estava familiarizado com os teoremas de Bayes e Laplace, e já imaginava que precisaria fazer uso deles. Ele entendeu que a experiência de todo um setor industrial ou econômico poderia ser usada como base para o histórico anterior e de uma companhia local para os novos dados, levando também em consideração a reação dos clientes locais dependendo do tamanho da companhia. Esse sistema bayesiano simplificado, criado por Whitney e chamado de Credibilidade (*Credibility*), foi o primeiro sistema de seguro social dos EUA, e continuou sendo usado por algumas décadas [3], p. 72].

1.5 AS OPOSIÇÕES DE KARL PEARSON, RONALD FISHER E JERZY NEYMAN

De volta ao mundo acadêmico, na Grã-Bretanha, dois grandes nomes da ciência da época, Karl Pearson e Ronald Aylmer Fisher, mesmo brigando entre si, tinham em comum algo que marcou todas as suas carreiras: eram ferozes opositores do bayesianismo. Junto com o peso dos seus nomes, a influência antibayesiana sobre a nascente ciência estatística

perdurou por muitas décadas. Assim como se fazia antes, o critério para justificar a rejeição a Bayes era a subjetividade. A visão de Pearson era que o uso prático do Teorema de Bayes era visto como temporário, já que ceticismo e desconfiança ainda eram posturas bastante comuns no meio acadêmico antes, durante e mesmo por um tempo depois da Segunda Guerra Mundial. Para ele, havia uma esperança de que algo melhor surgisse na ciência: “Se a ciência não puder medir o grau de probabilidade envolvido [na previsão da experiência passada para a futura], pior para a ciência. O homem prático apegar-se-á aos seus métodos apreciativos até que ela [a ciência] possa, ou então aceitará os resultados da probabilidade inversa da marca Bayes/Laplace até que [métodos] melhores apareçam” [12], p. 3, tradução livre]. Aparentemente, Pearson buscava um método alternativo de fazer probabilidades inversas, um que não fosse subjetivo. Contudo, ele não via essa alternativa, mesmo entre seus pares [13], pp. 203–204].

Ronald Fisher, da Universidade de Cambridge, formulou uma teoria abrangente e rigorosa da estatística e a colocou em favor da sua agenda antibayesiana. Em 1925, ele publicou sua *magnum opus* intitulada *Statistical Methods for Research Workers* que foi um marco desse período antibayesiano. A ciência experimental, que não fazia uso de métodos estatísticos, passou a incorporar os novos métodos estatísticos de Fisher. Expressões comuns hoje como análise de variância, teste de significância, valores p foram criações dele, e é praticamente impossível que se faça qualquer trabalho estatístico ainda hoje sem usar terminologia inventada por Fisher. Em *Statistical Methods*, ele deixa claro o que é a ciência estatística, e nela, o seu posicionamento quanto ao bayesianismo: “será suficiente neste esboço geral do escopo da Ciência Estatística reafirmar minha convicção pessoal, que tenho defendido em outros lugares, de que a teoria da probabilidade inversa se baseia em um erro e deve ser totalmente rejeitada” [14], p. 9, tradução livre]. Boa parte do mundo acadêmico, por muitos anos, não precisou lidar com questionamentos bayesianos devido à influência do pensamento e da obra de Fisher.

Um dos filhos de Karl Pearson, Egon Pearson, também contribuiu para a manutenção do sentimento antibayesiano do seu tempo. Em 1933, ele se juntou a Fisher e a um matemático polonês, Jerzy Neyman, para desenvolver o teste de hipóteses hoje chamado de Neyman-Pearson, uma teoria que acabou por ser uma das mais influentes da ciência estatística do século XX. Contudo, o antibayesianismo de Neyman se resumia a quando as probabilidades a priori eram desconhecidas. Nesse caso, ele julgava que o uso de probabilidades uniformes era ilegítimo [15], p. 286]. Essa postura não mudou quando, não muito antes da Segunda Guerra Mundial, ele se mudou para a Universidade de Berkeley nos EUA. Neyman fez do seu teste de hipóteses Neyman-Pearson uma espécie de mascote dessa universidade, além de torná-la um reduto antibayesiano.

1.6 RESSURGIMENTO DISCRETO E A CONTRIBUIÇÃO DE HAROLD JEFFREYS

Durante as décadas de 1920 e 1930, os posicionamentos antibayesianos de Fisher, Egon Pearson e Neyman atraíam toda a atenção. Mesmo em meio a essa forte corrente de oposição, houve também pensadores que resgataram – ou reformularam – a abordagem subjetiva da probabilidade.

Emile Borel, matemático francês, afirmou em 1924 que as crenças subjetivas pudessem ser quantificadas pelo uso de probabilidade e, de modo prático, verificáveis através de apostas. Dois anos depois, um matemático e filósofo britânico de nome Frank P. Ramsey afirmou algo muito semelhante. Um de seus maiores interesses era em como tomar decisões diante de incerteza. Na Universidade de Cambridge, Ramsey sugeriu que a probabilidade fosse baseada em crenças pessoais que poderiam ser quantificadas por alguma aposta. Ele ainda introduziu uma função de utilidade e de maximização da utilidade esperada falando sobre medida de crença como base para algum tipo de ação, ou seja, Ramsey usou a probabilidade como método para decisão e comportamento, algo que nem mesmo Bayes ou Laplace fizeram. Além de Borel e Ramsey, Bruno de Finetti, considerado o melhor matemático italiano do século XX, é creditado por colocar a subjetividade de Bayes em uma base matemática sólida. De Finetti também considerava que crenças subjetivas poderiam ser quantificadas.

Mas foi Harold Jeffreys, professor de astronomia da Universidade de Cambridge, que conseguiu manter viva a chama bayesiana nas primeiras décadas do séc. XX em meio aos ataques antibayesianos. Pode-se atribuir à personalidade calma e cavalheiresca de Jeffreys a razão de ele e Ronald Fisher se tornarem amigos, ainda que discordassem fortemente sobre Bayes.

Jeffreys estudava a formação da Terra e dos planetas para entender a origem do sistema solar. Havia nesse campo da ciência uma enorme quantidade de problemas que poderiam ser abordados através da probabilidade inversa de causas. Sua pesquisa em geofísica e astrofísica era ampla e reunia evidências de diversos outros campos que, quando combinadas, eram sujeitas a relações matemáticas e então extrapoladas para buscar explicar algum fenômeno geológico, como por exemplo, determinar o possível epicentro de um terremoto, ou uma possível composição química do interior do planeta. Tais inferências não eram feitas com absoluta certeza, mas sim com graus de confiança que eram modificados para se adaptar a novas informações, que representa o *modus operandi* bayesiano clássico. Isso fica ainda mais evidente quando Jeffreys escreve, numa revisão de um livro de física: “são justamente as proposições duvidosas que constituem a parte mais interessante da ciência; todo avanço científico começa numa transição da ignorância completa, passando por um estágio de conhecimento parcial baseado em evidências, tornando-se gradualmente mais conclusivo, para um estágio de certeza prática” [16], p. 1021, tradução livre]. Curiosamente, Jeffreys também era muito interessado em histórias de detetive e fazia muitas anotações sobre os alibis e as motivações dos personagens para explicar seus comporta-

mentos. O modo bayesiano de pensar não se restringia apenas às ciências naturais para ele, aparentemente. A probabilidade inversa de causas de Laplace estava se tornando algo útil nas mãos de Jeffreys e, nisso, ele acabou sendo considerado o fundador da estatística Bayesiana moderna, sendo o primeiro após Laplace a aplicar a teoria bayesiana a uma série de problemas científicos [3, p. 89].

Dada a proximidade entre Jeffreys e Fisher e seus posicionamentos opostos quanto ao bayesianismo, foi iniciado um debate que durou dois anos na Royal Society. Eles eram os principais estatísticos do mundo, e cada um usou os métodos mais adequados ao seu campo. No entanto, nenhum deles conseguia ver o ponto de vista do outro. Logo depois do debate, Jeffreys escreveu sua maior obra, *Theory of Probability*, que foi por muitos anos a única explicação sistemática de como aplicar Bayes a problemas científicos. Contudo, o debate Fisher-Jeffreys terminou de forma inconclusiva mas, na prática, Jeffreys perdeu. O frequentismo prevaleceu e quase fez desaparecer Bayes. Ao fim da década de 1930, próximo ao início da Segunda Guerra Mundial, a probabilidade inversa das causas parecia estar morta para os estatísticos, pois a maioria deles estava usando as ideias desenvolvidas pelo trio antibayesiano Pearson-Fisher-Neyman.

1.7 ALAN TURING, CLAUDE SHANNON E INÍCIO DA ERA DA INFORMAÇÃO

Na década de 1930, ao menos entre os estatísticos em campo, o Teorema de Bayes estava morto. Eram anos em que se aproximava a Segunda Guerra Mundial, e os líderes das nações em conflito começaram a se questionar como tomar decisões importantes em meio a tantas incertezas e informações incompletas. Trabalhando de modo secreto em seus países, grandes mentes desse período histórico (e por que não, do século) repensaram a importância do bayesianismo para responder a essa questão.

Especialistas em probabilidade eram raros, e os grandes nomes da estatística até então que poderiam contribuir na guerra em favor dos Aliados, isto é, Jeffreys, Fisher e Neyman, não foram considerados. Analisar informações de guerra acabou sendo um trabalho sob responsabilidade de físicos, biólogos e matemáticos puros, entre os quais não predominava qualquer conhecimento sob o *status* científico do bayesianismo. No Reino Unido, a Government Code and Cypher School (GC&CS, a agência britânica de inteligência para codificação e decodificação de mensagens) recrutou em secreto alguns intelectuais para trabalhar na decodificação de mensagens secretas dos países do Eixo. Dentre esses intelectuais, havia um chamado Alan Mathison Turing, que viria a ser o protagonista da inteligência britânica de guerra, além também de ser considerado o pai da computação moderna. À época, em setembro de 1939, ele tinha 27 anos. Após alguns meses de trabalho, ele já conseguia decifrar mensagens secretas da aeronáutica alemã, transformando a GC&CS em uma máquina de decifragem de mensagens inimigas. Os alemães desenvolveram um dispositivo para codificar mensagens intitulado Enigma e, para decodificar mensagens das Enigmas, Turing desenvolveu um mecanismo eletromecânico chamado de

“Bombe”, com o qual se poderia testar com enorme rapidez os mecanismos internos de uma Enigma, para que se compreendesse seu processo codificador interno. Era mais fácil testar combinações e eliminar as que fossem erradas, do que achar as mais prováveis. Por isso, as Bombes de Turing testavam combinações que *não poderiam* gerar as mensagens.

Houve sucesso ao desvendar códigos da aeronáutica alemã, mas ninguém ainda havia tentado decodificar os códigos da marinha alemã. A marinha nazista operava as Enigmas mais complexas de todas as forças armadas do Eixo. Sabendo que ninguém estava buscando decodificar os códigos dessas Enigmas mais avançadas, Turing decidiu trabalhar nisso. Ele iniciou tentando reduzir o número de testes que uma Bombe precisava realizar. Através de um processo manual que ele apelidou de *Bamburismus*, Turing poderia atribuir uma probabilidade a uma sequência de letras escolhida por ele, mensurar sua confiança na validade delas por métodos bayesianos adicionando novas informações à medida que apareciam, e desse modo atualizar e refinar a probabilidade da sequência de letras. Nisso, ele conseguiria reduzir drasticamente a quantidade de testes para cada Bombe e reduzir o tempo de decodificação das mensagens. Em época de guerra, tempo economizado significava mais decisões acertadas por parte dos líderes e, potencialmente, mais vidas salvas.

Entende-se que uma das maiores contribuições de Turing para a defesa britânica tenha sido o uso de uma unidade de medida de confiança, que foi chamada de “ban”. Turing descobriu, enquanto formalizava seu processo de atualização das probabilidades, que não seria possível sistematizar suas comparações sem uma métrica definida. Segundo Irving John Good, matemático e criptanalista assistente de Turing, a unidade ban foi definida como sendo “a menor mudança no valor de uma evidência que seja perceptível à intuição humana” [17, p. 394]. Turing usava essa unidade de medida de confiança para estimar quanto de *informação* seria necessária para resolver algum tipo de problema. Contrariamente ao processo frequencista de decidir quantas *observações* precisariam ser feitas, Turing estimava o quanto de evidência era necessária para considerar um problema resolvido.

O desafio de descobrir as configurações internas de uma Enigma que codificava mensagens é considerado tipicamente um problema clássico da probabilidade inversa de causas. No entanto, o sistema de Turing era reconhecidamente britânico, sem traços aparentes que apontem o sistema ter sido inspirado nos sistemas originais de Bayes ou de Laplace. Contudo, o fato de Turing ter estudado matemática pura e não estatística parece ter assegurado que ele não fosse afetado por atitudes antibayesianas. Turing, quando questionado pelo seu assistente Good se o processo *Bamburismus* era bayesiano, confirmou a ideia, e isso sugeriu que Turing também já conhecia o Teorema de Bayes. É possível que, na GC&CS, só Turing e Good tivessem percebido que o *Bamburismus* era um processo bayesiano [3, p. 105].

Turing, ao viajar para os EUA em missão de cooperação, trabalhou com Claude Shannon nos Laboratórios Bell. Shannon também estava usando o Teorema de Bayes em seus

projetos de guerra. Ambos tinham um grande interesse por criptografia, mas Shannon estava à época trabalhando em sua teoria da comunicação e informação e suas possíveis aplicações para a criptografia. Shannon descobriu que tanto mensagens criptografadas como também os ruídos das linhas de telefone poderiam ser analisadas pelo mesmo tipo de matemática. Nesse sentido, a teoria da informação e a teoria da criptografia tinham problemas similares, contudo, o propósito das teorias era reverso: enquanto a primeira buscava diminuir a incerteza da informação, a segunda busca o aumento dessa incerteza. Em ambas, Shannon adotou a abordagem bayesiana. Com todas essas descobertas, Shannon também criou sua unidade de medida de informação, chamada de “bit” [18, p. 380].

Com o uso das métricas de informação, a abordagem bayesiana permitia analisar se algo foi, ou pode ser, aprendido ou não: bastava quantificar e comparar os níveis de informação dos termos *a priori* e *a posteriori* no Teorema de Bayes. Se o termo posterior contiver uma quantidade de informação muito diferente daquela do termo anterior, então algo foi aprendido sobre o sistema em estudo. Mas se as quantidades de informação de ambos dos termos forem as mesmas, então nada pode ser aprendido. Assim, também se inspirando na teoria da probabilidade de Kolmogorov, Shannon chegou à conclusão de que num sistema secreto ideal, a diferença da quantidade de informação entre o termos bayesianos *a priori* e *a posteriori* é exatamente zero. Tal conclusão foi inicialmente confidencial, especialmente logo após o fim da Segunda Guerra Mundial e o início da Guerra Fria. Contudo, ainda hoje se desenvolvem extensões das descobertas de Shannon pelo uso de técnicas bayesianas, tal foi o impacto dessa abordagem e sua aplicabilidade [3, p. 116].

1.8 FUNDAMENTAÇÃO MODERNA DO BAYESIANISMO

Pouco tempo depois que a Alemanha se rendeu na Segunda Guerra Mundial, o primeiro ministro britânico Winston Churchill deu ordens para que quaisquer evidências dos esforços de decodificação de mensagens secretas que levaram à vitória dos Aliados na guerra fossem destruídas. Isso significava que praticamente todos os dispositivos eletromecânicos construídos (isto é, as Bombes e todos os outros decorrentes de melhorias e aperfeiçoamentos), além de todo conhecimento produzido no período, precisavam ser destruídos.

As implicações diretas de todo esse material para a vitória britânica na Segunda Guerra Mundial não eram conhecidas até 1973, quando começaram a aparecer alguns registros do arcabouço teórico produzido por Turing, incluindo o seu bayesianismo. Aquela decisão política distorceu a percepção do valor da criptanálise para a Guerra Fria que se iniciava. Ainda assim, com o aval de Turing, seu antigo assistente Irving John Good desenvolveu teorias e métodos bayesianos de criptanálise, se tornando um dos reavivadores do bayesianismo no período pós-guerra. Contudo, fora dos círculos da criptografia, poucos conheciam o Teorema de Bayes, e a percepção acadêmica sobre ele permaneceu a mesma

dos anos anteriores à guerra. As teorias de Ramsey, De Finetti e Jeffreys permaneciam desconhecidas. [3, p. 130]

Para aqueles que mantinham simpatia à abordagem bayesiana, o frequentismo parecia uma espécie de Frankenstein de postulados, testes e procedimentos, sem qualquer unidade conceitual oriunda naturalmente das definições básicas da probabilidade. Com o crescimento do número de profissionais da estatística na década de 1950, cresceu também o número de publicações e conferências, dentro dos quais, pouco a pouco, algumas atenções se voltaram para o Teorema de Bayes. A primeira publicação que inaugurou o reavivamento do bayesianismo foi autorada por Good, intitulada *Probability and the Weighing of Evidence* e publicada em 1950.

Quando Good proferiu sua palestra na conferência da Royal Statistical Society, Dennis Lindley, matemático de Cambridge, estava presente. Só no ano de 1994, ao se recordar daquela palestra, Lindley diz a respeito de Good que “ele não tinha transmitido bem suas ideias a nós. Nós deveríamos ter prestado muito mais atenção ao que ele dizia pois ele estava muito à frente de nós em muitos pontos” [19, p. 312, tradução livre]. Este mesmo Lindley, retornando a Cambridge após a guerra, tinha interesse na estatística em geral e, de modo específico, em transformá-la em um ramo respeitável da matemática, com um corpo completo de pensamento baseado em axiomas e teoremas demonstrados à mesma maneira que outras ciências já estabelecidas [19, p. 308]. O esforço de Lindley resultou na sua importante publicação de 1953, *Statistical Inference* [20].

Além de Lindley, outro intelectual que se converteu ao bayesianismo foi Jimmie Savage, da Universidade de Chicago. Pela descoberta dos trabalhos de Borel, Ramsey e de De Finetti, ele se tornou um dos mais ardentes defensores da abordagem bayesiana, não apenas como um ramo da estatística, mas sim como um olhar sobre o todo da estatística. Repensando as ideias de seus influenciadores e adicionando as suas próprias, para ele, a probabilidade subjetiva era uma métrica de confiança. Opiniões subjetivas e experiência profissional nos campos da ciência, do direito, da engenharia ou quaisquer outros precisavam ser quantificados e incorporados em análises estatísticas [3, p. 150]. Adicionando aos mais recentes trabalhos de Good e Lindley, Savage produziu sua *Foundations of Statistics* em 1954. Coube a ele fundamentar logicamente a abordagem bayesiana da probabilidade.

Como esses teóricos responderam à principal objeção frequentista da probabilidade, que era a da objetividade científica? Se diferentes pesquisadores trazem olhares e opiniões diferentes uns dos outros à análise de dados, como é possível garantir objetividade? Para Savage, a quantidade de dados inicial sobre um fenômeno em estudo é baixa, o que significaria grande incerteza sobre os possíveis resultados desse fenômeno, em função dos pressupostos subjetivos que cada pesquisador traz consigo. À medida que essa quantidade de dados aumenta, os pesquisadores tendem a convergir em favor de alguma opinião, da mesma maneira que se imagina idealmente quando cientistas em campo concordam quando evidências relevantes se acumulam em favor de certa teoria ou de alguma combinação de teorias existentes. Quando se tem poucos dados, todos são subjetivistas ainda

que não tenham consciência disso e, por isso, discordam entre si. Contudo, conforme novas informações são incorporadas e compartilhadas igualmente entre os pesquisadores, a tendência deles é de concordância e nisso, gradualmente se convertem em objetivistas. Lindley concorda: é assim que a ciência é feita [3, p. 150]. A subjetividade da abordagem bayesiana não está na abordagem em si, mas é tão somente uma fase da investigação científica: é a fase inicial, quando necessariamente há pouca informação e não se espera que essa informação esteja compartilhada.

INTRODUÇÃO À TEORIA DA PROBABILIDADE

É necessário fazer uma axiomatização da teoria da probabilidade, estabelecendo definições que sejam relevantes ao estudo posterior de probabilidades condicionais, a qual leva ao Teorema de Bayes. Essa axiomatização significa que a teoria de probabilidade aqui exposta será desenvolvida exclusivamente a partir de definições básicas e de relações entre elas, com inspiração nas teorias de Kolmogorov ([21]) e Carnap ([22]). Essas definições também serão feitas já considerando-se sua possível aplicação às ciências. Por isso, é possível que exista mais de uma interpretação concreta das definições abstratas aqui expostas, conforme a ciência à qual se aplicam.

2.1 AXIOMATIZAÇÃO DE ESPAÇOS PROBABILÍSTICOS

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 que e_1, e_2, e_3, \dots sejam elementos que representem de modo único os possíveis resultados de um experimento, sem ambiguidade ou multiplicidade na representação, então chama-se de espaço amostral ao conjunto de todos esses elementos, que será denotado por $\Omega = \{e_1, e_2, e_3, \dots\}$.*

O espaço amostral é, na prática, o universo de possibilidades a serem cogitadas como resultados de um experimento, realizado sob condições bem conhecidas e estabelecidas. Representa o contexto geral de estudo. Esse conjunto precisa ser elaborado de modo que exista um único elemento desse conjunto a ser clara e explicitamente mapeado ao resultado do experimento após a realização deste, não podendo haver ambiguidade ou multiplicidade de elementos nesse mapeamento. As condições de realização do experimento, uma vez estabelecidas, não podem ser alteradas, uma vez que a mudança dessas condições implicaria também na potencial mudança do próprio espaço amostral, ou seja, dos possíveis resultados do experimento.

Definição 3 (Evento). *Um evento E é algum subconjunto do espaço amostral Ω , ou seja, pode ser qualquer conjunto E em que $E \subset \Omega$. É um possível resultado do experimento definido para Ω .*

Tal definição de evento é, por assim dizer, uma representação matemática de algo que pode ou não acontecer, tendo-se realizado o experimento definido para Ω . Se o resultado do experimento puder ser mapeado para algum dos elementos que pertença ao evento, então diz-se que aquele evento *aconteceu*, ou que houve *sucesso* do evento. Caso contrário,

se o resultado do experimento for mapeado para um elemento que não pertença ao evento, então se diz que aquele evento *não aconteceu* ou que houve *falha* do evento.

Definição 4 (Evento Elementar). *Classifica-se um evento E como sendo elementar caso ele seja um conjunto unitário.*

O evento elementar é, na prática, a representação singular de um único resultado possível do experimento. Exemplos de eventos elementares são $\{e_1\}$, $\{e_2\}$ ou $\{e_5\}$. Eventos que não são elementares são quaisquer outros subconjuntos do espaço amostral que não sejam unitários, como por exemplo, $\{e_1, e_2\}$, $\{e_3, e_4, e_5\}$, ou também $\emptyset = \{\}$ ou mesmo o próprio $\Omega = \{e_1, e_2, e_3, \dots\}$.

Definição 5 (Espaço de Eventos). *O espaço de eventos S é o conjunto de todos os possíveis eventos em Ω ; isto é, o conjunto de todos os subconjuntos de Ω .*

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, ou seja,*

$$P(\{e_i\}) \geq 0, \forall i \in \mathbb{N}.$$

2. *Para $\forall E \in S$ em que $\#(E) > 1$, ou seja, $E = \{e_{k_1}, e_{k_2}, \dots\}$, 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 sempre um, ou seja,*

$$P(\Omega) = P\left(\bigcup_{i=1}^{\infty} \{e_i\}\right) = \sum_{e \in \Omega} P(\{e\}) = 1.$$

A função de probabilidade é a função que associa um evento a um valor que será chamado de *probabilidade*. Essa função – mais especificamente, a lei da função – precisa ser concebida pelo observador de modo a satisfazer às três condições mencionadas acima. Nesse sentido, os valores das probabilidades dos eventos são arbitrários, podendo obedecer a critérios que satisfaçam ao observador segundo algum propósito experimental. Um exemplo de critério que é costumeiramente 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 .*

Com a definição acima de espaço probabilístico, que faz uso de todas as outras definições previamente feitas, tem-se aí uma axiomatização da teoria da probabilidade e que poderá ser utilizada para a modelagem de experimentos. Vejamos alguns exemplos.

Exemplo 1 (Cara e Coroa). *Lança-se uma única vez uma moeda com duas faces e observa-se a face superior.*

Um dos experimentos mais simples do mundo real, no qual se possa modelar um espaço probabilístico é o de lançar uma única vez uma moeda com duas faces: cara e coroa. Para estabelecer o espaço probabilístico desse experimento, é necessário conceber seus dois ingredientes: um conjunto Ω e uma função $P(E)$, segundo a Definição 7. Vejamos:

- Segundo a Definição 2, os possíveis resultados desse experimento são as possíveis faces que aparecem no lado superior da moeda após o lançamento, ou seja, são as faces Cara (K) ou Coroa (C). Matematicamente falando, tem-se $\Omega = \{K, C\}$, com $\#(\Omega) = 2$.
- Segundo a Definição 6, a função de probabilidade $P(E)$ a ser concebida precisa ter uma lei que satisfaça as condições mencionadas, um domínio S (segundo a Definição 5) e um contradomínio \mathbb{R} . Se seguirmos o Princípio da Indiferença para atribuir os valores de probabilidade aos eventos elementares de Ω – assumindo que se trata de uma moeda honesta –, pode-se conceber que $P(\{K\}) = P(\{C\}) = \frac{1}{\#(\Omega)} = \frac{1}{2}$.

Vamos garantir que as três condições para o estabelecimento da lei da função de probabilidade estejam sendo satisfeitas com esses valores. Dado que $\Omega = \{K, C\}$, então todos os eventos elementares são $\{K\}$ e $\{C\}$, com os quais as condições serão verificadas:

1. A probabilidade de qualquer um dos eventos elementares é realmente maior ou igual a zero, já que $P(\{K\}) = P(\{C\}) = \frac{1}{2}$. A condição 1 foi satisfeita.
2. O evento $E = \{K, C\}$, que é o único evento concebível em Ω que tenha mais de um elemento, é idêntico ao próprio Ω , ou seja, é um evento que tem probabilidade 1 e, por isso, é garantido que vá acontecer. Logo, $P(E) = 1$. Observe também que, se somarmos as probabilidades dos eventos elementares que compõem o evento E , chegamos a $P(\{K\}) + P(\{C\}) = \frac{1}{2} + \frac{1}{2} = 1$.

Portanto, $P(E) = P(\{K\}) + P(\{C\})$, e a condição 2 também foi satisfeita.

3. No item anterior, vimos que a soma das probabilidades de todos os eventos elementares resulta em $P(\{K\}) + P(\{C\}) = 1$, e portanto, a condição 3 é satisfeita também.

Uma observação importante a ser feita sobre essa concepção da função de probabilidade é que, ao adotar o critério do Princípio da Indiferença para atribuir os valores de probabilidade aos eventos elementares $\{K\}$ e $\{C\}$, houve um pressuposto experimental de que a moeda seja honesta. Esse pressuposto é, como se percebe, bastante razoável e intuitivo, pois é o que costumeiramente se espera de um sorteio realizado com moedas.

Exemplo 2 (Dado de 6 faces). *Lança-se uma única vez um dado com 6 faces, cada face numerada com um número de 1 a 6, e observa-se a face superior.*

Vejamos como este experimento pode ser modelado probabilisticamente segundo a Definição 7.

- Segundo a Definição 2, os possíveis resultados desse experimento são os números naturais de 1 a 6, ou seja, $\Omega = \{1, 2, 3, 4, 5, 6\}$, com $\#(\Omega) = 6$.
- Segundo a Definição 6 e seguindo o Princípio da Indiferença para atribuir as probabilidades aos eventos elementares de Ω – assumindo que se trata de um dado honesto –, podemos conceber $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{\#(\Omega)} = \frac{1}{6}$.

Vejamos se as três condições para a função probabilística estabelecidas na Definição 6 são satisfeitas:

1. A probabilidade de qualquer um dos eventos elementares é $P(\{e\}) = \frac{1}{6}$ e, portanto, é sempre maior que zero.
2. Para qualquer evento E , tem-se sempre

$$P(E) = \sum_{e \in E} P(\{e\}) = \sum_{i=1}^{\#(E)} \frac{1}{6} = \frac{\#(E)}{6}.$$

que é a soma das probabilidades de todos os eventos elementares que constituem o evento E .

3. Na condição anterior, usando o próprio espaço amostral como exemplo, tem-se $E = \Omega$, de modo que

$$P(E) = P(\Omega) = \frac{\#(\Omega)}{6} = \frac{6}{6} = 1.$$

2.2 PARTIÇÃO DO ESPAÇO AMOSTRAL E PROBABILIDADE CONDICIONAL

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

1. $A_i \in \mathcal{S}, \forall i \in \mathbb{N}$.
2. $A_i \cap A_j = \emptyset, \forall i, j \in \mathbb{N}$ e $i \neq j$, ou seja, os eventos são dois a dois mutuamente exclusivos em Ω .
3. $\bigcup_{i=1}^n A_i = \Omega$, ou seja, a união de todos os eventos resulta no próprio espaço amostral.

Neste ponto, há uma consideração importante a ser feita. A noção matemática de partição traz à tona a discussão da classificação dos elementos que pertencem ao espaço amostral. Matematicamente falando, uma partição do espaço amostral Ω não é única – ou seja, não é determinada unicamente pelo espaço amostral –, podendo haver muitas delas. A possibilidade de existência de uma partição específica \mathcal{P} dependerá do *critério* de classificação dos elementos pertencentes a Ω . A fonte de informação desse critério é sempre “exterior” ao espaço amostral e, com esse critério, deverá ser sempre possível encaixar cada um dos elementos de Ω a um único evento A_i que faz parte daquela partição \mathcal{P} . Em resumo, se $e \in \Omega$, então $\exists! n \in \mathbb{N} \mid e \in A_n$.

É muito comum em problemas da probabilidade lidar com partições do espaço amostral a serem identificadas depois de uma devida compreensão do contexto. Por isso, será necessário reconhecer em cada problema quais seriam as informações que garantem a existência de um espaço probabilístico (e por conseguinte, de um espaço amostral Ω) e quais informações estabelecem critérios para identificação de partições desse Ω . Nesse sentido, alguns questionamentos costumam ser muito úteis para se compreender a natureza do problema a ser solucionado:

- Qual o conjunto ou agrupamento que representa o espaço amostral, e quem são seus elementos?
- Que critérios foram usados para classificar os elementos desse espaço amostral, e como se expressam como partições?

Mais adiante, na Seção [4.1](#), será exemplificado o uso desses questionamentos e da Definição [8](#). Vejamos agora algumas consequências dessa definição:

Corolário 8.1. *Todos os eventos elementares $\{e_i\}$, $i \in \mathbb{N}$, constituem uma partição de Ω .*

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 , é interessante notar que 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:

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

que é a equação da Lei da Probabilidade Total.

Corolário 8.2. *Os eventos A e \bar{A} (complementar de A) constituem uma partição de Ω .*

Considerando um espaço amostral $\Omega = \{e_{k_1}, e_{k_2}, \dots, e_{w_1}, e_{w_2}, \dots\}$ e um possível evento dele $A = \{e_{k_1}, e_{k_2}, \dots\}$, o evento complementar de A seria $\bar{A} = \{e_{w_1}, e_{w_2}, \dots\}$, com $k_i \neq w_j$ para $\forall i, j \in \mathbb{N}$. Também, de acordo com os Itens 2 e 3 da Definição [6](#):

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

De modo que chegamos à equação da probabilidade do evento complementar:

$$1 = P(A) + P(\bar{A}) \implies P(\bar{A}) = 1 - P(A). \quad (3)$$

Em resumo, para qualquer evento $A \subset \Omega$, também haverá um evento complementar $\bar{A} \subset \Omega$, dos quais sempre se poderá dizer que:

- Se A aconteceu, então \bar{A} não aconteceu, ou seja, o sucesso de A implica em falha de \bar{A} . E vice-versa: a falha de A implica em sucesso de \bar{A} .
- A soma das probabilidades de A e de \bar{A} é sempre 1.
- Conhecer a probabilidade de sucesso de A implica em conhecer também a probabilidade para a sua falha (ou o sucesso de \bar{A}), o que pode até ser entendido como um modo *dialético* de resolver problemas da probabilidade.

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*

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

Uma interpretação comum da probabilidade condicional é que ela representa uma mudança de espaço amostral e, portanto, do contexto de estudo. De modo mais prático, ela troca o universo de possibilidades de resultados de um experimento (ou seja, Ω) por outro universo de possibilidades (Ω'), que é “menor” em comparação ao original. A consequência imediata disso é que as probabilidades de certos eventos, outrora calculadas em relação a certo espaço amostral, poderão ter valores diferentes em outro espaço amostral. Um evento E pode se tornar mais provável ou menos provável (ou até impossível) dependendo do tipo de mudança do espaço amostral.

Outra interpretação possível da probabilidade condicional é a de que ela busca obter a probabilidade percebida por observadores que tenham espaços amostrais diferentes. Esses outros observadores tem “campos de visão” reduzidos, isto é, são subconjuntos do espaço amostral do observador original. O exemplo abaixo ilustra bem este caso:

Exemplo 3 (Um dado de 6 faces e dois observadores enviesados). *Lança-se um dado com 6 faces, cada face numerada com um número de 1 a 6, e duas pessoas observam a face superior. Uma delas só aceitará um número maior que 3; a outra só aceitará um número primo.*

A diferença entre este exemplo e o Exemplo 2 são as condições para cada uma das pessoas. Aqui, já se suspeita que os resultados possíveis desse experimento para cada pessoa não podem ser os mesmos, pois há uma diferença das condições. Enquanto o observador do Exemplo 2 aceita qualquer um dos resultados do dado, as pessoas deste exemplo “reduziram” seus possíveis resultados.

Chamemos de Ω_1 ao espaço amostral do primeiro observador e de Ω_2 ao espaço amostral do segundo observador. Portanto $\Omega_1 = \{4; 5; 6\}$ e $\Omega_2 = \{2; 3; 5\}$. Vamos também considerar que exista um observador “neutro”, isto é, o mesmo do Exemplo 2. Assumindo o Princípio da Indiferença para o dado e para todos os observadores, as probabilidades dos eventos elementares para cada um deles será $P_1(\{e\}) = \frac{1}{3}$, $P_2(\{e\}) = \frac{1}{3}$ e $P(\{e\}) = \frac{1}{6}$ respectivamente, para o primeiro e segundo observador e para o observador neutro, sendo $\{e\}$ um evento elementar qualquer dos seus respectivos espaços amostrais.

Suponhamos agora que se deseja calcular a probabilidade de observar um número par como resultado.

Para o primeiro observador, o evento observado foi $E_1 = \{4; 6\}$, então:

$$P_1(E_1) = \frac{\#(E_1)}{\#(\Omega_1)} = \frac{2}{3}.$$

Para o segundo observador, o evento observado seria $E_2 = \{2\}$, logo:

$$P_2(E_2) = \frac{\#(E_2)}{\#(\Omega_2)} = \frac{1}{3}.$$

Já para o observador neutro, o evento seria $E = \{2; 4; 6\}$, de modo que:

$$P(E) = \frac{\#(E)}{\#(\Omega)} = \frac{3}{6} \implies P(E) = \frac{1}{2}.$$

Note que cada observador, calculando a probabilidade do evento desejado, não usa informações de nenhum outro observador. Os valores usados para os cálculos individuais das probabilidades só dependem de informações também individuais. Contudo, é possível que esses valores sejam calculados de outra forma. A noção de probabilidade condicional permite, por exemplo, que sejam usadas apenas informações do observador neutro para calcular alguma probabilidade percebida por qualquer outro observador. Para o observador neutro, tanto o primeiro como o segundo observador deste exemplo terão probabilidades condicionais e, na Equação (4), os valores a serem usados no lado direito da igualdade na expressão farão uso apenas de informações do observador neutro.

Vejamos como seria o cálculo da probabilidade do mesmo evento segundo o observador neutro, porém condicionado ao primeiro observador. Note que *para o observador neutro, o espaço amostral do primeiro observador (isto é, Ω_1) é, na verdade, um evento*, uma vez que $\Omega_1 \subset \Omega$. Portanto, presumindo também o Princípio da Indiferença e usando apenas informações do observador neutro, teremos

$$P(\Omega_1) = \frac{\#(\Omega_1)}{\#(\Omega)} = \frac{3}{6} \implies P(\Omega_1) = \frac{1}{2}.$$

Tem-se ainda $E \cap \Omega_1 = \{4; 6\} \implies P(E \cap \Omega_1) = \frac{2}{6} = \frac{1}{3}$, e então:

$$P(E|\Omega_1) = \frac{P(E \cap \Omega_1)}{P(\Omega_1)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

que coincide com $P_1(E_1)$, que era a probabilidade do mesmo evento calculada apenas com as informações do primeiro observador. Obtivemos, portanto, a probabilidade do evento E em Ω_1 usando apenas informações do observador neutro.

Já o cálculo da probabilidade do evento pelo observador neutro condicionado ao segundo observador seria análogo. O espaço amostral Ω_2 desse observador é também um evento para o observador neutro (pois $\Omega_2 \subset \Omega$) e, por isso,

$$P(\Omega_2) = \frac{\#(\Omega_2)}{\#(\Omega)} = \frac{3}{6} \implies P(\Omega_2) = \frac{1}{2}.$$

Tem-se ainda $E \cap \Omega_2 = \{2\} \implies P(E \cap \Omega_2) = \frac{1}{6}$, e obteríamos

$$P(E|\Omega_2) = \frac{P(E \cap \Omega_2)}{P(\Omega_2)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

que também coincide com a probabilidade do evento calculada com as informações do segundo observador, $P_2(E_2)$.

2.3 TEOREMA DE BAYES

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

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

de onde deduzimos o

Corolário 9.1 (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*

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

Esta relação matemática foi deduzida iniciando com axiomatização dos espaços probabilísticos e fazendo uso de definições bastante intuitivas da teoria dos conjuntos. No entanto, a utilidade dessa simples relação vai além da matemática. Com ela, torna-se possível pensar e solucionar um tipo diferente de problemas na probabilidade. 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 – na história do teorema de Bayes descrita no capítulo 1 – de *probabilidade inversa*, e que por sua vez pode ser vista como uma *probabilidade de causas* se o evento H estiver associado a alguma “causa” (aristotelicamente falando) do evento E em estudo.

Exemplo 4 (Dois dados de 6 faces e três observadores). *Lançam-se dois dados de 6 faces, cada face numerada com um número de 1 a 6, e uma pessoa (observador neutro) observa os números das faces superiores. Outras duas pessoas também observam os dados. A primeira só aceita resultados em que ambas as faces dos dados sejam números primos; a segunda só aceita resultados em que a soma das faces dos dados seja um número primo.*

É necessário, antes de tudo, montar o modelo probabilístico deste experimento. Para isso, vamos conceber o espaço amostral Ω do observador neutro como conjunto de elementos em forma de par ordenado $(x; y)$, sendo x o número da face superior de um dado e y é o número da face superior do outro dado. Então, $\#(\Omega) = 6 \times 6 = 36$, sendo que

$$\Omega = \{(1; 1); (1; 2); \dots; (1; 6); (2; 1); (2; 2); \dots; (6; 6)\}.$$

Assumindo o Princípio da Indiferença, concebe-se também a função de probabilidade

$$P(\{(x; y)\}) = \frac{1}{\#(\Omega)} = \frac{1}{36}, \forall x, y \in \{1; 2; 3; 4; 5; 6\}.$$

Suponhamos que se deseja estudar e relacionar as probabilidades percebidas pelos observadores neste experimento. Chamando de \mathbb{P} ao conjunto dos números primos naturais, então podemos escrever que o espaço amostral do primeiro observador D é $\Omega_D = \{(x; y) \in \Omega \mid x, y \in \mathbb{P}\}$, e o espaço amostral do segundo observador S é $\Omega_S = \{(x; y) \in \Omega \mid x + y \in \mathbb{P}\}$. Note que $\Omega_D, \Omega_S \subset \Omega$ e, por isso, para o observador neutro, os espaços amostrais dos outros dois observadores são interpretados como eventos.

Pela tabela a seguir, as informações referentes aos espaços amostrais envolvidos ficam um pouco mais claras. Pode-se dizer que o observador D , a princípio, estaria interessado apenas na coluna da esquerda – com os possíveis resultados \mathbf{x} do primeiro dado – e também na linha do topo – com os possíveis resultados \mathbf{y} do segundo dado. Já o observador S estaria mais interessado no meio da tabela, onde estão os possíveis resultados da soma $\mathbf{x} + \mathbf{y}$ de acordo com a linha e a coluna.

O observador neutro notaria rapidamente que $\#(\Omega_D) = 9 \implies p(\Omega_D) = \frac{9}{36} = \frac{1}{4}$, $\#(\Omega_S) = 15 \implies p(\Omega_S) = \frac{15}{36} = \frac{5}{12}$ e que $\#(\Omega_S \cap \Omega_D) = 4 \implies p(\Omega_S \cap \Omega_D) = \frac{4}{36} = \frac{1}{9}$. Com essas informações, ele consegue calcular qualquer probabilidade, inclusive aquelas

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

Tabela 1: Possíveis resultados para o Exemplo 4

que seriam percebidas pelos outros observadores, como por exemplo, a probabilidade de que o observador D observe um resultado cuja soma seja um número primo. Do ponto de vista desse observador neutro, essa probabilidade é condicional e será expressa por

$$P(\Omega_S|\Omega_D) = \frac{P(\Omega_S \cap \Omega_D)}{P(\Omega_D)} = \frac{\frac{1}{9}}{\frac{1}{4}} = \frac{4}{9}.$$

Do ponto de vista do observador D , dos seus $\#(\Omega_D) = 9$ possíveis resultados, a probabilidade calculada acima realmente faz sentido, já que apenas 4 deles teriam uma soma que é número primo.

O que aconteceria se fosse desejado calcular a probabilidade condicional “trocada”, isto é, a probabilidade de que o observador S notar que os números que ele somou também são ambos primos? Aqui entra o Teorema de Bayes que, do ponto de vista do observador neutro, expressará essa outra probabilidade assim:

$$P(\Omega_D|\Omega_S) = \frac{P(\Omega_D) \times P(\Omega_S|\Omega_D)}{P(\Omega_S)} = \frac{\frac{1}{4} \times \frac{4}{9}}{\frac{5}{12}} = \frac{4}{15}.$$

Tal número faz sentido para o observador S , pois das suas $\#(\Omega_S) = 15$ possibilidades, só 4 delas realmente vieram da soma de dois números primos.

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,*

$$O(E) = \frac{P(E)}{P(\bar{E})} = \frac{P(E)}{1 - P(E)}. \quad (6)$$

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 $\frac{p}{q}$ 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 .

Corolário 10.1 (Razão de Chances Condicional). *Sejam dois eventos E e H do espaço amostral Ω , com $P(H) \neq 0$ e $P(E) \neq 0$. Chama-se de razão de chances condicional de H em E à razão entre as probabilidades condicionais de sucesso e de falha de H em E , isto é,*

$$O(H|E) = \frac{P(H|E)}{P(\bar{H}|E)}. \quad (7)$$

Usando o teorema de Bayes para calcular a probabilidade condicional da falha de H em E , teríamos

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

A partir da Equação 7, que é uma divisão entre as probabilidades de sucesso e de falha de H em E , cada uma delas expressa respectivamente pela Equação 5 e pela equação logo acima, pode-se assim escrever:

$$O(H|E) = \frac{\frac{P(H) \times P(E|H)}{P(E)}}{\frac{P(\bar{H}) \times P(E|\bar{H})}{P(E)}}$$

de onde obtemos uma outra versão da razão de chances condicional, chamada de

Corolário 10.2 (Razão de Chances com Fator de Bayes). *Sejam dois eventos E e H do espaço amostral Ω , com $P(H) \neq 0$ e $P(E|\bar{H}) \neq 0$. Então a razão de chances de H em E com fator de Bayes é dada por*

$$O(H|E) = \frac{P(H)}{P(\bar{H})} \times \frac{P(E|H)}{P(E|\bar{H})} = O(H) \times B(E|H). \quad (8)$$

O termo $B(E|H) = \frac{P(E|H)}{P(E|\bar{H})}$ é chamado de *fator de Bayes* de E em relação a H .

Observe que, matematicamente, o que diferencia a razão de chances simples da razão de chances condicional é o fator multiplicativo $B(H|E)$, o fator de Bayes.

SIMULAÇÃO COMPUTACIONAL DO EXPERIMENTO DE BAYES

Far-se-á, neste capítulo, uma simulação computacional do experimento de Bayes. Tal experimento foi descrito na página 7. Para simula-lo, foi necessário dividir o espaço da superfície da mesa em um número finito, n , de regiões verticais para que possamos localizar as bolas lançadas. Essas bolas serão lançadas na mesa também na direção vertical, não importando o sentido delas (de baixo para cima ou de cima para baixo).

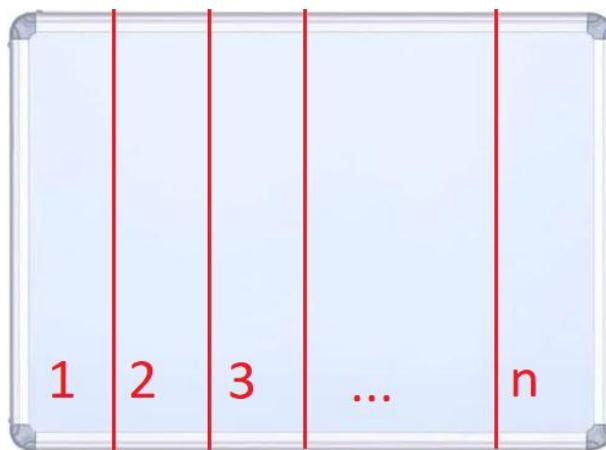


Figura 1: Ilustração da mesa dividida em n regiões verticais

Antes de irmos à didática dessa simulação, é preciso mostrar que há um modelo probabilístico nesse experimento, isto é, um espaço amostral Ω e uma função de probabilidade $P(E)$, segundo a Definição 7. Vejamos:

- Os possíveis resultados de um lançamento de bola são as regiões numeradas de 1 a n . Por isso, tem-se $\Omega = \{1, 2, \dots, n\}$, com $\#(\Omega) = n$.
- Assumindo o Princípio da Indiferença (i.e., a posição final da bola pode ser qualquer uma das n regiões indistintamente), a função de probabilidade $P(E)$ pode ser simplesmente $P(\{1\}) = P(\{2\}) = \dots = P(\{n\}) = \frac{1}{\#(\Omega)} = \frac{1}{n}$. Quando o conjunto for elementar, far-se-á a simplificação de $P(\{e\})$ por $P(e)$.

No entanto, o experimento de Bayes não é um simples lançamento de bola no espaço da mesa e localizá-la em alguma das n regiões. O objetivo do experimento é descobrir a posição da marcação a partir de informações colhidas a partir de próximos lançamentos de bola. O processo para fazer essa descoberta é o que se chama de inferência bayesiana que, como mencionamos anteriormente (página 3), é uma ferramenta matemática favorável ao desenvolvimento das competências listadas na BNCC.

Suponhamos, portanto, que a bola tenha sido lançada para que se faça a marcação, e esta tenha sido feita na região k . Então $k \in \Omega$, isto é, $k \in \{1, 2, \dots, n\}$. Somente o ajudante sabe em qual região da mesa a bola teria parado – pois ele fez a marcação na mesa –, mas Bayes não. Sabendo que os próximos lançamentos da bola não dependem do resultado de lançamentos anteriores, então nota-se que 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 se os lançamentos caem à esquerda ou à direita da marcação inicial. Conforme as informações que o ajudante dá, Bayes poderá, gradualmente, afirmar com maior confiança em qual região da mesa a marcação inicial havia sido feita.

O processo de inferência que Bayes fará começa pela avaliação das probabilidades de que os novos lançamentos caiam à esquerda ou à direita da marcação. O evento “cair à esquerda da marcação” pode ser representado por $Q = \{1, 2, \dots, k - 1\}$ e, por isso, $\#(Q) = k - 1$. De modo análogo, o evento “cair à direita da marcação” será representado por $D = \{k + 1, k + 2, \dots, n\}$, com $\#(D) = n - k$. Desse modo, tem-se as probabilidades de que cada lançamento novo caia à esquerda ou à direita, condicionados à marcação ter sido feita na região k :

$$P(Q|k) = \frac{\#(Q)}{\#(\Omega)} = \frac{k - 1}{n};$$

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

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 idéia semelhante à da razão de chances (ver Equação 6), para cada n lançamentos da bola, espera-se que $k - 1$ lançamentos caiam à esquerda da marcação e $n - k$ lançamentos caiam à direita. 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 se dará, 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 qual seria a posição da marcação.

No Anexo [A](#), exibe-se o código-fonte do algoritmo de simulação do experimento de Bayes, escrito em linguagem *Python*. Agora, passar-se-á às exemplificações da inferência fazendo uso do algoritmo de simulação do experimento.

3.1 UM LANÇAMENTO POR VEZ

Nesta seção, uma exemplificação do processo de inferência 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 mais fácil acesso, além da grande quantidade de exemplos que já ocorrem em livros didáticos. Aliás, no Anexo [B](#), tem-se um plano de aula que oferece uma sugestão de como uma aula seria organizada para o ensino do Teorema de Bayes, tendo em vista este experimento de Bayes, adaptado para um dado de 6 faces, e também ilustrando como ocorre o processo de inferência bayesiano. 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 arbitrária de lançamentos. Neste primeiro exemplo, a simulação ocorrerá de maneira que a inferência seja feita gradualmente.

Suponhamos que a marcação tenha sido feita na região 4, isto é, $k = 4$. Sendo assim, à esquerda da marcação há 3 regiões (ou seja, $\#(Q) = 3$) e, à direita há 2 regiões (ou seja, $\#(D) = 2$). Por isso, há uma probabilidade maior de que os lançamentos novos da bola na mesa caiam à esquerda da marcação do que à direita. De modo mais específico, essas probabilidades serão exatamente $P(Q|4) = \frac{3}{6}$ e $P(D|4) = \frac{2}{6}$ respectivamente. Espera-se, portanto, que a cada 6 lançamentos da bola, 3 deles caiam à esquerda e 2 caiam à direita. O que poderia acontecer na simulação desse experimento?

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 do experimento mental, se referirão, respectivamente, ao “ajudante” e ao próprio Bayes.

Inicialmente, a primeira informação que o usuário precisa é de saber em quantas regiões a mesa foi dividida. Abaixo segue as instruções iniciais do programa até o momento em que é solicitado o número de regiões em que a mesa será dividida:

Tela 1: Informando a quantidade de regiões na mesa

```

1  === ALGORITMO BAYESIANO ===
2  Para sair do programa, basta apertar ENTER sem fornecer informacao util
   ao programa em qualquer momento.
3  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 dele é justamente o de 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. Abaixo são exibidas as mensagens do programa até esse ponto:

Tela 2: Fazendo um lançamento

```

4 A mesa foi dividida igualmente em regioes numeradas de 1 a 6, da
   esquerda para a direita.
5 (quanto mais proximo de 1, a regioao estah mais a esquerda)
6 (quanto mais proximo de 6, a regioao estah mais a direita)
7 A bola foi lancada na mesa. O ajudante marcou a posicao dela em uma
   dessas regioes.
8
9 SEU OBJETIVO: descobrir em qual regioao a marcacao foi feita.
10 Agora o ajudante vai lancar a bola quantas vezes voce desejar para obter
   informacao.
11 Depois desses lancamentos, ele lhe dira em quantos deles a bola ficou a
   ESQUERDA ou a DIREITA da marcacao.
12
13 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 acima na Tela 2).

Após o programa “fazer o lançamento”, ele informa ao usuário quantos deles ocorreram à esquerda e à direita da marcação original. Ao receber a informação, o usuário pode tentar descobrir onde a marcação foi feita, e o programa informa se ele acertou ou não.

Tela 3: Resultado do lançamento

```

14 AJUDANTE: A bola foi lancada 1 vez(es) agora. Em relacao a marcacao,
15 1 foram a esquerda (100.0 % desses lancamentos)
16 0 foram a direita (0.0 % desses lancamentos)
17
18 AJUDANTE: Total de 1 lancamentos, sendo:
19 1 a esquerda (100.0 % do total)
20 0 a direita (0.0 % do total)

```

A partir de agora, a busca pela resposta se dará com as informações oferecidas. O que é possível inferir sobre o resultado acima?

Ainda não há certeza do local da 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 e 6 aquela mais à direita), então pode-se inferir com esse primeiro lançamento que, se a marcação tivesse sido feita na região 1, não poderiam ocorrer lançamentos à esquerda dela. Essa impossibilidade é matematicamente traduzida pelo valor zero de probabilidade, que é obtido relembrando as relações da página 32, com $n = 6$ e supondo $k = 1$: $P(Q|1) = \frac{1-1}{6} = 0$.

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.

É importante lembrar que a função de probabilidade original é $P(e) = \frac{1}{6}$, $\forall e \in \{1, 2, 3, 4, 5, 6\}$ (levando em conta o modelo probabilístico da página 31). Ao realizar o primeiro lançamento e obter a informação de que ele caiu à esquerda da marcação, é preciso montar uma função de probabilidade que estará mais atualizada com as informações disponíveis até então. Simbolizando essa informação por Q – isto é, o evento “cair à esquerda da marcação” ter ocorrido uma vez –, e lembrando que $P(Q|1) = 0$, passar-se-á a trabalhar com uma função de probabilidade atualizada, indicada por $P_1(E)$, que será entendida como $P(E)$ condicionada a Q , ou seja, $P_1(E) = P(E|Q)$. Essa função terá seus valores avaliados para cada elemento de Ω , fazendo uso do Teorema de Bayes (página 27). Avaliando primeiramente $P_1(1) = P(1|Q)$, tem-se:

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

Isso já era esperado pois, se a marcação estivesse na região 1, era impossível que qualquer lançamento posterior caísse à esquerda. Então, por isso, se fosse observado que um lançamento caísse à esquerda, então é impossível que a marcação estivesse na região 1, não importando o valor de $P(Q)$. Contudo, o valor de $P(Q)$ será importante na avaliação de $P_1(E)$ 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):

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

O valor de $P(D)$ (probabilidade de o lançamento cair à direita da marcação) seria obtido de modo similar. Pela simetria do problema, o resultado também seria $P(D) = \frac{15}{36}$.

O restante dos valores de probabilidade de $P_1(E)$ são assim encontrados:

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

Em suma, o que essa nova função de probabilidade $P_1(E)$ faz é oferecer valores de probabilidade dadas todas as informações disponíveis. Até agora, a única informação disponível é a condicional Q . 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 $P_1(E)$, ainda que exista incerteza, já é possível fazer tentativas mais objetivas, não baseadas na completa ignorância, pois a função indica que há regiões mais prováveis que outras. Suponhamos que essa tentativa seja feita para eliminar o outro extremo da mesa, a região 6.

Tela 4: Primeira tentativa

```

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

```

Note que $P_1(6) = \frac{5}{15}$ é a maior das probabilidades. Isso quer dizer que essa tentativa foi feita por ser a região mais provável das opções existentes, denotando maior objetividade na busca. Caso o usuário não acerte – conforme a Tela 4 foi o que realmente aconteceu –, o programa o informará de que a tentativa não foi bem sucedida, e oferecerá outra oportunidade de lançar a bola para que o usuário obtenha mais informação. Aqui se inicia um ciclo que se repetirá até que o usuário acerte em 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 esta restrição por $\overline{\{6\}}$ – isto é, a marcação não está na região 6 –, monta-se uma segunda função de probabilidade $F_2(E) = P_1(E|\overline{\{6\}})$, que assume os seguintes valores para cada evento elementar:

- $F_2(1) = 0.$
- $F_2(2) = P_1(2|\overline{\{6\}}) = P_1(\overline{\{6\}}|2) \frac{P_1(2)}{P_1(\overline{\{6\}})} = 1 \times \frac{P_1(2)}{1-P_1(6)} = \frac{\frac{1}{15}}{1-\frac{5}{15}} = \frac{\frac{1}{15}}{\frac{10}{15}} = \frac{1}{10}.$
- $F_2(3) = P_1(3|\overline{\{6\}}) = P_1(\overline{\{6\}}|3) \frac{P_1(3)}{P_1(\overline{\{6\}})} = \frac{\frac{2}{15}}{\frac{10}{15}} = \frac{2}{10}.$
- $F_2(4) = P_1(4|\overline{\{6\}}) = P_1(\overline{\{6\}}|4) \frac{P_1(4)}{P_1(\overline{\{6\}})} = \frac{\frac{3}{15}}{\frac{10}{15}} = \frac{3}{10}.$
- $F_2(5) = P_1(5|\overline{\{6\}}) = P_1(\overline{\{6\}}|5) \frac{P_1(5)}{P_1(\overline{\{6\}})} = \frac{\frac{4}{15}}{\frac{10}{15}} = \frac{4}{10}.$
- $F_2(6) = P_1(6|\overline{\{6\}}) = 0.$

Dando prosseguimento ao ciclo, após o usuário atualizar sua função de probabilidade, ele poderá solicitar um novo lançamento.

Tela 5: Lançamento sem informação nova

```

23 AJUDANTE: Quantas vezes devo lancar a bola? 1
24 AJUDANTE: A bola foi lancada 1 vez(es) agora. Em relacao a marcacao,
25 0 foram a esquerda (0.0 % desses lancamentos)
26 0 foram a direita (0.0 % desses lancamentos)
27
28 AJUDANTE: Total de 2 lancamentos, sendo:
29 1 a esquerda (50.0 % do total)
30 0 a direita (0.0 % do total)

```

Dessa vez, o programa informou que a bola caiu numa região que não é nem à esquerda e nem à direita da marcação. 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:

Tela 6: Segunda tentativa

```

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

```

O programa exibiu as mensagens da Tela 6 acima. 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 $F_3(E) = F_2(E|\overline{\{5\}})$, mais atualizada, cujos valores nos eventos elementares serão:

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

Essa função $F_3(E)$ produz resultados bem diferentes quando em comparação com a função $P(E)$, na qual nada se sabia ainda. A região 4 aparece com probabilidade $\frac{1}{2}$, a maior até então. A única região com a mesma probabilidade do início é a região 2.

Nesse momento, o programa oferece ao usuário uma nova oportunidade de lançamento.

Tela 7: Último lançamento

```

33 AJUDANTE: Quantas vezes devo lancar a bola? 1
34 AJUDANTE: A bola foi lancada 1 vez(es) agora. Em relacao a marcacao,
35 1 foram a esquerda (100.0 % desses lancamentos)
36 0 foram a direita (0.0 % desses lancamentos)
37
38 AJUDANTE: Total de 3 lancamentos, sendo:
39 2 a esquerda (66.7 % do total)
40 0 a direita (0.0 % do total)

```

Segundo a Tela 7 acima, esse último lançamento caiu à esquerda da marcação. Isto também é nova informação, pois houveram no todo até agora 2 lançamentos à esquerda. Essa condicional será simbolizada por Q^2 . Com ela, se monta uma função de probabilidade $P_2(E) = P(E|Q^2)$ a partir da qual se pode montar uma outra função, também atualizada, que incorpore as restrições já estabelecidas, isto é, $\overline{\{5,6\}}$. Para obter essa função, será necessário antes obter as probabilidades dos eventos elementares de $P_2(E)$, de modo muito análogo ao de obtenção de $P_1(E)$:

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

Agora, os valores da função $P_2(E) = P(E|Q^2)$ em seus eventos elementares:

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

- $P_2(5) = P(5|Q^2) = P(Q^2|5) \frac{P(5)}{P(Q^2)} = \frac{16}{36} \times \frac{36}{55} = \frac{16}{55}$.
- $P_2(6) = P(6|Q^2) = P(Q^2|6) \frac{P(6)}{P(Q^2)} = \frac{25}{36} \times \frac{36}{55} = \frac{25}{55}$.

Depois de tudo isso, torna-se possível obter as probabilidades dos eventos elementares da nova função $F_4(E)$, que incorpora a restrição $\overline{\{5, 6\}}$ a partir de $P_2(E)$. Isso significa que $F_4(E) = P_2(E|\overline{\{5, 6\}})$:

- $F_4(1) = F_4(5) = F_4(6) = 0$.
- $F_4(2) = P_2(2|\overline{\{5, 6\}}) = P_2(\overline{\{5, 6\}}|2) \frac{P_2(2)}{P_2(\overline{\{5, 6\}})} = 1 \times \frac{P_2(2)}{1 - P_2(5) - P_2(6)} = \frac{\frac{1}{55}}{\frac{14}{55}} = \frac{1}{14}$.
- $F_4(3) = P_2(3|\overline{\{5, 6\}}) = P_2(\overline{\{5, 6\}}|3) \frac{P_2(3)}{P_2(\overline{\{5, 6\}})} = \frac{\frac{4}{55}}{\frac{14}{55}} = \frac{4}{14}$.
- $F_4(4) = P_2(4|\overline{\{5, 6\}}) = P_2(\overline{\{5, 6\}}|4) \frac{P_2(4)}{P_2(\overline{\{5, 6\}})} = \frac{\frac{9}{55}}{\frac{14}{55}} = \frac{9}{14}$.

Após essa longa sequência de cálculos só para que fosse obtida uma função de probabilidade completamente atualizada com todas as informações disponíveis até então, o usuário está novamente pronto para fazer mais uma tentativa objetiva, conforme sua função o indica. A região com maior probabilidade de ter a marcação é a 4, e é esta que o usuário informará ao programa na sua próxima tentativa.

Tela 8: Tentativa correta

```

41 AJUDANTE: Em qual regioao voce acha marcacao foi feita? 4
42 AJUDANTE: VOCE ACERTOU! A marcacao foi feita na regioao 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.

Qual seria a grande lição tirada deste primeiro exemplo? 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.

Usar a função de probabilidade como critério de inferência bayesiana pode servir para tomada de decisão, por exemplo. O estudante de ensino médio, que esteja em busca de aperfeiçoar sua opinião inicial sobre um determinado experimento, poderá usufruir do arcabouço teórico do Capítulo 2 – 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 do 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.

3.2 VÁRIOS LANÇAMENTOS POR VEZ

Agora uma outra exemplificação do processo de inferência será exibida, mas dessa vez serão feitos vários lançamentos em cada vez. Mantém-se a mesa com 6 regiões, para que a similaridade com o lançamento de dado de 6 faces também se mantenha.

Na Tela 9 abaixo, segue a configuração inicial até o momento em que o programa solicita pela primeira vez o número de lançamentos a serem realizados:

Tela 9: Iniciando outra simulação

```
1 === ALGORITMO BAYESIANO ===
2 Para sair do programa, basta apertar ENTER sem fornecer informacao util
   ao programa em qualquer momento.
3 Numero de regioes da mesa: 6
4 A mesa foi dividida igualmente em regioes numeradas de 1 a 6, da
   esquerda para a direita.
5 (quanto mais proximo de 1, a regioao estah mais a esquerda)
6 (quanto mais proximo de 6, a regioao estah mais a direita)
7 A bola foi lancada na mesa. O ajudante marcou a posicao dela em uma
   dessas regioes.
8
9 SEU OBJETIVO: descobrir em qual regioao a marcacao foi feita.
10 Agora o ajudante vai lancar a bola quantas vezes voce desejar para obter
   informacao.
11 Depois desses lancamentos, ele lhe dira em quantos deles a bola ficou a
   ESQUERDA ou a DIREITA da marcacao.
```

Nesse ponto, o programa já marcou alguma das 6 regiões da mesa, e o objetivo do usuário continua sendo o de descobrir onde essa marcação foi feita. Dessa vez, o processo de inferência não será gradual: de uma vez só, o usuário vai observar os resultados de uma quantidade arbitrária de lançamentos, que lhe darão as frequências reais do experimento. Ao compará-las com as frequências esperadas para cada possibilidade de marcação – já que a tendência é que as frequências reais se aproximem de alguma das frequências esperadas –, então ele terá uma boa confiança sobre o local da marcação inicial.

O usuário, fazendo bom uso da capacidade computacional da sua máquina, solicitará ao programa que faça 1000 lançamentos. Os resultados de todos esses lançamentos são exibidos na Tela 10 abaixo:

Tela 10: Resultado de 1000 lançamentos

```
12 AJUDANTE: Quantas vezes devo lancar a bola? 1000
13 AJUDANTE: A bola foi lancada 1000 vez(es) agora. Em relacao a marcacao,
14 167 foram a esquerda (16.7 % desses lancamentos)
```

```

15 648 foram a direita (64.8 % desses lançamentos)
16
17 AJUDANTE: Total de 1000 lançamentos, sendo:
18 167 a esquerda (16.7 % do total)
19 648 a direita (64.8 % do total)

```

O que os resultados sugerem sobre a localização da marcação? Conforme mencionado na página [32](#), as probabilidades de um lançamento cair à esquerda ou à direita tem valores bem determinados e que permitem obter as frequências esperadas para os lançamentos. Assim, para uma mesa dividida em 6 regiões de igual área ($n = 6$), pode-se obter as frequências esperadas caso a marcação estivesse em qualquer uma das regiões se forem determinadas as probabilidades dos lançamentos em cada caso:

- Para região 1, as probabilidades são $P(Q|1) = \frac{1-1}{6} = 0$ e $P(D|1) = \frac{6-1}{6} = \frac{5}{6}$. Espera-se portanto que para cada 6 lançamentos da bola, 5 deles caiam à direita, isto é, aproximadamente 83% dos lançamentos. A bola nunca cairá à esquerda.
- Para a região 2, as probabilidades são $P(Q|2) = \frac{1}{6}$ e $P(D|2) = \frac{4}{6}$. A expectativa é de que a cada 6 lançamentos, 1 caia à esquerda (aproximadamente 17%) e 4 caiam à direita (aproximadamente 67%). Em termos de razão de chances, isso significa 4 vezes mais lançamentos à direita do que à esquerda.
- Para a região 3, $P(Q|3) = \frac{2}{6}$ e $P(D|3) = \frac{3}{6}$. Então a frequência esperada é de que a cada 6 lançamentos, 2 caiam à esquerda (aproximadamente 33%) e 3 caiam à direita (50%), o que significa $\frac{3}{2} = 1.5$ vezes mais lançamentos à direita do que à esquerda.
- Para a região 4, $P(Q|4) = \frac{3}{6}$ e $P(D|4) = \frac{2}{6}$. A frequência esperada é de, a cada 6 lançamentos, 3 sejam à esquerda (50%) e 2 à direita (aproximadamente 33%). Isso significa $\frac{3}{2} = 1.5$ vezes mais lançamentos à esquerda do que à direita.
- Para a região 5, $P(Q|5) = \frac{4}{6}$ e $P(D|5) = \frac{1}{6}$ com frequência esperada de, em 6 lançamentos, 4 caiam à esquerda (aproximadamente 67%) e 1 à direita (aproximadamente 17%). Haveria 4 vezes mais lançamentos à esquerda do que à direita.
- Para a região 6, $P(Q|6) = \frac{5}{6}$ e $P(D|6) = 0$. Em 6 lançamentos, espera-se que 5 deles caiam à esquerda, ou seja, 83% aproximadamente. A bola nunca cairá à direita.

Nota-se que as frequências esperadas da marcação em cada região oferecem informações inequívocas. Não há duplicidade de frequências esperadas de regiões diferentes. Se as frequências reais do experimento se aproximarem de alguma das frequências esperadas para certa região, então a marcação realmente deve estar naquela região, e não em outra. E sem grandes dificuldades, uma comparação simples leva à inferência de que, nessa simulação, a marcação deve estar na região 2. Esta é a região que o usuário, portanto, informará ao programa em sua tentativa.

Tela 11: Resultado de 1000 lançamentos

```

20 AJUDANTE: Em qual regioa voce acha marcacao foi feita? 2
21 AJUDANTE: VOCE ACERTOU! A marcacao foi feita na regioa 2.

```

Observe que a quantidade grande de lançamentos facilitou muito ao usuário realizar sua inferência, já que as frequências reais se aproximam visivelmente das frequências esperadas para alguma marcação. Contudo, na vida real, ao estudar algum problema mais complexo, nem sempre será possível realizar quantidade tão grande de experimentos (neste exemplo, os lançamentos da bola) para obter novas informações sobre certo fenômeno. Apesar disso, há uma percepção importante a ser feita. Se a tendência é que frequências reais se aproximam das esperadas, então pode-se usar esse critério de comparação *em conjunto* com o processo gradual já ilustrado e exemplificado na seção anterior.

3.3 COMPARANDO PARADIGMAS

Uma importante percepção a partir das inferências estatísticas realizadas nas simulações acima é que elas foram feitas de maneiras diferentes, *se assemelhando aos dois paradigmas da inferência estatística*: o frequentista e o bayesiano.

Na primeira simulação, o único critério para a inferência foi a função de probabilidade, que era atualizada conforme novas informações eram obtidas sobre o experimento. Já na segunda simulação, o critério é o de comparação entre as frequências reais dos resultados do experimento e as frequências esperadas, levando-se em conta um número considerável de lançamentos. Tais decisões de metodologia da inferência foram previamente feitas com este propósito de ilustrar a diferença desses paradigmas.

Observa-se, portanto, que os dois critérios usados nas simulações realmente produzem processos diferentes, o que provavelmente tenha sido a razão das contendas históricas entre aqueles grandes nomes da ciência mencionados no Capítulo 1. No entanto, os dois paradigmas podem muito bem ser usados em conjunto, ou, algum deles ser priorizado em relação ao outro dependendo das condições do experimento.

Considere, a fim de ilustração desse uso conjunto dos paradigmas, a primeira simulação. Nela, desde o seu início até o terceiro lançamento, o critério usado foi apenas a função de probabilidade (e portanto, fazia-se uso do paradigma bayesiano). Nesse momento, havia sido feita uma nova atualização da função de probabilidade. Contudo, o custo de atualização para essa nova função foi maior – basta lembrar a quantidade de cálculos – para que todas as informações disponíveis até então fossem incorporadas. Nesse ponto, já seria razoável considerar o uso do paradigma frequentista, comparando as frequências reais com as esperadas. Seria possível notar que, naquele momento, dos 3 lançamentos que tinham sido feitos, só 1 deles foi inconclusivo e os outros 2 haviam caído à esquerda. Isso por si mesmo já tem significado em termos de frequências esperadas: a marcação realmente deve estar mais à direita do que à esquerda. Levando em consideração que, naquele momento, só as regiões 2, 3 e 4 estavam com probabilidade maior que zero, então

imaginar que a marcação estivesse na região 4 já seria perfeitamente razoável, sem que fosse feita uma nova atualização da função de probabilidade.

Se por questões de praticidade ou limitações experimentais o custo de se fazer uma inferência pela atualização da função de probabilidade ficar muito grande – isto é, o paradigma bayesiano se tornar caro –, então abre-se espaço para o critério de comparação das frequências – ou seja, para o paradigma frequentista. O mesmo valeria para a situação reversa: se o custo de uma inferência pelo frequentismo ficar alto – o número de observações a serem feitas ficar alto demais –, então troca-se pelo paradigma bayesiano, dentro do qual o que determina a inferência não é o número de observações, mas sim a função de probabilidade já atualizada com toda a informação disponível.

PROPOSTAS PARA O ENSINO MÉDIO

No capítulo final desta monografia, buscar-se-á algumas propostas para o ensino da probabilidade e estatística e, em especial do Teorema de Bayes, no currículo do Ensino Médio, levando-se em conta a discussão historiográfica do primeiro capítulo, a teórica do segundo capítulo e a sistemática-computacional do terceiro capítulo.

4.1 USO DE TABELAS DE CONTINGÊNCIA

Considerando o arcabouço teórico da probabilidade – em especial, a noção de Probabilidade Condicional (conforme a Definição 9), a noção de partição de um espaço amostral (conforme a Definição 8) e a Lei da Probabilidade Total –, passemos a explicar o uso de Tabelas de Contingência (doravante referida apenas como “tabela”) para compreender como elas beneficiam uma visão mais completa da realidade e dos problemas em diversos campos da experiência humana e, em especial, nas ciências. Uma abordagem e uso avançados da tabela – cujas descrições ficam de fora do escopo desta monografia – poderão ser encontrados em Devore [23, p. 569] ou em DeGroot [24, p. 641].

Eis abaixo o esquema geral de uma tabela:

Critério 2 Critério 1	$\overline{C_2}$	C_2	Total
$\overline{C_1}$	$P(\overline{C_1} \cap \overline{C_2})$	$P(\overline{C_1} \cap C_2)$	$P(\overline{C_1})$
C_1	$P(C_1 \cap \overline{C_2})$	$P(C_1 \cap C_2)$	$P(C_1)$
Total	$P(\overline{C_2})$	$P(C_2)$	100% ou 1

Tabela 2: Estrutura geral de uma Tabela de Contingência

A tabela traz consigo um método esquemático de resolução de problemas, facilitando ao estudante que ele enxergue o quadro geral das informações disponíveis e das não disponíveis a ele, facilitando a busca daquilo que ainda é desconhecido e/ou relevante para solucionar o problema.

Para compreender como ela deve ser construída, é preciso inicialmente dar atenção aos questionamentos sobre a natureza do problema, cujas respostas informam quem representa o espaço amostral, isto é, os seus elementos, e também sobre os critérios usados para classificação desses elementos no espaço amostral. Vejamos:

- **Qual o conjunto ou agrupamento que representa o espaço amostral, e quem são seus elementos?** Todos os valores de probabilidade se referem a esses

elementos. Na tabela, o valor da probabilidade do espaço amostral Ω é representada na célula onde se escreve 100% ou 1.

- **Que critérios foram usados para classificar os elementos desse espaço amostral, e como se expressam como partições?** Todos os elementos do problema poderão ser classificados segundo dois critérios diferentes (e teoricamente independentes um do outro). Na tabela, esses dois critérios foram referidos como “Critério 1” (representado nas linhas) e “Critério 2” (representado nas colunas). Segundo o “Critério 1”, todo elemento $e \in \Omega$ deverá satisfazê-lo (isto é, $e \in C_1$) ou não satisfazê-lo (isto é, $e \notin C_1 \implies e \in \overline{C_1}$). Assim, as células $P(C_1)$ e $P(\overline{C_1})$ indicam, respectivamente, as probabilidades do evento satisfazer e não satisfazer ao “Critério 1”. Procedimento análogo valerá para o “Critério 2”: se $e \in \Omega$, então $e \in C_2$ ou $e \in \overline{C_2}$; e as células $P(C_2)$ e $P(\overline{C_2})$ indicam, respectivamente, as probabilidades de o elemento satisfazer e não satisfazer ao “Critério 2”.

Tendo respondido a esses questionamentos, o preenchimento do restante da tabela se torna possível quando compreendermos a disposição das probabilidades nela indicadas. A tabela permite a integração de dois critérios teoricamente independentes e, por conta disso, qualquer elemento e poderá pertencer a quatro possíveis eventos, resultantes da combinação daqueles critérios. Cada um desses quatro eventos terão suas probabilidades indicadas na tabela:

- $P(C_1 \cap C_2)$: probabilidade do elemento satisfazer a ambos os critérios;
- $P(C_1 \cap \overline{C_2})$: probabilidade de satisfazer a “Critério 1” mas não a “Critério 2”;
- $P(\overline{C_1} \cap C_2)$: probabilidade de satisfazer a “Critério 2” mas não a “Critério 1”;
- $P(\overline{C_1} \cap \overline{C_2})$: probabilidade de não satisfazer a nenhum dos critérios.

Há três importantes considerações a serem feitas aqui e que fazem parte do procedimento de preenchimento da tabela a fim de mantê-la consistente com o arcabouço teórico da probabilidade:

4.1.1 Observância da probabilidade do evento complementar

A Equação 3 do evento complementar precisa ser incorporada à estrutura da tabela. Para ilustrar, tomemos os eventos associados somente ao “Critério 1”, C_1 e $\overline{C_1}$. Esses eventos são complementares e, portanto, a soma das suas probabilidades deve ser 1, ou 100%. Note que essas três probabilidades mencionadas (C_1 , $\overline{C_1}$ e a soma 1 ou 100%) estão todas posicionadas verticalmente uma ao lado da outra na tabela, à direita, com a soma mais abaixo. Visualmente e didaticamente, isto é muito vantajoso, o que torna o preenchimento dos valores das probabilidades muito mais interativo. O estudante não

precisa buscar $P(C_1)$ e $P(\overline{C_1})$ separadamente: basta achar uma delas, e a outra já estará determinada.

De modo análogo, para o “Critério 2” e seus eventos C_2 e $\overline{C_2}$, que também são complementares, suas probabilidades e a sua soma estão dispostas horizontalmente uma ao lado da outra na tabela, abaixo, com a soma mais à direita. A mesma vantagem visual e didática ocorre, e o estudante não precisa buscar as probabilidades $P(C_2)$ e $P(\overline{C_2})$ separadamente. Pode-se até entender esta estratégia como uma forma de *dialética*.

Critério 2 Critério 1	$\overline{C_2}$	C_2	Total
$\overline{C_1}$	$P(\overline{C_1} \cap \overline{C_2})$	$P(\overline{C_1} \cap C_2)$	$P(\overline{C_1})$
C_1	$P(C_1 \cap \overline{C_2})$	$P(C_1 \cap C_2)$	$P(C_1)$
Total	$P(\overline{C_2})$	$\oplus P(C_2)$	100% ou 1

Tabela 3: Soma das probabilidades de eventos complementares

4.1.2 Observância da Lei da Probabilidade Total

É preciso também incorporar Lei da Probabilidade Total (Equação 2). Sem perda de generalidade, considere de início a probabilidade de satisfação do “Critério 1”, $P(C_1)$. Um elemento que satisfaça a este critério tem possibilidade de ter satisfeito ou não ao outro critério (“Critério 2”), pois já se assumiu que os critérios sejam teoricamente independentes um do outro. Como C_2 e $\overline{C_2}$ constituem uma partição de Ω , a Lei da Probabilidade Total se aplica, resultando em $P(C_1 \cap \overline{C_2}) + P(C_1 \cap C_2) = P(C_1)$. Essas três probabilidades estão dispostas na mesma linha da tabela, no meio dela, e o resultado $P(C_1)$ da soma aparece à direita. De maneira muito semelhante, a vantagem visual e didática do item anterior também se verifica. Neste caso, a determinação de dois desses três valores – $P(C_1 \cap \overline{C_2})$, $P(C_1 \cap C_2)$ ou $P(C_1)$ – implica na determinação do outro, também agilizando o preenchimento da tabela.

Critério 2 Critério 1	$\overline{C_2}$	C_2	Total
$\overline{C_1}$	$P(\overline{C_1} \cap \overline{C_2})$	$\oplus P(\overline{C_1} \cap C_2)$	$P(\overline{C_1})$
C_1	$P(C_1 \cap \overline{C_2})$	$\oplus P(C_1 \cap C_2)$	$P(C_1)$
Total	$P(\overline{C_2})$	$P(C_2)$	100% ou 1

Tabela 4: Lei da Probabilidade Total nas linhas

Tal procedimento também funcionará para a linha onde aparece a probabilidade $P(\overline{C_1})$ como resultado da Lei da Probabilidade Total. O resultado $P(\overline{C_1})$ aparece na mesma linha e à direita das probabilidades $P(\overline{C_1} \cap \overline{C_2})$ e $P(\overline{C_1} \cap C_2)$, e a soma dessas duas resulta na primeira.

Para as colunas, procedimento análogo funcionará. As probabilidades $P(\overline{C_2})$ e $P(C_2)$, que aparecem abaixo na tabela, são resultado da soma das duas probabilidades que aparecem acima delas em suas respectivas colunas. Isto é, aplicando a Lei da Probabilidade Total, tem-se $P(\overline{C_1} \cap \overline{C_2}) + P(C_1 \cap \overline{C_2}) = P(\overline{C_2})$ e $P(\overline{C_1} \cap C_2) + P(C_1 \cap C_2) = P(C_2)$.

Critério 2 Critério 1	$\overline{C_2}$	C_2	Total
$\overline{C_1}$	$P(\overline{C_1} \cap \overline{C_2})$	$P(\overline{C_1} \cap C_2)$	$P(\overline{C_1})$
C_1	$P(C_1 \cap \overline{C_2})$	$P(C_1 \cap C_2)$	$P(C_1)$
Total	$P(\overline{C_2})$	$P(C_2)$	100% ou 1

Tabela 5: Lei da Probabilidade Total nas colunas

4.1.3 Observância da Probabilidade Condicional

Para as probabilidades que envolvem os dois critérios simultaneamente, será necessário incorporar da relação da Probabilidade Condicional (Equação 4). Sem perda de generalidade, vamos considerar primeiro as probabilidades condicionais envolvendo C_1 e C_2 . Será possível obter essa probabilidade de dois modos possíveis. Se usarmos C_1 como condicional, teremos:

$$P(C_2|C_1) = \frac{P(C_1 \cap C_2)}{P(C_1)} \implies P(C_1 \cap C_2) = P(C_1) \times P(C_2|C_1).$$

Ou, se usarmos C_2 como condicional, teremos:

$$P(C_1|C_2) = \frac{P(C_1 \cap C_2)}{P(C_2)} \implies P(C_1 \cap C_2) = P(C_2) \times P(C_1|C_2).$$

Ambas as expressões acima usam probabilidade conjunta $P(C_1 \cap C_2)$ – que aparece na tabela –, e não parece haver nelas nenhum indício para dar preferência a uma ou outra. A escolha de uma delas só será possível considerando-se as informações do problema a ser solucionado. É neste ponto onde qualquer uma das probabilidades condicionais – $P(C_1|C_2)$ ou $P(C_2|C_1)$ – poderão ser usadas, seja como informação oferecida pelo contexto do problema ou como informação a ser obtida para solucionar o problema.

Quanto às probabilidades condicionais envolvendo $\overline{C_1}$ ou $\overline{C_2}$, basta proceder de modo análogo para obter alguma das probabilidades conjuntas restantes e que também aparecem na tabela, ou seja, $P(C_1 \cap \overline{C_2})$, $P(\overline{C_1} \cap C_2)$ ou $P(\overline{C_1} \cap \overline{C_2})$. Qualquer uma destas probabilidades conjuntas farão uso das probabilidades condicionais com $\overline{C_1}$ ou $\overline{C_2}$ e o mesmo tipo de escolha mencionado anteriormente se aplica aqui. Usar alguma das duas expressões dessas probabilidades conjuntas é uma decisão a ser tomada pelo contexto do problema, verificando-se as informações já obtidas ou a serem obtidas.

Encerra-se aqui a exposição do uso de tabelas de contingência. Dada a generalidade dessa exposição, exemplificar o uso da tabela fará com que a vantagem do seu ensino

Critério 2 Critério 1	$\overline{C_2}$	C_2	Total
$\overline{C_1}$	$P(\overline{C_1}) \times P(\overline{C_2} \overline{C_1})$ ou $P(\overline{C_2}) \times P(\overline{C_1} \overline{C_2})$	$P(\overline{C_1}) \times P(C_2 \overline{C_1})$ ou $P(C_2) \times P(\overline{C_1} C_2)$	$P(\overline{C_1})$
C_1	$P(C_1) \times P(\overline{C_2} C_1)$ ou $P(\overline{C_2}) \times P(C_1 \overline{C_2})$	$P(C_1) \times P(C_2 C_1)$ ou $P(C_2) \times P(C_1 C_2)$	$P(C_1)$
Total	$P(\overline{C_2})$	$P(C_2)$	100% ou 1

Tabela 6: Probabilidades Condicionais

seja mais facilmente percebida. Muitos problemas continuam podendo ser solucionados pelos cálculos tradicionais da probabilidade, numa disposição visual linear tradicional. Contudo, uma visão mais completa desses problemas será obtida pelo uso da tabela, além da própria disposição visual das informações e das relações entre elas ficar mais acessível.

Exemplo 5 (Estou gripado ou não, eis a questão). *Uma pessoa está com dor de cabeça e a garganta inflamada. Fazendo uma rápida pesquisa na internet, ela descobre que entre todas as pessoas que realmente estão gripadas, 90% delas tem esses mesmos sintomas. Além disso, essa mesma pessoa também descobre que, a cada ano, 5% da população pega gripe, e que 20% da população sente dores de cabeça e garganta inflamada em algum momento desse período. Qual é a probabilidade de essa pessoa realmente estar gripada?*

Relembrando a Definição [8](#), há que se reconhecer quais informações nesse problema garantem a existência de um espaço amostral e dos critérios para identificar partições desse espaço. Vejamos:

- *Qual o conjunto ou agrupamento que representa o espaço amostral, e quem são seus elementos?* Nesse problema, o objeto de estudo de probabilidade é a “pessoa”, que faz parte de um conjunto maior chamado de “população”. Portanto, os elementos do espaço amostral seriam pessoas, e o espaço amostral seria a população.
- *Que critérios foram usados para classificar os elementos desse espaço amostral, e como se expressam como partições?* Frases como “5% da população pega gripe” e “20% da população sente dores de cabeça e garganta inflamada” oferecem informações sobre uma parte do espaço amostral e, nisso, incluem em si mesmas algum tipo de critério. Na primeira frase, os elementos (pessoas) foram classificados pelo critério de “estar gripado”. Pela lógica básica, é impossível que uma pessoa esteja ao mesmo tempo gripada e não-gripada e, com esse critério, pode-se garantir que G (estar gripado) e \overline{G} (não estar gripado) constituem de fato uma partição do espaço amostral Ω (a população). Já na segunda frase, os elementos foram classificados pelo critério de “ter sintomas” e, analogamente à frase anterior, garante-se que S (ter sintomas) e \overline{S} (não ter sintomas) constituem uma outra possível partição de Ω .

Note que a probabilidade desejada, a da pessoa estar gripada, está condicionada ao fato de ela já ter sintomas. Desse modo, usando a notação simbólica estabelecida, o objetivo é calcular $P(G|S)$. Pelo Teorema de Bayes:

$$P(G|S) = \frac{P(G) \times P(S|G)}{P(S)}. \quad (9)$$

Façamos a interpretação das probabilidades do lado direito da igualdade acima. $P(G)$ se refere à probabilidade de uma pessoa qualquer da população estar gripada sem qualquer condição prévia, isto é, $P(G) = 5\% = 0.05$. De modo análogo, $P(S)$ representa a probabilidade de uma pessoa qualquer apresentar sintomas sem condições prévias, ou seja, $P(S) = 20\% = 0.20$ e, finalmente, $P(S|G)$ representa a probabilidade de uma pessoa já gripada ter os sintomas, isto é, $P(S|G) = 90\% = 0.90$. Usando estes valores no Teorema de Bayes, obtemos

$$P(G|S) = \frac{0.05 \times 0.90}{0.20} = 0.225$$

e portanto, a probabilidade de uma pessoa sintomática realmente estar gripada é de 22.5%, o que representa possivelmente um valor abaixo do que seria esperado sem uma devida análise.

Observe que, usando o Teorema de Bayes de modo direto, foi possível obter a probabilidade desejada sem grandes esforços. Contudo, é possível estudar o problema de modo mais amplo, se usarmos uma tabela e todo o processo descrito na Seção 4.1. Utilizando as informações oferecidas no problema e interpretando-os para a tabela, pode-se preencher uma porção dela como a seguir:

Gripe \ Sintomas	Sintomas		Total
	\bar{S}	S	
\bar{G}	?	?	?
G	?	$90\% \times 5\%$ $= 4.5\%$	5%
Total	?	20%	100%

Tabela 7: Tabela inicial para o Exemplo 5

Observe que só foi possível escrever $P(G \cap S) = 4.5\%$ por conta da Probabilidade Condicional $P(G \cap S) = P(G) \times P(S|G) = 0.05 \times 0.90 = 0.045$.

Depois desse preenchimento inicial e lembrando a probabilidade do evento complementar e da Lei da Probabilidade Total, torna-se possível preencher o restante da tabela e obter um quadro mais amplo do problema:

Por essa tabela, todas as probabilidades possíveis poderão ser obtidas. Voltando ao problema do exemplo, o objetivo era obter a probabilidade de uma pessoa realmente estar gripada, desde que ela tenha apresentado sintomas, isto é, $P(G|S)$. Essa probabilidade

Sintomas \	\bar{S}	S	Total
Gripe			
\bar{G}	79.5%	15.5%	95%
G	0.5%	4.5%	5%
Total	80%	20%	100%

Tabela 8: Tabela preenchida para o Exemplo 5

condicional, estando diretamente relacionada com a probabilidade conjunta $P(G \cap S)$ e com $P(S)$, ambas exibidas na tabela, é então calculada:

$$P(G|S) = \frac{P(G \cap S)}{P(S)} = \frac{0.045}{0.200} = 0.225.$$

Desse modo, o Teorema de Bayes foi usado de modo implícito. Além disso, qualquer uma daquelas probabilidades condicionais restantes (envolvendo \bar{G} ou \bar{S}) poderia ser calculada com a mesma facilidade. Um exemplo disso seria calcular também a probabilidade de uma pessoa que não apresenta sintomas também não estar gripada, isto é, $P(\bar{G}|\bar{S})$. Isso seria obtido usando a tabela num cálculo similar:

$$P(\bar{G}|\bar{S}) = \frac{P(\bar{G} \cap \bar{S})}{P(\bar{S})} = \frac{0.795}{0.800} = 0.99375.$$

Esse número indica que quase todas as pessoas (99.375%) que não tem aqueles sintomas também não estejam gripadas.

Vê-se, portanto, com o uso da tabela, que o cálculo das diversas combinações de probabilidades pode ser facilmente realizado, considerando-se as devidas observâncias teóricas da probabilidade e da própria disposição visual das informações. Para estudantes do ensino médio que visam aprovação em concursos vestibulares, a disposição visual facilitada das informações torna-se um fator mais importante para resolução de problemas nas provas abertas (popularmente chamadas de “canetão”) e, por isso, a tabela é uma opção bastante recomendável para atender ao critério de clareza de soluções.

4.2 DIMINUIR VIESES DE JULGAMENTO PROBABILÍSTICO COM FATOR DE BAYES

Um dos grandes benefícios do paradigma bayesiano da probabilidade é a possibilidade que ele oferece para evitar alguns vieses de julgamento, que *aparentam* ser naturais dada a representatividade dos elementos num espaço amostral.

4.2.1 Fator de Bayes em Tversky & Kahneman

O exemplo a seguir, inspirado em Tversky & Kahneman ([25]), busca demonstrar como o Fator de Bayes (Equação 8) torna mais viável levar em conta outros fatores para fazer algum julgamento probabilístico, diminuindo vieses. Isto é especialmente útil ao buscar analisar problemas socialmente ou culturalmente sensíveis.

Exemplo 6 (Qual é a profissão de Steve?). *Certo indivíduo foi descrito com as seguintes palavras: “Steve é muito tímido e retraído, sempre prestativo, mas com pouco interesse nas pessoas ou no mundo real. Uma alma mansa e íntegra, ele necessita de ordem e estrutura e possui uma paixão por detalhes”. Como as pessoas avaliariam a probabilidade de Steve ser ou não ser um cientista?*

Ao ouvir tal descrição de Steve, qual seria a primeira impressão, ou uma resposta majoritária das pessoas, para essa probabilidade? Considerando que há uma tendência de associar estereotipicamente aquelas características de personalidade a profissões que exijam um maior senso de organização, detalhes e alguma disciplina, é natural que a resposta à pergunta seja de que “é mais provável que Steve seja mesmo um cientista”. Afinal de contas, esse estereótipo é por muitas vezes retratado em filmes, novelas e na cultura popular. Este seria um julgamento probabilístico baseado naquilo que percebe através da cultura e, portanto, não pareceria irracional. A impressão inicial seria, portanto, de que Steve tem realmente mais chances de ser um cientista do que de não ser. Mas quanto?

Suponha que, após alguma discussão, as pessoas cheguem num acordo e digam que seja 3 vezes mais provável que Steve seja um cientista do que ele não seja. Isso sugere que, na visão dessas pessoas que discutiram, de cada 4 pessoas que tivessem aquela personalidade, 3 seriam cientistas. Na linguagem da probabilidade, este número seria reconhecido como a *razão de chances* de Steve ser um cientista. Portanto, levando em consideração a Definição 10, e que C seja o evento “ser cientista”, tem-se

$$O(C) = 3 = \frac{P(C)}{1 - P(C)} \implies P(C) = \frac{3}{4}.$$

Então, segundo aquelas pessoas, a probabilidade de Steve ser cientista seria de 75%.

No entanto, tudo isso seria um julgamento enviesado. Contudo, não seria enviesado em função dessa cultura que sugere que pessoas que tenham certas profissões tenham ou não certas características de personalidade, mas sim por levar *apenas* isso em consideração. Na avaliação da probabilidade de Steve ser cientista, quais poderiam ser outros fatores relevantes?

O fato de Steve ter características específicas de personalidade o faz ser parte de um subconjunto da população em geral. Considerando que T seja o evento de “possuir características como timidez, sempre prestativo, com pouco interesse em pessoas, alma mansa e íntegra”, a probabilidade de fato que se busca avaliar é, portanto, $P(C|T)$. Ao

avaliar essa probabilidade, se está avaliando também $P(\bar{C}|T)$, pois essas probabilidades estão relacionadas pela Lei do Evento Complementar (Equação 3). Além disso, elas estão relacionadas também pela Razão de Chances com Fator de Bayes (Equação 8). Esta sim será usada para explicitar ao menos um fator adicional na avaliação individual de $P(C|T)$:

$$\frac{P(C|T)}{P(\bar{C}|T)} = O(C|T) = \frac{P(C)}{P(\bar{C})} \times \frac{P(T|C)}{P(T|\bar{C})} = O(C) \times B(T|C). \quad (10)$$

O uso correto da razão de chances, observadas as probabilidades condicionais e as não-condicionais, explicita e relembra a necessidade de estabelecer qual é o espaço amostral sob o qual um observador “neutro” avaliaria as probabilidades não-condicionais. Um estudante atencioso, ao avaliar a relação matemática acima, e que fosse questionado sobre a probabilidade de Steve ser ou não ser cientista após ouvir as características de sua personalidade, notaria que ele próprio não é um observador neutro, e questionaria a impressão inicial das pessoas.

Para o estudante, seria necessário reavaliar a probabilidade se colocando como observador neutro. Ele entenderia aquele valor 3 sugerido pelas pessoas não como sendo uma razão de chances de ser um cientista, mas sim como o *fator de Bayes de uma razão de chances condicionada ao fato de ser um cientista* — ou seja, $B(T|C)$. Nisso, ele teria percebido dois pontos importantes. O primeiro deles é que o julgamento inicial das pessoas, na verdade, era baseado não no espaço amostral todo da população, mas sim que elas estavam pensando (isto é, condicionadas previamente) apenas nos cientistas do seu imaginário. Esse julgamento pode até ser considerado enviesado pela cultura segundo outros critérios, mas ao menos se toma consciência disso: o estereótipo dos cientistas no imaginário é de que existem cerca de 3 vezes mais cientistas tímidos do que não-tímidos. O estudante conseguiu perceber onde estão os julgamentos de probabilidade condicionados, algo que não estava tão claro na primeira impressão das pessoas.

O segundo ponto, e definidor da avaliação final da probabilidade, é que o valor de probabilidade que o estudante desejava obter não era simplesmente $P(C)$, mas sim $P(C|T)$, isto é, dentre todas as pessoas da população que sejam tímidas, qual a probabilidade de ela também ser um cientista? Percebendo agora que seu espaço amostral não consta apenas de um estereótipo de cientistas, o estudante precisará avaliar o fator $O(C) = \frac{P(C)}{P(\bar{C})}$, que se baseia no espaço amostral da população (ou seja, não-condicionado).

Suponha que, após uma pesquisa rápida na internet, o estudante tenha chegado à conclusão de que, em sua população, a proporção entre cientistas e não-cientistas seja de 7 para 10000. Em terminologia de probabilidade, isto seria precisamente a razão de chances $O(C)$ mencionada no parágrafo anterior, fator com o qual o estudante terá condições de avaliar as chances de Steve ser cientista possuindo uma personalidade tímida, conforme a Equação 10:

$$O(C|T) = O(C) \times B(T|C) = \frac{7}{10000} \times 3 = \frac{21}{10000} \implies P(C|T) = \frac{21}{10021}.$$

A probabilidade $P(C|T)$ acima tem um valor muito diferente ($\approx 0.21\%$) e muito menor, se comparado à probabilidade segundo a primeira impressão das pessoas.

O que o fator de Bayes possibilitou ao estudante fazer é *atualizar* a informação inicial multiplicando-a por esse fator. Uma interpretação não enviesada da condição T avaliaria a probabilidade desejada considerando C não como contexto geral de estudo (ou seja, como espaço amostral) do modo como as pessoas fizeram em sua impressão inicial, mas sim como subconjunto de algum outro espaço amostral “maior” que permita considerar a probabilidade dos elementos pertencerem ou não a C . Julgamentos de probabilidade que estejam assim estruturados tenderão a ser menos enviesados e mais objetivos, baseando-se não apenas em percepções culturais, mas também de fatores externos à percepção imediata do observador (como por exemplo, a proporção de certos elementos no espaço amostral).

4.2.2 Fator de Bayes em Gigerenzer et al.

Inspirado na experiência citada em Gigerenzer et al. ([26]), segue-se um exemplo bastante útil e importante de aplicação do Fator de Bayes, pelo seu contexto de diagnóstico médico.

Exemplo 7 (Câncer de mama). *Um ginecologista de determinada região recebe em seu consultório uma mulher, chamada Maria, que fez um teste de mamografia, reportando a ele que seu resultado foi positivo. A prevalência de câncer de mama naquela região é de 1%, isto é, de cada 100 mulheres na região, 1 tem câncer. Esse teste de mamografia tem sensibilidade de 90%, isto é, de cada 100 mulheres com câncer que são testadas, 90 delas teriam resultado positivo. Esse mesmo teste tem especificidade de 91%, isto é, de cada 100 mulheres sadias que fazem o teste, 91 delas teriam resultado negativo. Qual é a probabilidade de que Maria realmente tenha câncer?*

A experiência citada em Gigerenzer et al. ([26]) mostra que a primeira impressão é que a probabilidade de Maria ter câncer seja alta ou, pelo menos, que é mais provável que ela tenha câncer do que ela não tenha. Cerca de 60% dos ginecologistas daquela experiência julgaram que essa probabilidade era alta. Trata-se, portanto, não de pessoas do público em geral sem treinamento, mas sim pessoas com treinamento médico. Será que tal julgamento está correto? Maria deve ficar alarmada com seu resultado?

Chamando de T a probabilidade de uma mulher testar positivo no exame de mamografia e de C a probabilidade de uma mulher realmente ter câncer de mama, os valores de probabilidade citados no exemplo seriam identificados como $P(C) = 0.01$, $P(T|C) = 0.90$ e $P(\bar{T}|\bar{C}) = 0.91$. Invocando o Teorema de Bayes, juntamente com a Lei da Probabilidade Total, será possível avaliar a probabilidade de aquela mulher ter câncer caso ela teste positivo, ou seja, calcular $P(C|T)$:

$$P(C|T) = \frac{P(T|C)P(C)}{P(T)} = \frac{P(T|C)P(C)}{P(T|C)P(C) + P(T|\bar{C})P(\bar{C})} = \frac{O(C|T)}{O(C|T) + 1}. \quad (11)$$

Esse resultado usa a razão de chances com fator de Bayes (Equação 8), de onde se relembra que $O(C|T) = O(C)B(T|C) = \frac{P(C)}{P(\bar{C})} \frac{P(T|C)}{P(T|\bar{C})}$.

O fator $O(C)$, que representa a razão de chances de uma mulher daquela região ter câncer, é obtido:

$$O(C) = \frac{P(C)}{P(\bar{C})} = \frac{0.01}{1 - 0.01} = \frac{1}{99}.$$

Essa fração, segundo a interpretação bastante comum sugerida na Definição 10, é que naquela região há 1 mulher com câncer para cada 99 mulheres sadias. Esta proporção é uma boa opinião inicial sobre as chances de uma mulher ter câncer *sem quaisquer informações adicionais*, isto é, sem quaisquer outras condições ou restrições. Contudo, o fator de bayes $B(T|C)$ fará a *atualização* das chances de Maria (e não uma mulher qualquer) ter câncer, pois, *para ela, há informação específica*, representada no fato de ela testar positivo para câncer. Para Maria, por conta do seu teste positivo, tem-se:

$$B(T|C) = \frac{P(T|C)}{P(T|\bar{C})} = \frac{0.90}{1 - 0.91} = 10 \implies O(C|T) = \frac{1}{99} \times 10 = \frac{10}{99}.$$

Isto é, a informação nova representada pelo teste positivo da mamografia de Maria faz aumentar em 10 vezes a sua razão de chances de ter câncer. Dessa forma, pode-se finalmente obter a probabilidade real de ela ter câncer, retornando à Equação 11:

$$P(C|T) = \frac{O(C|T)}{O(C|T) + 1} = \frac{\frac{10}{99}}{\frac{10}{99} + 1} = \frac{10}{109} \approx 0.092.$$

A probabilidade real de Maria ter câncer é, portanto, de 9.2%. O julgamento inicial dos ginecologistas, cuja primeira impressão era de que a probabilidade dela ter câncer fosse alta, é errônea. Ainda com o resultado positivo do teste, não é possível afirmar que seja mais provável que Maria tenha câncer do que ela não tenha. Eis aí um bom motivo para se buscar matematicamente uma interpretação mais objetiva do que significam testes médicos: ainda que tais testes ofereçam porcentagens altas de sensibilidade (alta taxa de detecção) e de especificidade (baixa taxa de falsos positivos), somente uma avaliação bayesiana poderia afirmar com maior garantia se realmente há mais chances ou não de uma pessoa ter alguma enfermidade. Isto poderia amenizar fatores como estresse e ansiedade causados pelo desconhecimento do que o resultado positivo de um teste realmente quer dizer, matematicamente falando.

Se fosse construída uma tabela de contingência para avaliar todas as probabilidades envolvidas neste exemplo conforme as recomendações da Seção 4.1, o resultado seria aquele apresentado na Tabela 9. Por essa tabela, a mesma probabilidade seria obtida, porém num cálculo diferente de $P(C|T)$.

Câncer \ Teste	\bar{T}	T	Total
\bar{C}	90.09%	8.91%	99%
C	0.10%	0.90%	1%
Total	90.19%	9.81%	100%

Tabela 9: Tabela preenchida para o Exemplo 7

$$P(C|T) = \frac{P(C \cap T)}{P(T)} = \frac{0.0090}{0.0981} \approx 0.092.$$

4.3 CONSIDERAÇÕES FINAIS

Reforça-se aqui o uso da abordagem bayesiana da probabilidade (isto é, Teorema de Bayes e suas decorrências, como o fator de Bayes) para diminuir vieses de julgamento. No Exemplo 7 apresentado na seção anterior, diferentemente do Exemplo 6, percebe-se a utilidade da abordagem bayesiana para avaliação de questões não só culturais, mas também na saúde. O ensino da inferência bayesiana, pelo uso de informações adicionais para julgamentos probabilísticos, parece atender às competências listadas na BNCC para a Matemática e Suas Tecnologias, especialmente os itens 1, 2 e 3 [1, p. 531]. Chama a atenção o item 2 desse documento, com menção específica a “situações de saúde” e, relembrando o Exemplo 7, há legitimidade nele por representar um desafio do mundo contemporâneo, suscitando tanto profissionais de saúde como cidadãos comuns a “tomar decisões éticas e socialmente responsáveis”.

Além disso, a inferência bayesiana utilizada no capítulo anterior também precisaria fazer parte do ferramental matemático à disposição do estudante. A simulação do experimento mental de Bayes ilustrou como novas informações podem ser incorporadas a um modelo probabilístico para atualização dos valores de probabilidade. O uso constante do Teorema de Bayes, aliado ao acréscimo de informações (como por exemplo, a impossibilidade de certo resultado) foram fundamentais para que opiniões iniciais – possivelmente formuladas pela ignorância – sobre certo fenômeno fossem atualizadas e pudessem convergir para opiniões mais objetivas e baseadas no que fosse conhecido de fato. É dessa maneira que, ao avaliar certas hipóteses dentro de um conjunto de possíveis explicações de um fenômeno, algumas delas possam ser escolhidas ou eliminadas, de acordo com as probabilidades que cada uma assumir incorporando informações das evidências relevantes. Relembrando o item 3 da BNCC para a Matemática e Suas Tecnologias [1, p. 531], a inferência bayesiana corresponderia a uma estratégia de resolução de problemas em diversos contextos, dada sua aplicabilidade quase universal nas ciências. A plausibilidade dos resultados obtidos na inferência bayesiana estaria relacionada à adequação inicial do modelo probabilístico. Sugere-se que se estabeleçam nesse modelo, por exemplo, espaços amostrais que correspondam a possibilidades reais e, também, que se estabeleçam as par-

tições desse espaço amostral de modo a obedecer às definições teóricas da probabilidade, como se buscou mostrar nos exemplos desta monografia.

A simulação do experimento mental de Bayes serviu para demonstrar a plausibilidade da inferência bayesiana como um processo de sistematização de novas informações, e também que o paradigma frequencista não é um rival do bayesianismo. Os paradigmas tem suas utilidades conforme os propósitos e as limitações inerentes à própria pesquisa. A subjetividade do processo bayesiano não está no processo, mas sim como fase inicial da investigação, enquanto ainda não há ainda informações suficientes para diminuir incertezas e superar a subjetividade dos investigadores.

CÓDIGO-FONTE DA SIMULAÇÃO

```

1 from random import randint as aleatorio
2
3 while(True):
4     print("=== ALGORITMO BAYESIANO ===")
5     print("Para sair do programa, basta apertar ENTER sem fornecer informacao
        util ao programa em qualquer momento.")
6     try:
7         n_reg = int(input("Numero de regioes da mesa: "))
8         marc = aleatorio(1, n_reg)
9         print(f"A mesa foi dividida igualmente em regioes numeradas de 1 a {n_reg}
            }, da esquerda para a direita.")
10        print("(quanto mais proximo de 1, a regioao estah mais a esquerda)")
11        print(f"(quanto mais proximo de {n_reg}, a regioao estah mais a direita)")
12        print("A bola foi lancada na mesa. O ajudante marcou a posicao dela em
            uma dessas regioes.\n")
13        n_esq_t, n_dir_t, n_lanc_t = 0, 0, 0
14        while(True):
15            print("SEU OBJETIVO: descobrir em qual regioao a marcacao foi feita.")
16            print("Agora o ajudante vai lancar a bola quantas vezes voce desejar
                para obter informacao.")
17            print("Depois desses lancamentos, ele lhe dira em quantos deles a bola
                ficou a ESQUERDA ou a DIREITA da marcacao.")
18            n_esq, n_dir, n_lanc = 0, 0, int(input("\nAJUDANTE: Quantas vezes devo
                lancar a bola? "))
19            for i in range(n_lanc):
20                reg_lanc = aleatorio(1, n_reg)
21                n_esq, n_dir = n_esq+(reg_lanc<marc)*1, n_dir+(reg_lanc>marc)*1
22            n_esq_t, n_dir_t, n_lanc_t = n_esq_t+n_esq, n_dir_t+n_dir, n_lanc_t+n_lanc
23            print(f"AJUDANTE: A bola foi lancada {n_lanc} vez(es) agora. Em relacao
                a marcacao,")
24            print(f"{n_esq} foram a esquerda ({100*n_esq/n_lanc:.1f} % desses
                lancamentos)")
25            print(f"{n_dir} foram a direita ({100*n_dir/n_lanc:.1f} % desses
                lancamentos)")
26            print(f"\nAJUDANTE: Total de {n_lanc_t} lancamentos, sendo:")
27            print(f"{n_esq_t} a esquerda ({100*n_esq_t/n_lanc_t:.1f} % do total)")
28            print(f"{n_dir_t} a direita ({100*n_dir_t/n_lanc_t:.1f} % do total)")
29            tent = int(input("\nAJUDANTE: Em qual regioao voce acha marcacao foi
                feita? "))
30            if tent == marc:
31                print(f"AJUDANTE: VOCE ACERTOU! A marcacao foi feita na regioao {marc}.\n")

```

```
32     break
33 else:
34     print(f"AJUDANTE: VOCE ERROU. A marcacao nao foi feita na regioao {tent
        }. Vamos tentar de novo?\n")
35 except:
36     print("=== FIM ===")
37     break
```

SUGESTÃO DE PLANO DE AULA COM DADO DE 6 FACES

B.1 OBJETIVOS

1. **Gerais:** Compreender o Teorema de Bayes como ferramenta de inferência estatística.
2. **Específicos:**
 - (a) Aplicação do Teorema de Bayes em problemas do dia-a-dia.
 - (b) Entender como o Teorema de Bayes facilita a sistematização da busca de informações para fazer inferência estatística em meio a incerteza.

B.2 CONTEÚDOS

1. Lei da Probabilidade Total
2. Teorema de Bayes

B.3 RECURSOS

1. Quadro negro, apagador e giz; ou quadro branco, apagador e pincel.
2. Dois dados de 6 faces (preferencialmente, cada dado de cor diferente um do outro).
3. Caixa de papelão.

B.4 PROCEDIMENTOS

1. **Introdução:**
 - (a) Revisão da Lei da Probabilidade Total (LPT).
 - (b) Dedução do Teorema de Bayes a partir da fórmula da Probabilidade Condicional.
 - (c) Explicação do significado do teorema enquanto “probabilidade inversa”, ou “probabilidade de causas”.

2. **Desenvolvimento:** Reprodução do experimento mental de Bayes, adaptado para dados de 6 faces.

- (a) O professor, em posse dos dois dados, usa um deles para sortear um número de 1 a 6, que precisará ser adivinhado pelos alunos. Após o sorteio e sem os alunos estarem cientes de qual número foi sorteado, o professor esconde esse dado (dado escondido), que só será revelado quando os alunos adivinharem o número sorteado. O dado restante será usado para outros lançamentos (dado de uso).
- (b) Um aluno é escolhido para fazer novos lançamentos com o dado de uso.
- (c) O professor coloca a caixa de papelão com a abertura pra cima em um local na sala de aula de forma que nenhum aluno consiga ver o interior da caixa. O aluno escolhido lança o dado de uso nessa caixa. O professor, sendo o único que tem acesso à caixa, informa a todos os alunos se o resultado desse lançamento é “maior”, “menor” ou “igual” ao número sorteado.
- (d) Os alunos, em posse da nova informação, são orientados a calcular a probabilidade de o número sorteado ser 1, 2, 3, 4, 5 ou 6, fazendo uso da LPT e do Teorema de Bayes.
- (e) Com base nessas probabilidades, o aluno escolhido, em nome de todos os alunos, informa ao professor qual número eles consideram ser o número sorteado. O professor diz se o número informado pelo aluno está correto ou não.
- (f) Caso o número informado não esteja correto, o professor não revela o dado escondido, e este processo retorna à etapa (c) para fazer um novo lançamento do dado de uso e colher mais informações do professor.
- (g) Caso o número informado pelo aluno seja correto, o professor revela o dado escondido, comprovando que os alunos acertaram, e o processo é finalizado.

3. **Conclusão:**

- (a) Sugere-se que o professor oriente os alunos em todo o processo descrito no item anterior interpretando as probabilidades obtidas em cada passo e explicando o que elas significam, mostrando como realizar inferência bayesiana e, por conseguinte, a tomada de decisão dos alunos em meio a incerteza (ao tentarem adivinhar o resultado) pode ser racionalmente justificada pela probabilidade.
- (b) Caso haja tempo a disposição do professor, ele poderá reproduzir o experimento mental adaptado para dados de 6 faces uma segunda vez, dessa vez sem que ele próprio dê orientações e propondo que os alunos por si mesmos interpretem as probabilidades obtidas.

B.5 AVALIAÇÃO

Ao final da aula, o professor poderá deixar como tarefa de casa uma lista com 2 tipos de exercícios, cada qual com uma finalidade específica:

1. Memorização o Teorema de Bayes: oferecer valores numéricos de 3 dos 4 termos que aparecem na fórmula – $p(A)$, $p(B)$, $p(B|A)$ e $p(A|B)$ –, e pedir o valor numérico do outro termo.
2. Reconhecimento do teorema em problemas descritivos: narrar alguma situação mais próxima do cotidiano para exigir do aluno que ele reconheça quais informações representam os termos $p(A)$, $p(B)$, $p(B|A)$ e $p(A|B)$ do teorema, com finalidade de calcular outro termo que tenha significado prático no problema. Dois possíveis exemplos são os já mencionados anteriormente, Exemplo 5 e Exemplo 7.

BIBLIOGRAFIA

- [1] Ministério da Educação do Brasil. *Base Nacional Comum Curricular*. 2018. URL: http://basenacionalcomum.mec.gov.br/images/BNCC_EI_EF_110518_versaofinal_site.pdf.
- [2] Bradley Efron. “Bayes’ Theorem in the 21st Century”. Em: *Science* 340.6137 (2013), pp. 1177–1178. ISSN: 0036-8075. DOI: [10.1126/science.1236536](https://doi.org/10.1126/science.1236536).
- [3] 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.
- [4] Sharon Bertsch McGrayne. *Why Bayes Rules: The History of a Formula That Drives Modern Life*. URL: <https://www.scientificamerican.com/article/why-bayes-rules/>.
- [5] 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](https://doi.org/10.1098/rstl.1763.0053).
- [6] Académie Royale des Sciences (France). *Mémoires de mathématique et de physique, présentés à l’Académie royale des sciences, par divers sçavans & lûs dans ses assemblées*. Vol. 6. Paris, 1774, pp. 621–656. URL: <https://www.biodiversitylibrary.org/page/26738365>.
- [7] Académie Royale des Sciences (France). *Mémoires de mathématique et de physique, présentés à l’Académie royale des sciences, par divers sçavans & lûs dans ses assemblées*. Vol. 7. Paris, 1776, pp. 37–232. URL: <https://www.biodiversitylibrary.org/page/27019507>.
- [8] Pierre Simon Laplace. *A Philosophical Essay on Probabilities*. URL: <https://www.gutenberg.org/ebooks/58881>.
- [9] *The Empire of Chance: How Probability Changed Science and Everyday Life*. Cambridge University Press, 1989. DOI: [10.1017/CB09780511720482](https://doi.org/10.1017/CB09780511720482).
- [10] Karl Pearson. “Laplace”. Em: *Biometrika* 21.1-4 (1929), pp. 202–216. DOI: [10.1093/biomet/21.1-4.202](https://doi.org/10.1093/biomet/21.1-4.202).
- [11] Arthur Bailey. “Credibility procedures: Laplace’s generalization of Bayes’ rule and the combination of collateral knowledge with observed data”. Em: *Proceedings of the Casualty Actuarial Society* 37.67 (1950). URL: https://www.casact.org/sites/default/files/database/proceed_proceed50_1950.pdf.

- [12] Karl Pearson. “The fundamental Problem of practical Statistics”. Em: *Biometrika* 13.1 (1920). DOI: [10.2307/2331720](https://doi.org/10.2307/2331720).
- [13] Donald A. MacKenzie. *Statistics in Britain 1865–1930: The Social Construction of Scientific Knowledge*. Edinburgh University Press, 1981.
- [14] Ronald A. Fisher. *Statistical Methods For Research Workers*. Oliver e Boyd, 1950.
- [15] Wilfred Perks. “Some observations on inverse probability including a new indifference rule”. Em: *Journal of the Institute of Actuaries* 73.2 (1947), pp. 285–334. DOI: [10.1017/S0020268100012270](https://doi.org/10.1017/S0020268100012270).
- [16] “Notices respecting new books”. Em: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 46 (1923), pp. 1021–1025. DOI: [10.1080/14786442308565259](https://doi.org/10.1080/14786442308565259).
- [17] I. J. Good. “Studies in the History of Probability and Statistics. XXXVII A. M. Turing’s Statistical Work in World War II”. Em: *Biometrika* 66.2 (1979). DOI: [10.1093/biomet/66.2.393](https://doi.org/10.1093/biomet/66.2.393).
- [18] C. E. Shannon. “A Mathematical Theory of Communication”. Em: *Bell System Technical Journal* 27.3 (1948), pp. 379–423. DOI: [10.1002/j.1538-7305.1948.tb01338.x](https://doi.org/10.1002/j.1538-7305.1948.tb01338.x).
- [19] A. Smith. “A Conversation with Dennis Lindley”. Em: *Statistical Science* 10.3 (1995), pp. 305–319. DOI: [10.1214/ss/1177009940](https://doi.org/10.1214/ss/1177009940).
- [20] D. V. Lindley. “Statistical Inference”. Em: *Journal of the Royal Statistical Society: Series B (Methodological)* 15.1 (1953), pp. 30–65. DOI: [10.1111/j.2517-6161.1953.tb00123.x](https://doi.org/10.1111/j.2517-6161.1953.tb00123.x).
- [21] A. N. Kolmogorov. *Foundations of the theory of probability*. 2^a ed. Chelsea Publishing Co., 1956.
- [22] Rudolf Carnap. *Logical Foundations of Probability*. 2^a ed. The University of Chicago Press, 1962.
- [23] Jay L. Devore. *Probabilidade e Estatística para Engenharia e Ciências*. São Paulo: Cengage Learning, 2011.
- [24] Morris H. DeGroot e Mark J. Schervish. *Probability and Statistics*. 4^a ed. Addison-Wesley, 2011.
- [25] Amos Tversky e Daniel Kahneman. “Judgment under Uncertainty: Heuristics and Biases”. Em: *Science* 185 (1974), pp. 1124–1131. DOI: [10.1126/science.185.4157.1124](https://doi.org/10.1126/science.185.4157.1124).
- [26] Gerd Gigerenzer et al. “Helping Doctors and Patients Make Sense of Health Statistics”. Em: *Psychological Science in the Public Interest* 8.2 (2007), pp. 53–96. DOI: [10.1111/j.1539-6053.2008.00033.x](https://doi.org/10.1111/j.1539-6053.2008.00033.x).