Universidade de Brasília

INSTITUTO DE CIÊNCIAS HUMANAS

PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA

LEONILSON GOMES DE SOUZA

O BAYESIANISMO NA ESTATÍSTICA E NA FILOSOFIA DA CIÊNCIA – PROBLEMAS E APLICAÇÕES

Brasília/DF
2024

LEONILSON GOMES DE SOUZA

O BAYESIANISMO NA ESTATÍSTICA E NA FILOSOFIA DA CIÊNCIA – PROBLEMAS E APLICAÇÕES

Dissertação apresentada ao Programa de Pós-graduação em Filosofia, da Universidade de Brasília (UnB), como requisito para a obtenção do título de Mestre em Filosofia.

Área de concentração: Epistemologia, Lógica e Metafísica.

Orientador: Prof. Dr. Agnaldo Cuoco Portugal

Brasília/DF
2024

BANCA EXAMINADORA

____

Prof. Dr. Agnaldo Cuoco Portugal (UnB)
Orientador

____

Prof. Dr. Nelson Gonçalves Gomes (UnB)
Examinador Interno

____

Prof. Dr. André Luiz de Almeida Lisbôa Neiva (UFAL)
Examinador Externo

____

Prof. Dr. Alexandre Costa-Leite (UnB)
Examinador Suplente

Brasília, _____de ___________de 2024

*__À Santíssima Virgem Maria__, que "embora levasse uma vida comum a todos aqui na terra, cheia de preocupações e trabalhos familiares, estava sempre intimamente unida ao seu Filho e de uma forma inteiramente única cooperou na obra do Salvador". "Com sua caridade maternal cuida destes irmãos do seu Filho que ainda estão na sua peregrinação terrena e permanecem envolvidos em perigos e angústias até serem conduzidos à pátria bem-aventurada" (*Decreto ***Apostolicam Actuositatem****)*.

Probabilidade, arte que revela o cosmos em seu mistério profundo,
não exclui o divino, mas, ao contrário, o exalta,
pois Deus, em Sua sabedoria infinita,
tece a trama da vida, ocultando milagres,
como um véu sutil, que nos rodeia,
nos convidando a crer, mesmo entre as sombras,
na luz que brilha em cada instante,
na graça que permeia o cotidiano.
Assim, no entrelaçar do acaso e da fé,
percebemos o toque do Criador,
que em cada detalhe se faz presente,
despertando em nós a admiração,
e a certeza da presença do sagrado
nas probabilidades do nosso viver.
Autor desconhecido

Probabilidade é o conceito mais importante da ciência moderna, especialmente porque ninguém tem a menor noção do que isso significa.
Bertrand Russell

RESUMO

O teorema de Bayes, formulado no século XVIII, não seguiu uma trajetória contínua de desenvolvimento dentro da ciência. Por um longo período, o subjetivismo inerente à abordagem inferencial associada a esse teorema limitou sua aceitação nos principais periódicos científicos, com a probabilidade subjetiva sendo frequentemente alvo de críticas. Nos tempos modernos, no entanto, o avanço tecnológico e o consequente aumento da capacidade de processamento computacional, entre outros fatores, possibilitaram uma redescoberta e valorização da vasta aplicabilidade e da robustez que o teorema de Bayes pode oferecer para o raciocínio em situações de incerteza. Este trabalho visa explorar o desenvolvimento histórico dessa teoria e discutir seu impacto e crescente sucesso nas discussões filosóficas. No primeiro capítulo, apresentamos a teoria da probabilidade sob uma perspectiva matemática, estruturada para oferecer uma compreensão introdutória dos conceitos fundamentais que sustentam a teoria da probabilidade. No segundo capítulo, exploramos a probabilidade e a estatística sob a ótica da filosofia da ciência, abrangendo tópicos relevantes com ênfase em temas centrais para o frequencismo e o bayesianismo. Esperamos que esta exposição possa contribuir para uma compreensão mais clara acerca das implicações decorrentes do uso da estatística em problemas filosóficos.

Palavras-chave: Teorema de Bayes. Bayesianismo. Probabilidade Subjetiva.

ABSTRACT

Bayes' Theorem, formulated in the 18th century, did not follow a continuous trajectory of development within science. For a long period, the subjectivism inherent in the inferential approach associated with this theorem limited its acceptance in major scientific journals, with subjective probability often being the target of criticism. In modern times, however, technological advancements and the consequent increase in computational processing power, among other factors, have enabled a rediscovery and appreciation of the vast applicability and robustness that Bayes' Theorem can offer for reasoning in situations of uncertainty. This work aims to explore the historical development of this theory and discuss its impact and growing success in philosophical discussions. In the first chapter, we present probability theory from a mathematical perspective, structured to provide an introductory understanding of the fundamental concepts underpinning probability theory. In the second chapter, we explore probability and statistics from the perspective of the philosophy of science, covering relevant topics with an emphasis on central themes for frequentism and Bayesianism. We hope that this exposition can contribute to a clearer understanding of the implications arising from the use of statistics in philosophical problems.

Keywords: Bayes' theorem. Bayesianism. Subjective Probability.

Sumário

INTRODUÇÃO

O objetivo principal deste texto é analisar a teoria bayesiana da probabilidade, sua aplicação na filosofia, e compará-la com a estatística frequencista. A razão fundamental para essa comparação é transpor para a filosofia as questões discutidas e resolvidas na estatística, ampliando assim a compreensão de como a aplicação de métodos estatísticos pode contribuir para a resolução de problemas filosóficos relevantes. Esta abordagem visa mostrar que os métodos estatísticos, quando adequadamente interpretados e aplicados, podem fornecer insights valiosos para debates filosóficos, promovendo um entendimento mais profundo e rigoroso das questões epistemológicas e metodológicas.

Esta investigação fundamenta-se predominantemente na teoria das probabilidades, uma área com aplicações estabelecidas em estatística e que vem ganhando espaço na filosofia. Ao longo deste texto, torna-se evidente que existe uma certa confusão sobre o tema e sua aplicação, indicando que parte dessa confusão é de natureza conceitual. Isso ressalta a necessidade de uma compreensão aprofundada de vários conceitos fundamentais em probabilidade e estatística. Por exemplo, o conceito de frequência pode variar consideravelmente dependendo do contexto em que é considerado. É imperativo que uma investigação sobre um tema essencial como o bayesianismo incorpore uma seção aprofundada sobre a teoria das probabilidades. Tal seção é crucial para promover uma discussão eficaz sobre o assunto em questão.

É importante diferenciar estatística frequencista (ou clássica) de probabilidade frequencista, apesar de sua inter-relação. Entender essas distinções ajudará na compreensão dos conceitos subsequentes. No entanto, deve-se observar que não existe uma uniformidade nos textos teóricos sobre os conceitos aqui destacados, de modo que diferentes interpretações podem ser encontradas em diversos textos. O motivo para adotar uma apresentação particular desses temas ficará claro ao longo deste texto.

Além das distinções comuns em livros didáticos que abordam a teoria da probabilidade, incluindo as teorias clássica, frequencista, axiomática e bayesiana, os textos de filosofia frequentemente abordam distintos conceitos de probabilidade como físico e epistêmico.

Essas classificações têm o propósito de facilitar a apresentação e discussão dos temas, criando agrupamentos com propriedades comuns que auxiliam na compreensão. A escolha desses agrupamentos é influenciada tanto pelo desenvolvimento histórico da disciplina quanto, inevitavelmente, pelas preferências pessoais do autor.

Quanto à distinção entre probabilidade física e probabilidade epistêmica, apesar da falta de consenso sobre essa dicotomia, a escolha parece alinhar-se às principais discussões na área, além de ajudar a entender seu desenvolvimento subsequente. Este aspecto será apenas brevemente mencionado aqui, com uma discussão mais detalhada reservada para a segunda parte deste texto.

Os termos utilizados geralmente possuem uma razão específica e tendem a refletir seu significado. O termo *probabilidade física*, como o próprio nome sugere, está relacionado com situações que envolvem a compreensão física do objeto de estudo ou com eventos físicos em sistemas aleatórios que, sob certas condições, convergem a uma taxa específica. Para simplificar, consideremos o exemplo de uma moeda com duas faces idênticas[1]. Quando esta moeda é lançada, é intuitivo pensar que cada lado tem igual probabilidade de terminar voltado para cima. Portanto, a probabilidade de a face 1 ficar para cima é de 1 em 2, e o mesmo vale para a face 2.

---

[1] Se aprofundarmos a discussão, notamos que deveríamos fazer comentários adicionais sobre a forma do lançamento, a densidade homogênea da moeda, simetria etc. Não vamos nos ater a todos esses detalhes para que possamos ter uma leitura mais agradável e para que o texto possa ser adequadamente compreendido.

Similarmente, como será explorado mais adiante sob a ótica da probabilidade frequencista, observa-se que ao aumentar o número de lançamentos da moeda, a frequência com que cada face termina voltada para cima tende a se aproximar de metade do total de lançamentos. Importante notar que isso não significa que o total de faces 1 e o total de faces 2 estão cada vez mais próximos numericamente[2] e sim que a fração que relaciona o número de faces 1 (ou faces 2) sobre o total se aproxima de $\frac{1}{2}$.

Portanto, a chamada probabilidade física está relacionada à compreensão dos objetos físicos em análise ou a sistemas aleatórios que tendem a convergir para uma taxa específica. Por outro lado, a probabilidade epistêmica liga-se ao grau de crença de um indivíduo na ocorrência de um evento, baseando-se no conhecimento e informações disponíveis para essa pessoa[3]. Conforme Romeijn (2022) aponta, essa forma de probabilidade está relacionada à disposição de agir com base em tais suposições. Embora não exista uma correspondência direta, como exploraremos na segunda parte deste texto, esses dois tipos estão associados às duas principais abordagens estatísticas: a estatística clássica e a bayesiana.

A probabilidade epistêmica, vale notar, pode ser aplicada a qualquer tipo de afirmação e não necessita estar vinculada a processos aleatórios, simetrias, ou propensões, servindo para representar a plausibilidade subjetiva ou o grau em que uma afirmação é sustentada pela evidência disponível, tornando-se assim, como será discutido, mais apta para uso filosófico.

Como será mais bem discutido adiante, a teoria da probabilidade desempenha um papel central no raciocínio analítico, influenciando o pensamento. Embora exista um debate filosófico sobre o tema, parece haver uma relação direta entre o raciocínio correto e os princípios

---

[2] A intuição sugere que à medida que lançamos uma moeda repetidamente o número de ocorrências de *cara* tende a se igualar ao número de ocorrências de *coroa*. No entanto, essa percepção não reflete com precisão a realidade probabilística subjacente ao processo. Para um entendimento mais aprofundado e rigoroso deste fenômeno, recomenda-se a leitura do terceiro capítulo da obra de Feller (1968).

[3] Essa é apenas uma simplificação. A depender do autor, o conceito de probabilidade epistêmica apresenta uma variedade de nuances, alguns serão discutidos na segunda parte deste texto.

fundamentais da teoria da probabilidade. Isso explica, em parte, por que a probabilidade é relevante em diversas áreas de estudo.

Segundo Hájek (2002), os fundamentos da probabilidade estão conectados a questões centrais nos campos científico, social e filosófico, sendo sua interpretação um dos problemas fundamentais. Existem várias maneiras de abordar o tema, e a escolha muitas vezes depende do objetivo do texto e do gosto pessoal do autor. Aqui, opta-se por uma abordagem até certo ponto cronológica, com discussões históricas pontuais, o que apresenta grandes vantagens para a compreensão.

Apresentar o tema conforme surgiu historicamente ajuda a clarificar os problemas e as escolhas feitas. Frequentemente, percebe-se que confusões resultam de mudanças nas interpretações ao longo do tempo e essa forma de abordagem colabora para reduzir essas confusões, além de oferecer uma visão holística do tema.

Outro aspecto crucial, que merece ser antecipado, é que a construção axiomática da probabilidade, a ser discutida posteriormente, geralmente se baseia na teoria de conjuntos ou de proposições.[4] A escolha entre essas abordagens depende em grande parte da linha adotada pelo autor e de seu objetivo com o texto.[5]

Neste texto, optou-se pela construção baseada na teoria dos conjuntos, embora a visão proposicional seja utilizada de forma pontual na segunda parte, para melhor contextualizar o leitor sobre aspectos específicos do seu uso filosófico.

Compreender a estrutura deste texto requer atenção às numerosas decisões tomadas durante sua elaboração, dada a extensão e complexidade de cada parte do conteúdo. Optou-se por incluir elementos considerados essenciais para melhor compreensão da teoria da

---

[4] Satisfazer certa proposição equivale a pertencer a um determinado conjunto.
[5] Embora a discussão axiomática a partir da teoria de conjuntos seja a adotada por Kolmogorov e repetida por grande parte dos matemáticos, sua construção a partir de proposições apresenta vantagens em algums discussões filosóficas e por isso adotada por parte desses textos.

probabilidade e para seu adequado uso nas discussões filosóficas apresentadas na segunda parte. Conforme antecipado, adotamos uma abordagem histórica da teoria da probabilidade, explorando suas principais evoluções matemáticas até a formulação da teoria axiomática, momento em que o teorema de Bayes é introduzido. A decisão por introduzir as discussões sobre o teorema de Bayes no final da primeira parte, apesar de seu surgimento histórico anterior, teve o objetivo de aproveitar a teoria matemática formal para a formulação de um teorema mais sólido, garantindo a robustez dos fundamentos que sustentam essa parte da teoria, crucial para o desenvolvimento das discussões subsequentes.

A próxima seção deste texto examinará o uso filosófico do bayesianismo, discutindo suas teses centrais relacionadas à gradação da crença, à aderência dos agentes racionais ao cálculo de probabilidades e à atualização de crenças conforme os princípios de condicionalização. Nesse ponto, o alicerce construído na primeira parte do texto permitirá uma navegação mais clara pela teoria, evitando as confusões usuais que surgem devido às diferentes definições de conceitos na teoria matemática e filosófica, problema comum quando o estudo é realizado a partir de textos diversos. Antes, porém, retornaremos ao tema da probabilidade, para estabelecer o alicerce da discussão que se seguirá.

1 TEORIA DA PROBABILIDADE

Retornemos ao ponto de partida de nossa discussão. Como relata Cramér (1973), a teoria da probabilidade, atualmente um ramo relevante da matemática pura, com vastas aplicações em praticamente todas as áreas das ciências naturais, técnicas e sociais, teve origens bastante modestas. Suas raízes remontam a um estudo significativamente mais elementar do que o que conhecemos hoje, situando-se no contexto dos jogos de azar, uma prática apreciada há mais de quatro séculos.

Embora Girolamo Cardano e Galileo Galilei sejam amplamente reconhecidos por suas contribuições ao desenvolvimento inicial da probabilidade, é geralmente aceito que a teoria da probabilidade teve seu verdadeiro início com o trabalho de dois eminentes matemáticos franceses: Blaise Pascal (1623-62) e Pierre de Fermat (1601-65) (Tabak, 2016).

Pierre de Fermat, formado em direito e atuante como advogado, dedicava-se à matemática e a outras áreas do conhecimento em seu tempo livre (Tabak, 2016). Dada a complexidade dos temas que estudava, Fermat mantinha-se informado e aprimorava suas ideias por meio de correspondências com os mais destacados matemáticos de sua época. Muitas dessas cartas, ainda preservadas, revelam um homem modesto e empenhado em sua contínua busca pelo progresso matemático (Tabak, 2016).

Em contraste, Blaise Pascal participou de um dos mais renomados clubes de matemática da história, onde teve a oportunidade de aprender com alguns dos melhores matemáticos da Europa (Tabak, 2016). Desde a adolescência, Pascal demonstrou notável habilidade matemática e teve a oportunidade de conhecer Chevalier de Méré, com quem discutiu a base matemática de certos problemas relacionados a jogos (Tabak, 2016). Pascal procurou a ajuda de Fermat para resolver muitos desses problemas e, em 1654, iniciaram uma célebre troca de correspondências sobre jogos de azar (Tabak, 2016).

1.1 TEORIA CLÁSSICA DA PROBABILIDADE

Alguns dos problemas que Pascal e Fermat discutiram diziam respeito ao *problema da divisão de apostas*. A ideia é bastante simples. Suponha que dois jogadores façam apostas iguais em um jogo de azar. Suponha ainda que um jogador saia na frente do outro e então decida parar de jogar antes do final. Como deveriam ser divididos os valores das apostas? Para uma maior clareza, tomemos como referência um exemplo clássico discutido amplamente por diversos autores.

Dois jogadores participam de uma série de partidas justas, cada um com igual probabilidade de vencer, até que um deles alcance seis vitórias. Por motivos alheios ao jogo, este é interrompido quando um dos jogadores tinha cinco vitórias e o outro duas vitórias. Como deveria ser feita a divisão justa do valor apostado? A primeira impressão é a de que a divisão mais justa seria $\frac{2}{7}$ para o jogador com duas vitórias e $\frac{5}{7}$ para o jogador com cinco vitórias. Essa foi, inclusive, a solução de Paciuolo (1445-1517), conforme informa Gårding (1977).

Para Pascal e Fermat, o que outros viam como um problema de proporção era, na verdade, um problema de probabilidades. A solução por eles proposta resulta em $\frac{1}{16}$ para o jogador com duas vitórias e $\frac{15}{16}$ para o jogador com cinco vitórias. A solução de Fermat considerou todos os resultados possíveis de quatro jogadas (Gårding, 1977). Há $2 \times 2 \times 2 \times 2 = 16$ resultados possíveis, e o primeiro jogador vence em todos os casos, exceto em um, onde todas as quatro jogadas favorecem o oponente.

É relevante destacar que a solução proposta emprega uma fórmula simples em seu cálculo: a razão entre o número de casos favoráveis e o número de casos possíveis. Este tema será explorado de maneira mais detalhada quando abordarmos a definição clássica de probabilidade.

Antes de avançarmos, é importante ressaltar que esse raciocínio foi criticado, com base no fato de que nem todos os resultados devem ser jogados até o fim, por exemplo, quando o primeiro jogador vence a primeira partida. No entanto, esta crítica foi adequadamente rebatida, argumentando que nada muda se todos os resultados forem jogados até o fim (Gårding, 1977).

Se ainda restarem dúvidas, considere qual seria a probabilidade de o jogador com menos vitórias ganhar todas as partidas restantes (a única forma de ser vitorioso). Para que isso ocorra, haveria apenas um caso favorável em 16 possíveis $(2 \times 2 \times 2 \times 2)$, resultando em $\frac{1}{16}$. Esse resultado também pode ser obtido pelo produto das probabilidades $(\frac{1}{2})$ de todas as jogadas, resultando em $\frac{1}{16}$.

Nas correspondências entre Pascal e Fermat, vários problemas análogos a este foram solucionados. No entanto, poucos meses após essas trocas, Pascal cessou suas atividades matemáticas. Anos mais tarde, Fermat tentou um encontro com Pascal, mas este recusou a proposta (Tabak, 2016).

Pascal, desgastado pela vida parisiense, adotou uma postura ascética, direcionando suas preocupações para questões de moralidade, religião e fé (Hacking, 2001). Ele é frequentemente reconhecido como um dos primeiros grandes filósofos existencialistas cristãos (Hacking, 2001). O argumento atualmente conhecido como a aposta de Pascal é apenas um exemplo dentre seus numerosos e breves pensamentos sobre a condição humana, moralidade, espiritualidade e religião (Hacking, 2001). Uma de suas célebres máximas é: "O coração tem suas razões, que a razão desconhece".

Antes de prosseguirmos, é importante entender que a probabilidade, naquela época, era ainda incipiente e seus métodos de cálculo não abrangiam uma ampla gama de situações em que hoje a probabilidade se faz presente.

A metodologia de cálculo utilizada na época de Pascal e Fermat, atualmente denominada definição clássica de probabilidade, focava-se no conceito de equiprobabilidade de eventos[6]. Esse conceito era considerado primitivo (Gnedenko, 1978), não sendo submetido a uma definição formal. O tema é de fácil compreensão. Suponha que tenhamos um dado (cubo), feito de material perfeitamente homogêneo (e sólido). A ideia é que cada face tem igual chance de aparecer voltada para cima quando o dado é lançado[7]. Em virtude da simetria do cubo, nenhuma das faces teria preferência objetiva sobre as demais. Esse exemplo pode ser relacionado ao conceito de probabilidade física, que será aprofundado posteriormente, juntamente com a probabilidade epistêmica[8]. A partir dessa compreensão da realidade física, podemos dizer que a chance de uma das faces, por exemplo o número 1, aparecer voltada para cima após o lançamento do dado é de um em seis (um sexto). Essas ideias são amplamente discutidas em temas relacionados a jogos de azar, extensamente explorados nos séculos XVII e XVIII.

Note que, conforme colocada a questão, podemos discutir probabilidade mesmo antes do lançamento dos dados. Por essa razão, falamos em probabilidade a priori. Como já afirmado,

[6] Conforme elucidado por Gnedenko (1978, p. 23), a definição clássica de probabilidade restringe-se à noção de probabilidade igual, um princípio fundamental da teoria que não requer uma definição formal. Essa abordagem nos proporciona uma compreensão mais clara da concepção de probabilidade na época, refletindo uma perspectiva menos rigorosa em comparação com os padrões teóricos contemporâneos.
[7] Estamos assumindo que o terreno é completamente plano e que não há possibilidade de a aresta do cubo aparecer voltada para cima. Embora, em razão da simetria do cubo, se bem definido o que significa “face voltada para cima” essas questões não fariam realmente diferença. No entanto, é importante reconhecer o interesse particular dos alunos em explorar essas questões, que, embora não alterem substancialmente o resultado, podem gerar discussões relevantes.
[8] Muitos autores argumentam que a definição de probabilidade clássica não corresponde adequadamente ao conceito de probabilidade física. Um dos fatores que sustenta essa visão é o princípio clássico de atribuir probabilidades iguais aos eventos quando não há informações suficientes para distingui-los. De toda forma, exemplos clássicos como o lançamento de dados, de moedas e a escolha de cartas, entre outros, quando expostos da forma como fizemos no exemplo do cubo, parecem se aproximar do conceito de probabilidade física.

não há consenso sobre o uso dos termos, e, por vezes, seu uso na filosofia difere do uso na matemática. Todavia, parece que o termo a priori é mais adequado para essa situação do que nos casos associados à frequência, que, embora possam ser discutidos antecipadamente, sua compreensão é formada a posteriori, dada sua dependência de testes anteriores. Claro que, ao aprofundarmos a discussão, perceberemos a necessidade de compreensão prévia também para o cálculo nos casos clássicos (sobre a gravidade, homogeneidade etc.), daí a razão pela qual esse conceito não avança muito sem críticas.

Aqui, cabe apenas mais um apontamento. Um conceito relacionado, mas distinto, surgirá ao discutirmos o teorema de Bayes, ponto central deste texto. Nesse contexto, a probabilidade prévia, também conhecida pelo termo *prior* e algumas vezes referida de forma atécnica como *probabilidade a priori* ou *distribuição a priori*, pode ser formada por compreensão física do processo, por frequência (considerada física por muitos) ou pela crença do agente. Isso antecipa o motivo de termos afirmado que, para a linha desenvolvida neste texto, não há coincidência no sentido dos termos *bayesiano* e *epistêmico*.

Voltemos ao cálculo de probabilidade previamente discutido (cubo). Estamos perante o que ficou conhecido como definição clássica de probabilidade. A ideia central dessa definição advém do entendimento de que a probabilidade é igualmente distribuída entre todos os resultados possíveis. Assim, o cálculo da probabilidade clássica se resume à razão entre os casos favoráveis (aqueles em que o evento ocorre) e os casos possíveis (todos os casos, independentemente da ocorrência do evento). Este princípio fundamental foi primeiramente apresentado na obra de Marquês de Laplace (1840)[9].

[9] Embora o *Ensaio Filosófico sobre Probabilidades* de Laplace seja amplamente reconhecido, Gillies (2000) coloca em dúvida sua originalidade, ao mesmo tempo em que admite sua importância fundamental para a difusão das ideias da interpretação clássica da probabilidade entre matemáticos e filósofos do século XIX. Segundo Gillies (2000), a teoria clássica da probabilidade emergiu como um produto do Iluminismo europeu, integrando várias ideias representativas dessa época, em particular, destaca-se a constante referência à mecânica newtoniana e a subsequente crença no determinismo universal. O *Ensaio Filosófico sobre Probabilidades* de Laplace apresenta uma das formulações mais célebres da tese do determinismo universal (Gillies, 2000).

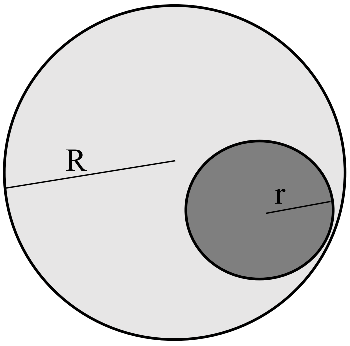
Para discutir os casos favoráveis e possíveis, partimos do pressuposto de que todos os casos têm igual chance de ocorrer, ou seja, são equiprováveis. Essa é a essência da interpretação clássica, defendida por Moivre e Laplace. Versões iniciais dessa interpretação podem ser encontradas nos trabalhos de Pascal, Bernoulli, Huygens e Leibniz, conforme exposto por Hájek (2002).

Outro ponto que merece destaque é a atitude a ser adotada quando não há informações suficientes para determinar a probabilidade dos eventos. Esta questão é de grande importância no contexto do bayesianismo. No momento, é suficiente mencionar que, na probabilidade clássica, era comum estabelecer, na ausência de informações adicionais, que todos os eventos possíveis deveriam ser considerados igualmente prováveis. Esse é o chamado princípio da razão insuficiente ou da indiferença, cuja discussão será retomada ao tratarmos do princípio da máxima entropia, na segunda parte deste texto.

1.2 TEORIA GEOMÉTRICA

No entanto, essa abordagem simplista, que se fundamenta em um conjunto finito de casos favoráveis e possíveis, revelou-se inadequada para a modelagem de uma ampla gama de problemas enfrentados pelos matemáticos. Como observado por Gnedenko (1978, p. 33), nas etapas iniciais do desenvolvimento da teoria da probabilidade, especialmente no que diz respeito à definição clássica, as suas limitações tornavam a aplicação em contextos em que os resultados possíveis eram infinitos particularmente desafiadora, mesmo ao considerar eventos com probabilidade igual. Essa inadequação gerou a necessidade de uma expansão do conceito de probabilidade.

Vejamos um exemplo ilustrativo da primeira extensão do conceito de probabilidade. Considere uma região R em um plano, contendo uma sub-região r. A probabilidade de um ponto escolhido aleatoriamente na região *R* estar localizado na sub-região *r* será analisada para ilustrar a aplicação desta extensão conceitual.



É intuitivo pensar que, em áreas uniformes, a probabilidade de um ponto estar em uma sub-região seria a razão entre a área da sub-região (casos favoráveis) e a área da região total (casos possíveis). Note-se que, neste contexto, não estamos lidando com uma quantidade finita de eventos, diferindo assim do exemplo anterior. Contudo, a noção de probabilidades iguais persiste, pois consideramos que cada ponto do plano tem igual probabilidade.

Diversos problemas interessantes são resolvidos com essa abordagem da probabilidade geométrica. Exemplos clássicos incluem cálculos de probabilidades relacionados a horários, como a probabilidade de pegar um trem em situações em que há múltiplas categorias de trens passando em intervalos regulares, ou problemas de encontros marcados, onde as pessoas esperam apenas por um período limitado.

A teoria da probabilidade geométrica tem sido objeto de diversas críticas, especialmente no que tange à definição da probabilidade de eventos. Gnedenko (1978) ressalta que muitos estudiosos argumentam ser inviável definir a probabilidade de maneira objetiva de forma independente do método de cálculo. Joseph Bertrand, um destacado matemático francês do século XIX, figura entre os críticos mais notórios. Bertrand apresentou, em seus trabalhos, uma série de problemas envolvendo probabilidade geométrica, demonstrando de forma precisa que em muitos desses problemas os resultados obtidos podem variar significativamente a depender do método de solução utilizado (Gnedenko, 1978).

A tentativa de ultrapassar a definição clássica de probabilidade enfrentou um desafio fundamental adicional: o conceito de equiprobabilidade. Mesmo em contextos finitos, ao considerar eventos elementares que se supõe serem simétricos, era imperativo combiná-los segundo regras específicas para constituir os casos favoráveis e os casos possíveis, sempre com o objetivo de preservar a simetria entre eles e, assim, conduzir os cálculos com a premissa de que as probabilidades são iguais (Cramer, 1973). Após dominar os jogos mais simples, a tarefa de lidar com a equiprobabilidade torna-se consideravelmente mais complexa. Cramér (1973) observa que, mesmo indivíduos com bom treinamento matemático frequentemente cometem erros ao tentar resolver problemas desse tipo.

Devido à complexidade inerente à equiprobabilidade, as discussões históricas sobre probabilidade frequentemente exibem uma ampla gama de opiniões sobre a forma correta de determinar os casos possíveis e favoráveis para um determinado problema (Cramér, 1973). Um exemplo dessa controvérsia é observado na análise de uma das questões de Méré, conforme discutido por Cramér (1973). Outro exemplo, mencionado anteriormente, refere-se ao problema da divisão de apostas. Considere a situação em que dois apostadores, A e B, participam de um jogo utilizando uma moeda perfeitamente simétrica. O jogo envolve dois lançamentos da moeda. A vitória de A ocorre se, em pelo menos um dos lançamentos, o resultado for *cara*; caso contrário, a vitória pertence a B. Qual é a probabilidade de A vencer neste cenário?

Remetendo à discussão sobre o *problema da divisão de apostas,* é possível identificar ao menos duas soluções para o jogo descrito. A primeira abordagem envolve a listagem dos possíveis resultados: *cara-cara, cara-coroa, coroa-cara* e *coroa-coroa*. Nesse contexto, fica claro que a probabilidade de A vencer é $\frac{3}{4}$, uma vez que em três dos resultados ele ganha, enquanto B vence apenas em um, *coroa-coroa*. Segundo Cramér (1973), essa foi a solução inicialmente proposta por Fermat.

Entretanto, conforme relatado por Cramér (1973), o matemático contemporâneo Roberval apresentou uma discordância em relação a essa solução. Roberval argumentou que apenas três cenários deveriam ser considerados, pois um resultado inicial favorável (cara) eliminaria a necessidade de um segundo lançamento, já que A teria vencido imediatamente. Assim, os casos relevantes seriam *cara*, *coroa-cara* e *coroa-coroa*. Como A vence nos dois primeiros cenários, a probabilidade de vitória seria $\frac{2}{3}$ e não $\frac{3}{4}$, como discutido anteriormente.

Ainda segundo Cramér (1973), D'Alembert também apresentou objeções semelhantes às regras amplamente aceitas da teoria da probabilidade, propondo que deveríamos superar a noção de simetria (equiprobabilidade) utilizada na solução de Fermat. Essas controvérsias ressaltam a complexidade inerente ao uso do princípio da equiprobabilidade e a inadequação da definição clássica de probabilidade, a qual carece de critérios claros e objetivos para a resolução de problemas que, à primeira vista, parecem simples. É relevante observar que a limitação da definição clássica só foi amplamente reconhecida e examinada com rigor considerável após um longo período (Cramér, 1973).

Como se pode observar, a definição clássica de probabilidade apresenta desafios significativos para os matemáticos. Este ponto é especialmente evidente, conforme salientado por Gnedenko (1978), ao abordar problemas de maior complexidade, particularmente aqueles que emergem nas ciências naturais. Por exemplo, não é possível abordar a probabilidade de um átomo de uma substância radioativa sofrer desintegração dentro de um intervalo de tempo específico utilizando unicamente considerações de simetria que sustentam os princípios de equiprobabilidade (Gnedenko, 1978). Situações análogas ocorrem em diversos problemas práticos, como as probabilidades associadas a eventos de nascimento e morte.

1.3 TEORIA FREQUENCISTA

O desenvolvimento histórico de uma ciência, em geral, não segue um percurso linear e frequentemente integra conceitos oriundos de outros contextos. Como será discutido posteriormente, uma nova e crucial ideia emergiu, desempenhando um papel fundamental na evolução da teoria da probabilidade: a frequência relativa.

Antes de discutirmos a frequência relativa, é útil resumir a evolução histórica até este ponto, o que certamente ajudará na compreensão das discussões subsequentes.

Retomando os fundamentos da teoria clássica, é perceptível que, com o passar do tempo, a terminologia e os métodos analíticos da teoria da probabilidade — inicialmente desenvolvidos com o objetivo de formalizar uma teoria matemática para os jogos de azar — demonstraram uma eficácia notável na resolução de uma ampla gama de problemas de natureza diversificada, que transcendem os limites da definição clássica de probabilidade (Cramér, 1973).

Entre os exemplos elencados por Cramér (1973), sobressaem-se as estatísticas demográficas e a teoria matemática do seguro de vida, duas áreas profundamente interconectadas que experimentaram um significativo avanço ao longo do século XVIII. Cramér (1973) ilustra essas aplicações através da análise da distribuição dos sexos entre recém-nascidos em séries observacionais. Esse fenômeno pode ser relacionado com uma sequência repetida de um jogo de azar, onde cada nascimento resulta em um de dois possíveis desfechos: menino ou menina.

De maneira análoga, ao monitorar um grupo de indivíduos pertencentes a uma faixa etária específica durante um intervalo de tempo determinado, como um ano, a sobrevivência de um indivíduo ao término desse período pode ser interpretada como um evento binário que se assemelha a um jogo de azar, com duas possíveis categorias de desfecho: *vivo* ou *morto* (Cramér, 1973).

É relevante salientar a observação de um fenômeno notável: com a aplicação intensiva desta metodologia estatística, os índices de frequência obtidos a partir dessas observações, tais como a porcentagem de meninos entre os recém-nascidos, demonstram uma tendência de estabilização ao longo do tempo, de maneira análoga aos índices de frequência observados em jogos de azar (Cramér, 1973).

Com base nessa observação, surgiu de maneira natural a associação da noção de probabilidade com essas sequências de eventos, em um paralelo com a probabilidade observada em jogos de azar. Dessa forma, era possível conceber a probabilidade de um indivíduo de uma faixa etária específica vir a falecer dentro de um intervalo de tempo determinado de forma análoga à probabilidade de se obter um número específico em um lançamento de dados (Cramér, 1973).

Parecia, então, lógico aplicar a terminologia e as regras de cálculo da teoria dos jogos de azar a este novo contexto. Dessa aplicação, emergiram novos resultados significativos e aplicáveis, incluindo métodos sofisticados para o cálculo de tabelas de sobrevivência, valores atuariais de anuidades e prêmios de seguros de vida, entre outros (Cramér, 1973).

Durante a expansão do alcance da teoria da probabilidade, não se dedicou atenção adequada às questões previamente identificadas relativas à definição clássica de probabilidade (Cramér, 1973). O cálculo probabilístico continuou a depender da contagem de casos favoráveis em relação aos casos possíveis, exigindo uma simetria que garantisse a equiprobabilidade. Essa abordagem manteve as dificuldades previamente apontadas (Cramér, 1973).

Apesar da ampliação dos domínios de aplicação da teoria das probabilidades, a problemática da interpretação dessa definição persiste, particularmente no que se refere à concepção de casos possíveis e favoráveis fora do contexto dos jogos de azar, como nos exemplos de nascimentos ou mortalidade (Cramér, 1973).

Como Cramér (1973) aponta, o produto desse processo de extensão encontra sua expressão concisa na obra de Laplace, *Théorie Analytique des Probabilités* (1812), um tratado monumental que integra de maneira magistral tanto a teoria matemática dos jogos de azar quanto uma vasta gama de questões científicas e práticas pertinentes à probabilidade.

Entretanto, conforme Cramér (1973), ao examinar as definições fundamentais, Laplace adotou uma abordagem não crítica, sugerindo que ele acreditava que a definição tradicional de probabilidade era universalmente aplicável. Laplace percebia qualquer aplicação como se fosse análoga a um jogo de azar, no qual os resultados possíveis se dividem de maneira inerentemente simétrica. Contudo, ele não forneceu diretrizes para lidar com contextos que transcendessem a esfera dos jogos de azar (Cramér, 1973).

É inegável que a contribuição de Laplace teve um impacto profundo e duradouro no desenvolvimento da teoria das probabilidades. No entanto, um problema substancial surgiu a partir dessa influência: a admiração pelos avanços e pelos resultados práticos significativos obtidos não foi acompanhada por uma devida consideração das fragilidades subjacentes nos fundamentos conceituais (Cramér, 1973). Esse cenário culminou em uma expansão rápida e incessante das aplicações da teoria da probabilidade ao longo do século XIX, enquanto a estrutura teórica matemática subjacente permanecia praticamente estática (Cramér, 1973).

Em virtude dessa estagnação, a estreita relação entre a teoria da probabilidade e a análise matemática geral foi gradualmente se enfraquecendo (Cramér, 1973). A teoria matemática da probabilidade permaneceu praticamente inalterada em um contexto histórico marcado por uma crescente demanda por precisão lógica e rigor matemático, cada vez mais evidente em outros ramos da matemática (Cramér, 1973).

Esse distanciamento teórico da probabilidade em relação ao desenvolvimento rigoroso da matemática não passou despercebido pelos matemáticos contemporâneos, que enfatizavam as limitações inerentes à definição clássica de probabilidade e a necessidade de construir a teoria sobre fundamentos mais sólidos (Cramér, 1973).

Ao longo do tempo, tornou-se evidente a necessidade de uma revisão teórica crítica e robusta dos fundamentos da teoria da probabilidade, visando resolver os problemas identificados e alinhar a teoria com o rigor e a precisão lógica característicos de outras áreas da matemática (Cramér, 1973). O primeiro aspecto a ser submetido a uma análise mais rigorosa foi o conceito de equiprobabilidade, empregado na definição clássica da teoria, que já havia demonstrado ser problemático. O objetivo era integrar um critério que aprimorasse essa definição (Cramér, 1973).

Os estudiosos concordavam amplamente que muitas das aplicações práticas significativas da teoria da probabilidade não se ajustavam ao conceito de casos igualmente possíveis, conforme estipulado pela definição clássica (Cramér, 1973). Essa limitação torna-se particularmente evidente em exemplos aparentemente simples, como a probabilidade de uma pessoa de uma idade específica falecer em um determinado ano (Cramér, 1973). Neste contexto, torna-se desafiador identificar situações em que os casos possam ser considerados igualmente prováveis (Cramér, 1973).

Em resposta a esses argumentos, diversos autores buscaram um novo critério para substituir a definição clássica de probabilidade. Propuseram uma nova abordagem baseada na estabilidade das frequências relativas, considerada uma substituição mais natural (Cramér, 1973). Esses autores sugeriram que o quociente de frequência de um evento específico em uma série de observações fosse tomado como a probabilidade do evento (Cramér, 1973). Defenderam que essa definição alternativa apresentava vantagens substanciais em relação à definição clássica, convertendo as regras tradicionais em meras ferramentas práticas para a realização de cálculos probabilísticos (Cramér, 1973).

Para aprofundar a compreensão dessa transição, é necessário revisitar o contexto histórico que a precede. A concepção de utilizar o quociente de frequência de um evento em uma sequência de observações como uma estimativa probabilística emergiu a partir de extensas séries de observações empíricas sobre a ocorrência de eventos. Conforme antecipado, verificou-se que, para diversos fenômenos, a distribuição das ocorrências e não ocorrências de um evento tende a seguir uma lei estatisticamente estável, quando se realizam um número substancial de experimentos sob condições controladas e uniformes (Gnedenko, 1978).

Para ilustrar, considere-se μ como o número de ocorrências de um evento A após n experimentos independentes. Observa-se que, para um n suficientemente grande, a razão $\frac{\mu}{n}$ converge para um valor quase constante na maioria das séries, com a frequência de grandes desvios diminuindo à medida que o número de experimentos aumenta (Gnedenko, 1978).

A estabilidade observada nos fenômenos demográficos foi inicialmente identificada em contextos de natalidade. De acordo com Gnedenko (1978), desde épocas remotas, havia um reconhecimento de que a proporção de nascimentos de meninos em relação ao total de nascimentos permanecia relativamente constante ao longo dos anos. Registros censitários realizados na China, datados de mais de dois mil anos antes de Cristo, já indicavam essa proporção como sendo de aproximadamente $\frac{1}{2}$ (Gnedenko, 1978).

Durante os séculos XVII e XVIII, emergiram diversos estudos pioneiros sobre estatísticas populacionais. Além da observação da estabilidade na razão de nascimentos de meninos em comparação com meninas, foram identificadas outras leis estatísticas consistentes. Entre essas leis, destacam-se a porcentagem de mortes em idades específicas dentro de grupos populacionais determinados (como categorias econômicas), bem como a distribuição de características físicas, como altura, largura do tórax e comprimento do passo, entre outros parâmetros (Gnedenko, 1978).

Essa verificação da estabilidade da frequência relativa após um grande número de eventos aleatórios sugere a existência de leis subjacentes que regulam a manifestação dos fenômenos, independentemente da influência do experimentador, como evidenciado pela constância observada na frequência relativa (Gnedenko, 1978).

Adicionalmente, a constatação de que a frequência relativa converge para o valor da probabilidade clássica quando se trata de um evento para o qual a definição clássica é aplicável, desde que o número de experimentos seja suficientemente elevado, sugere que, de forma mais geral, existe uma constante em torno da qual a frequência relativa tende a oscilar (Gnedenko, 1978). Considerando que tal constante representa uma característica numérica objetiva do fenômeno em questão, é razoável denominá-la como a probabilidade do evento aleatório em análise (Gnedenko, 1978).

Assim, consideramos que um evento A possui uma probabilidade quando se observam as seguintes condições, conforme estabelecido por Gnedenko (1978): é possível conduzir um número ilimitado de experimentos mutuamente independentes sob condições idênticas, onde o evento A pode ou não ocorrer em cada teste. Em um número suficientemente grande de experimentos, a frequência relativa com que o evento *A* ocorre em quase todas as séries dentro de um extenso conjunto de experimentos se desvia apenas ligeiramente de uma constante específica, que geralmente é desconhecida.

A probabilidade de um evento aleatório, quando definida sob essa perspectiva, é conhecida como probabilidade estatística (Gnedenko, 1978). Segundo Gnedenko (1978, p.42), as frequências relativas exibem as seguintes propriedades:

- A frequência relativa de um evento certo é 1;
- A frequência relativa de um evento impossível é zero;
- Se o evento aleatório *C* é a soma de um número finito de eventos mutuamente exclusivos $A_1, A_{2,\ldots}A_n$, então sua frequência relativa é igual à soma das frequências relativas de cada um dos eventos $A_1, A_{2,\ldots}, A_n$.

Essas características das frequências relativas conduzem, de maneira natural, às seguintes propriedades da definição de probabilidade estatística, conforme abordado por Gnedenko (1978, p. 43):

- A probabilidade do evento certo é 1;
- A probabilidade do evento impossível é zero;
- Se um evento aleatório *C* é a soma de um número finito de eventos mutuamente exclusivos $A_1, A_{2,\ldots}, A_n$, cada um com uma probabilidade associada, então a probabilidade de C ocorrer é igual à soma das probabilidades dos eventos A_i, ou seja, $P(C) = P(A_1) + \mathrm{P}(A_2) + \cdots + P(A_n)$.

A definição estabelecida parece oferecer vantagens consideráveis em virtude de sua simplicidade e da facilidade para o desenvolvimento de suas propriedades. Adicionalmente, preserva um caráter objetivo e mantém uma certa independência em relação ao observador (Gnedenko, 1978). Entretanto, essa definição não está isenta de desafios. Em especial, sua natureza mais descritiva do que formalmente matemática pode limitar sua aplicabilidade. Ademais, ela não esclarece adequadamente as condições e os fenômenos específicos nos quais a frequência relativa pode ser considerada estável (Gnedenko, 1978).

A teoria da frequência da probabilidade surgiu inicialmente no século XIX, sendo impulsionada pela escola de Cambridge, representada por Robert Leslie Ellis e John Venn, e pode ser interpretada como uma resposta empirista britânica ao racionalismo continental[10] defendido por Laplace e seus seguidores[11] (Gillies, 2000). John Venn foi o primeiro a formular claramente essa teoria em 1866, e, posteriormente, em 1928, Richard Martin Edler von Mises, matemático e filósofo austríaco que mais tarde se tornou professor em Harvard, desenvolveu-a de maneira sistemática (Hacking, 2001).

Neste estágio, é relevante analisar a interpretação de probabilidade proposta por Richard von Mises, amplamente adotada nas ciências naturais. Von Mises, um dos principais teóricos da probabilidade frequencista, exerceu uma influência considerável sobre a compreensão e aplicação da probabilidade tanto na ciência quanto na filosofia. Sua abordagem, centrada na frequência de eventos em longas séries de experimentos repetidos, é conhecida como interpretação frequencista da probabilidade.

A interpretação frequentista da probabilidade, conforme formulada por von Mises, está alicerçada em dois princípios fundamentais: a noção de uma série de ensaios repetidos e o

10 É crucial enfatizar a divergência existente nesse ponto. De acordo com Verburgt (2014), não se constituiu uma escola britânica coesa no âmbito do frequencismo durante o século XIX, e os trabalhos de Ellis e Venn não podem ser considerados como uma reação empirista britânica contra a teoria tradicional da probabilidade.

11 Durante o apogeu do empirismo, em grande parte promovido pelo Círculo de Viena, a teoria da probabilidade ganhou notável relevância (Gillies, 2000). Inicialmente, o epicentro dessa corrente de pensamento estava na Europa continental; contudo, após a dispersão dos membros do Círculo de Viena, a teoria se difundiu entre os países de língua inglesa (Gillies, 2000). Nesse período, dois intelectuais intimamente ligados ao Círculo de Viena, Hans Reichenbach e Richard von Mises, realizaram desenvolvimentos significativos na teoria da probabilidade (Gillies, 2000).

conceito de um limite para a frequência relativa (Gnedenko, 1978). Von Mises definiu a probabilidade de um evento como a frequência relativa desse evento em uma **sequência infinita** de experimentos realizados sob condições idênticas (Gnedenko, 1978).

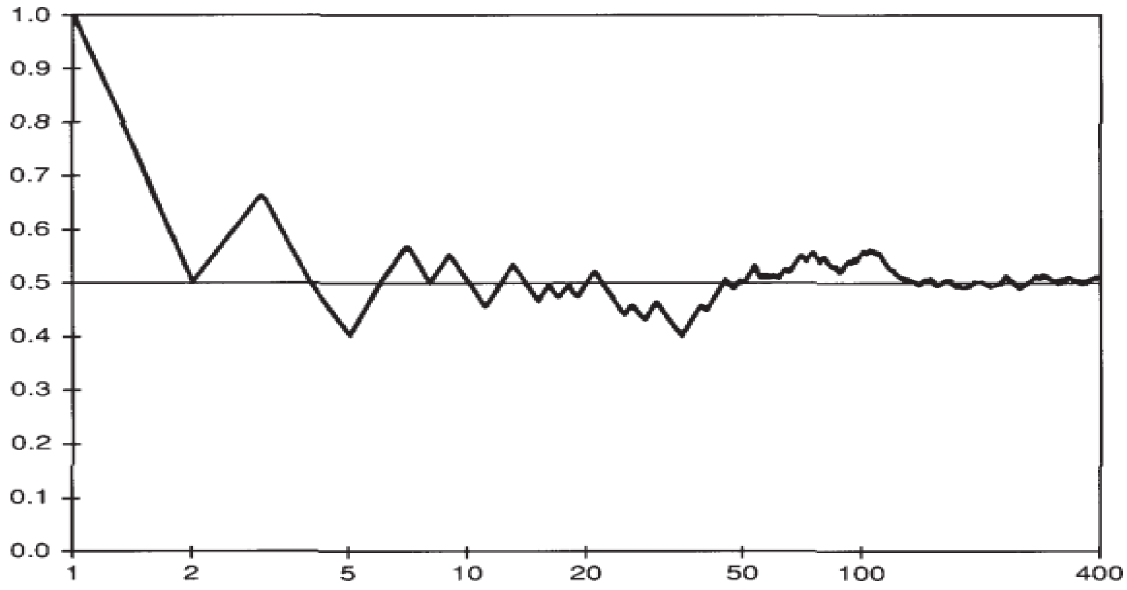
Considerando o exemplo clássico do lançamento de uma moeda justa, a probabilidade frequencista de obter *cara* seria definida como o limite da frequência relativa de caras em uma sequência infinita de lançamentos da moeda. Matematicamente, podemos expressar isso da seguinte forma:

$$P(cara) = \lim_{n\to\infty}\left(\frac{\mathrm{n(cara)}}{n}\right),$$ [12]

onde n é o número total de lançamentos e n(cara) o número de vezes que a moeda caiu com o lado *cara* voltado para cima.

Por exemplo, se lançarmos a moeda 100 vezes e obtivermos 55 caras, a frequência relativa seria 0,55. À medida que continuarmos os lançamentos, aumentando o número de ensaios, a frequência relativa tende a se aproximar de 0,5, assumindo que a moeda é justa e que os lançamentos são independentes. A seguir, um gráfico extraído de Gillies (2000, p. 93) ilustra esse exemplo.

[12] $\lim_{x\to\infty} f(x) = L$ significa que para cada número real $\varepsilon > 0$, dado arbitrariamente, podemos encontrar $\delta > 0$ de modo que se tenha $|f(x) - L| < \varepsilon$ (Lima, 1995), em outras palavras, é sempre possível aproximar a função do valor L, bastando que se escolha um x suficientemente grande. Assim, o que a fórmula atribuída a von Mises nos diz é que para n suficientemente grande, podemos nos aproximar do resultado P(cara) tanto quanto quisermos.



Some empirical evidence for the law of stability of statistical frequencies. The frequency ratio of heads in a sequence of tosses of a coin (logarithmic scale for the abscissa).

Von Mises adota a definição mencionada anteriormente como sua concepção de probabilidade, argumentando que qualquer definição *a priori* seria inadequada. Ele defende que uma definição empírica é crucial para satisfazer as exigências das ciências, como a matemática e a filosofia (Gnedenko, 1978).

A abordagem de von Mises enfrenta diversas críticas. Uma das principais é a impraticabilidade das sequências infinitas e a dificuldade em tratar eventos únicos, questões que serão exploradas com maior detalhe na segunda parte deste texto. Ademais, existe a crítica concernente à dificuldade de assegurar que o limite de frequências convergirá para um valor constante, uniforme em todas as sequências e repetições do experimento (Ross, 2010). John Maynard Keynes[13], defensor da teoria lógica da probabilidade, rejeitava a abordagem frequencista, considerando a ideia de frequência a longo prazo como desprovida de sentido

---

[13] Ao ingressar no King's College, Cambridge, como estudante de graduação em 1903, John Maynard Keynes foi introduzido à sociedade secreta conhecida como os Apóstolos (Gillies, 2000). Naquele período, a sociedade contava com membros de destacada relevância, como Bertrand Russell e G.E. Moore, cuja influência exerceu um papel determinante na formação do pensamento de Keynes (Gillies, 2000).

(Hacking, 2001). Para Keynes, o conceito de longo prazo era meramente metafórico e desconectado da realidade concreta (Hacking, 2001).

Adicionalmente, Hacking (2001) argumenta que a ideia de frequência limite representa uma forma extrema da probabilidade frequentista, idealizando uma série de resultados de testes reais em um contexto aleatório. Esta abordagem privilegia o observável, em detrimento das causas ou estruturas subjacentes (Hacking, 2001). Frequências relativas estáveis emergem apenas se uma configuração tiver certas propriedades físicas ou geométricas subjacentes, sendo possível existir essa estrutura mesmo sem ensaios reais (Hacking, 2001).

A partir dessa perspectiva, surge uma nova ideia que enfatiza a inclinação, disposição ou propensão da configuração aleatória, o que é particularmente adequado ao considerarmos processos estocásticos naturais, como o decaimento radioativo (Hacking, 2001). Esta abordagem inovadora da probabilidade como propensão foi desenvolvida pelo filósofo austríaco Karl Popper[14], renomado professor da London School of Economics e um dos filósofos mais influentes do século XX. Este tema será retomado na segunda parte deste texto.

Finalmente, é crucial enfatizar que a probabilidade frequencista representa o alicerce da estatística clássica ou frequencista. Em termos simplificados, esta abordagem fundamenta-se na frequência com que eventos são observados para inferir sobre características de uma população. O pressuposto central é que as probabilidades dos eventos podem ser determinadas pela proporção de ocorrências observadas em um grande número de experimentos repetidos sob condições idênticas, como discutido previamente. A partir dessa premissa, são desenvolvidos modelos que representam situações factuais de interesse, possibilitando uma gama de inferências variadas.

[14] Embora Karl Popper seja amplamente reconhecido como o precursor da interpretação da propensão, é importante destacar que a concepção dessa ideia já estava presente nos escritos de Charles Sanders Peirce, conforme indicado por Hájek (2002).

As distribuições de probabilidade desempenham um papel central na estatística frequencista ao descrever como os valores de uma variável aleatória são distribuídos. Exemplos comuns incluem a distribuição normal, a distribuição binomial e a distribuição de Poisson. Cada uma dessas distribuições é caracterizada por parâmetros específicos que determinam sua forma e propriedades (Morettin & Bussab, 2010). Segundo Morettin e Bussab (2010), a ideia fundamental é que, com suposições apropriadas, podemos criar um modelo teórico que reproduza de maneira razoável a distribuição de frequências observadas, mesmo sem observar diretamente o fenômeno aleatório de interesse.

Um modelo estatístico configura uma estrutura matemática que delineia o comportamento de uma variável ou de um conjunto de variáveis. As distribuições de probabilidade vinculadas a esses modelos formam a base da inferência estatística, permitindo a quantificação de probabilidades, a execução de testes de hipóteses e a formulação de previsões. Dentre essas distribuições, a normal é particularmente proeminente na estatística frequencista, sendo frequentemente adotada como alicerce para uma variedade de modelos estatísticos.

Apesar da ampla utilização da estatística clássica, ela não escapa de críticas, um tema a ser explorado na segunda parte deste texto. Antes de abordarmos essa discussão, introduziremos a teoria axiomática, crucial para a formalização rigorosa e o substancial desenvolvimento da teoria moderna da probabilidade.

1.4 TEORIA AXIOMÁTICA

Desde o meio do século XVII, houve uma série de tentativas para definir a probabilidade com base em frequências relativas, como abordado anteriormente, e para fundamentar a teoria a partir dessa definição (Cramér, 1973). No decorrer do século XX, essas investigações começaram a ser progressivamente moldadas pela tendência à axiomatização, um traço distintivo da matemática moderna (Cramér, 1973). Nesse contexto, a probabilidade de um evento foi reconfigurada como uma magnitude numérica atribuída ao evento, a qual deve obedecer a propriedades básicas estabelecidas por axiomas—ou seja, proposições fundamentais aceitas sem necessidade de demonstração (Cramér, 1973). Assim, a definição de probabilidade passou a ser interpretada como uma quantidade numérica que deve satisfazer determinados axiomas (Cramér, 1973).

Com os recentes avanços, foi reestabelecida a íntima conexão previamente existente entre a teoria da probabilidade e a matemática moderna (Cramér, 1973). Atualmente, a teoria da probabilidade é reconhecida como um subcampo da matemática pura, imersa em um diálogo contínuo e intercâmbio de conceitos com outras áreas matemáticas, enquanto seu escopo de aplicação continua a se expandir (Cramér, 1973). Conforme elucidado por Courant e Robbins (1996), o sistema axiomático é fundamentado na necessidade de evitar uma regressão infinita na prova de teoremas. Em um sistema dedutivo, a demonstração de um teorema envolve a sua derivação lógica a partir de proposições previamente aceitas. Para evitar uma cadeia infinita de provas dessas proposições, utilizam-se axiomas, que são afirmações aceitas como verdadeiras sem a necessidade de prova adicional. A partir desses axiomas, todos os teoremas são deduzidos por meio de raciocínios lógicos rigorosos (Courant & Robbins, 1996).

Quando os fatos dentro de um domínio científico são organizados de modo que todos derivem de um conjunto específico de proposições fundamentais, diz-se que o campo possui uma estrutura axiomática (Courant & Robbins, 1996). Embora a seleção dos axiomas possa ser arbitrária, o método axiomático demonstra uma eficácia notável quando os postulados são simples e limitados em número (Courant & Robbins, 1996). No contexto da teoria da probabilidade, a escolha dos axiomas foi meticulosamente elaborada para abarcar e refletir integralmente o desenvolvimento prévio da teoria.

Embora, no contexto dos sistemas formais contemporâneos, os axiomas sejam definidos como um conjunto de sentenças dentro da linguagem do sistema, sem a necessidade de que sejam verdades evidentes ou derivadas de fatos previamente estabelecidos (Mortari, 2016), a situação muda quando consideramos o desenvolvimento de uma ciência matemática. Nesse contexto, a formulação de axiomas que fundamentam uma teoria abrangente não representa um estágio inicial, mas sim o resultado de um extenso processo de acumulação de dados e análise lógica para identificar os princípios fundamentais subjacentes (Gnedenko, 1978). Esse padrão foi implementado na elaboração dos axiomas da geometria elementar, e a teoria da probabilidade trilhou um percurso análogo, embora sua construção axiomática tenha surgido de maneira significativamente mais tardia (Gnedenko, 1978).

Embora Andrei Nikolaevich Kolmogorov seja amplamente reconhecido, Serge N. Bernstein foi um precursor crucial no estabelecimento de uma teoria axiomática da probabilidade (Gnedenko, 1978). Sua abordagem, que enfatizava a comparação qualitativa entre eventos aleatórios, forneceu a base para formulações mais sistemáticas e reconhecidas, como a de Kolmogorov (Gnedenko, 1978). Kolmogorov (1956) fundamentou sua teoria em estruturas formais mais consolidadas, incluindo a teoria da probabilidade, a teoria dos conjuntos, a teoria da medida e a teoria das funções de uma variável real (Gnedenko, 1978).

A formulação de Kolmogorov para a teoria das probabilidades é, de fato, um marco fundamental, mas a discussão sobre suas axiomatizações e interpretações é rica e complexa. Hájek (2002) destaca a ausência de consenso na escolha da melhor axiomatização, sugerindo que a rigidez em seguir os axiomas de Kolmogorov pode limitar a aceitação de outras abordagens igualmente válidas. Ele argumenta que diferentes contextos e aplicações podem demandar diferentes estruturas axiomatizadoras, e rejeitar teorias que não se alinham estritamente com Kolmogorov pode ser uma limitação desnecessária. A visão de Wesley Salmon, que favorece uma axiomatização alternativa, enfatiza a importância de considerar múltiplas perspectivas na fundamentação da teoria da probabilidade. Essa pluralidade é fundamental para enriquecer o debate filosófico e matemático sobre o assunto, permitindo que novas ideias e abordagens possam emergir e ser exploradas.

Embora se considerem as observações de Hájek (2002) sobre a possibilidade de outras axiomatizações e sua assertiva de que seria lamentável desconsiderar uma teoria apenas por não aderir rigidamente aos axiomas de Kolmogorov, a discussão se torna ainda mais complexa diante da violação efetiva desses axiomas. Essa questão está intrinsecamente ligada ao probabilismo, que propõe a utilização da probabilidade como um método para abordar e justificar o conhecimento e a ação em situações de incerteza (Neiva, 2022). Este tópico será tangenciado na segunda parte deste texto.

Superada a discussão anterior, é fundamental ressaltar que a abordagem de Kolmogorov, marcada por sua simplicidade e rigor, consolidou-se como uma referência significativa na literatura filosófica e matemática, razão pela qual foi a metodologia escolhida para este estudo.

Tabak (2016) ressalta a importância de Andrei Nikolaevich Kolmogorov como um matemático de relevância ímpar, cujas contribuições foram essenciais para a edificação da teoria moderna da probabilidade. Nascido em Tambov, Rússia, em 25 de abril de 1903, Kolmogorov faleceu em Moscou em 20 de outubro de 1987. Como um professor prolifico, ele deixou um legado de mais de trezentas publicações científicas e orientou a formação de mais de sessenta doutorandos. Sua vasta contribuição científica abrangeu diversas áreas do saber, incluindo topologia, teoria da medida, sistemas dinâmicos e estatística (Alencar, 2014). Embora sua atuação tenha sido multifacetada, Kolmogorov é amplamente reconhecido por seu desenvolvimento da teoria axiomática da probabilidade e da teoria da complexidade (Alencar, 2014). Em sua obra seminal "Foundations of the Theory of Probability" (1956), Kolmogorov formulou os axiomas fundamentais da probabilidade, os quais serão objeto de análise na sequência.

A abordagem axiomática da probabilidade não apenas engloba as definições clássica e frequencista como casos específicos, mas também as supera, incorporando um rigor metodológico em sua formulação (Gnedenko, 1978). Essa metodologia possibilitou a construção de uma estrutura lógica coerente para a teoria da probabilidade contemporânea, atendendo às exigências da ciência moderna e alinhando-se com os desenvolvimentos matemáticos da época (Gnedenko, 1978).

A teoria moderna da probabilidade representa um domínio sofisticado da matemática que se dedica ao estudo da incerteza e dos fenômenos aleatórios. Sua fundamentação lógica está intrinsecamente vinculada à lógica clássica, porém é expandida para abarcar a natureza probabilística dos eventos. Essa expansão é realizada de maneira rigorosa e formal, através de conceitos, axiomas e teoremas que sustentam a estrutura teórica da probabilidade. O presente texto se propõe a examinar alguns dos fundamentos teóricos que são cruciais para a compreensão da estruturação e evolução deste campo filosófico.

Antes de tratarmos do axiomas, importante destacar o que vem a ser um espaço de probabilidade, que é uma estrutura matemática formada por um triplo $(\Omega, \mathscr{F}, P)$, onde:

Espaço Amostral (Ω): Conjunto que contém todos os resultados possíveis de um experimento aleatório.

σ-Álgebra ($\mathscr{F}$): Um conjunto de eventos (subconjuntos de Ω) que satisfaz certas propriedades (fechado sob complementos e uniões contáveis).

Medida de Probabilidade (P): Trata-se de função que associa a cada evento em $\mathscr{F}$ um número real do intervalo [0,1] satisfazendo os axiomas de Kolmogorov destacados a seguir.

A teoria da probabilidade, tal como formulada por Andrei Kolmogorov em 1933, repousa sobre três axiomas fundamentais que estabelecem um espaço de probabilidade:

1. **Axioma da Não-Negatividade:**

$P(A) \geq 0, \quad \forall A \in F$

2. **Axioma da Normalização:**

$P(\Omega)=1$

3. **Axioma da Aditividade (ou Aditividade Contável):**

$P\left(\bigcup_{\boldsymbol{i=1}}^{\infty} A_{\boldsymbol{i}}\right) = \sum_{\boldsymbol{i}=1}^{\infty} P(\boldsymbol{A}_i)$**, se** $\boldsymbol{A_i} \cap \boldsymbol{A_j}$**=0 para i≠j**

Onde Ω é o espaço amostral, $\mathscr{F}$ é uma σ-álgebra de subconjuntos de Ω, e P é a medida de probabilidade.

No âmbito do tratamento axiomático proposto por Kolmogorov, o conceito de evento aleatório é reconfigurado, deixando de ser uma noção primitiva, como foi anteriormente discutido, para se tornar um conceito derivado de fundamentos mais essenciais, conforme se espera em uma teoria axiomática (Gnedenko, 1978).

Conforme delineado por Gnedenko (1978), a formulação dos axiomas pode gerar a exigência de uma hipótese suplementar para o avanço da teoria, a qual é referida como o axioma estendido da adição. Este axioma assume uma relevância fundamental, pois, na teoria da probabilidade, frequentemente nos deparamos com eventos que se podem desagregar em uma infinidade de subeventos, conforme já foi explorado na discussão anterior (Gnedenko, 1978).

No contexto específico deste texto, a hipótese adicional proposta é desnecessária, uma vez que, para simplificação, iniciamos a análise a partir do axioma estendido da adição. É importante ressaltar que adotamos como axioma a contagem infinita de eventos disjuntos. Em diversas obras, como a de Kolmogorov (1956), parte-se de uniões finitas de eventos disjuntos, o que torna necessária a inclusão de uma hipótese adicional para que se possa abordar adequadamente os casos infinitos.

Para complementar o argumento, é pertinente destacar que Gnedenko (1978) demonstra a equivalência[15] entre o axioma estendido[16] da adição e o axioma da continuidade. Essa equivalência revela-se de grande importância, especialmente considerando a necessidade de operar em um ambiente contínuo. Sob a perspectiva da teoria dos conjuntos, nossa definição axiomática de probabilidade consiste na introdução de uma medida não negativa, totalmente aditiva e normalizada no conjunto Ω, a qual é definida sobre todos os elementos do conjunto $\mathscr{F}$ (Gnedenko, 1978).

Conforme debatido, ao estabelecer o conjunto de probabilidade, devemos considerar não apenas o conjunto inicial dos eventos elementares, denotado por Ω, mas também o conjunto de eventos aleatórios $\mathscr{F}$ e a função de probabilidade P que é definida sobre este último (Gnedenko, 1978). A tripla {Ω, $\mathscr{F}$, P} é designada como espaço de probabilidade[17], desempenhando um papel fundamental na formalização rigorosa da teoria da probabilidade.

---

[15] Essa equivalência permite flexibilidade no desenvolvimento teórico, pois ambos os axiomas podem ser usados de forma intercambiável para deduzir propriedades importantes da probabilidade.

[16] O axioma da adição estendido é um componente essencial na teoria da probabilidade, especialmente no contexto da medida de probabilidade em espaços infinitos. Trata-se de um desdobramento natural do axioma da adição que possui fundamental importância, uma vez que muitos problemas práticos envolvem considerações acerca de um número infinito de eventos possíveis.

[17] Como complemento, é importante destacar a diferença entre o espaço de probabilidade discreta e o espaço de probabilidade contínua (Kolmogorov, 1956) (Cramer, 1963). Para um espaço de **probabilidade discreta**, onde Ω é finito ou contável, a σ-álgebra $\mathscr{F}$ é frequentemente o conjunto de todas as partes de Ω, e a medida de probabilidade P é definida diretamente sobre os pontos de Ω. Nesse contexto, a aditividade completa é uma extensão natural da aditividade finita e seu manejo não exige muita complexidade, tendo em vista que $\sum_{i=1}^{\infty} P(A_i)$ é uma soma de uma série contável de números não negativos. A continuidade é normalmente garantida pela natureza discreta dos eventos, onde sequências decrescentes de eventos tendem a se tornar vazias ou a convergir para um evento específico de probabilidade zero.

Em um espaço de **probabilidade contínua**, onde Ω é um conjunto não contável, como um intervalo da reta real, a σ-álgebra $\mathscr{F}$ é tipicamente a σ-álgebra de Borel, e a medida de probabilidade P é frequentemente definida via uma função densidade de probabilidade $f(x)$ para variáveis aleatórias contínuas. Nesse contexto, a aditividade completa garante que a integral da densidade de probabilidade sobre a união de conjuntos

Neste estágio, resta evidente que existem condições fundamentais que devem ser satisfeitas para a determinação da probabilidade de um evento. Quando não são impostas restrições adicionais, além das condições estabelecidas acima, dizemos que a probabilidade em questão é incondicional (Gnedenko, 1978).

Entretanto, há contextos em que o cálculo da probabilidade requer a consideração de condições adicionais, como a ocorrência prévia de um outro evento, cuja probabilidade é maior que zero. Nesses casos, essa probabilidade é classificada como probabilidade condicional, sendo comumente denotada pelo símbolo P(A/B). Conforme observado por Gnedenko (1978), é pertinente afirmar que as probabilidades incondicionais também possuem um caráter condicional. Isso se deve ao fato de que, conforme discutido anteriormente, a teoria da probabilidade admite, desde sua fundamentação, a presença de um conjunto específico de condições que orienta o cálculo probabilístico.

O conceito de probabilidade condicional reveste-se de grande relevância para o presente debate. A sua compreensão é fundamental, uma vez que a formulação e as análises subsequentes relacionadas ao teorema de Bayes, e, por extensão, ao bayesianismo, são intrinsicamente vinculadas a uma interpretação precisa da probabilidade condicional.

O objetivo a seguir é avançar na compreensão do teorema de Bayes, com fundamento nos conceitos que exploramos até o presente momento. Na abordagem clássica da probabilidade, a formulação geral para o cálculo da probabilidade condicional se apresenta de forma simples, fundamentada nos chamados casos favoráveis e casos possíveis, conforme já destacado. A construção desenvolvida no contexto axiomático apresenta uma analogia

disjuntos é a soma das integrais individuais $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \int_{U_{i=1}^{\infty}} f(x)\,dx = \Sigma_{i=1}^{\infty} \int_{A_i} f(x)\,dx$. A continuidade é vital para assegurar que a medida de probabilidade de uma sequência de eventos decrescentes que convergem para um conjunto vazio também converge para zero $\lim_{i\to\infty} P_{(A_i)} = \lim_{i\to\infty} \int_{A_i} f(x)\,dx = 0$

significativa, uma vez que esta abordagem integra teorias preexistentes. A dificuldade maior fica por conta do caso contínuo, para o qual já indicamos o caminho a ser seguido em linhas anteriores. Assim, tendo em vista a já discutida proposta deste texto, vamos avançar nesse ponto a partir da noção clássica de casos favoráveis e casos possíveis, de forma a não tornar demais densa essa parte do texto. Em função da clareza e do usual rigor, adotaremos a exposição de Gnedenko (1978), conforme apresentado nas páginas 53 e seguintes de sua obra, complementando-a com notas explicativas sempre que necessário. Importa ressaltar que diversas obras apresentam essa discussão inclusive mais didática, ao apresentar diagramas que facilitam a visualização. Não seguiremos essa linha, unicamente para evitar o excesso de figuras nessa parte do texto.

Para as discussões acerca da probabilidade condicional e, posteriormente, do teorema de Bayes, Gnedenko (1978) parte de um conjunto de ocorrências $E_1, E_{2,\dots} E_n$ que sejam exaustivas[18], mutualmente exclusivas[19] e igualmente prováveis[20] das quais *m* são favoráveis ao evento A, k favoráveis ao evento B e r favoráveis ao evento $A\ e\ B$ $(r \leq k, r \leq m)$[21].

Em seguida, supõe-se o interesse de calcular a probabilidade de A ocorrer, dada a ocorrência de B. Se o evento B ocorreu, isso implica que ao menos um dos eventos E_j favorável a B ocorreu. Sob essa condição, r, e somente r[22], dos eventos E_j são favoráveis à ocorrência de A. Então:

(I) $$P(A/B) = \frac{r}{k} = \frac{\frac{\mathrm{r}}{\mathrm{n}}}{\frac{\mathrm{k}}{\mathrm{n}}} = \frac{P(A \cap B)}{P(B)}$$

[18] O termo *exaustivo* significa que todos os casos possíveis são considerados, não deixando nenhuma ocorrência de fora. De forma mais técnica, a união de todos os eventos cobre todo o espaço amostral ($E_1 U E_2 \dots U E_n = \Omega$). Note que isso significa que podemos construir qualquer conjunto do nosso universo a partir desses Es, como seria o caso do conjunto A e do conjunto B.
[19] Eventos são mutualmente exclusivos (ou disjuntos) se não podem ocorrem simultaneamente, em outras palavras, não há intersecção entre eles ($E_1 \cap E_2 \dots \cap E_{\mathrm{n}} = \emptyset$). Disso resulta que não contamos duas vezes quanto utilizamos uniões desses conjuntos para formar um conjunto maior.
[20] Estamos tratando de probabilidade clássica, o que explica a equiprobabilidade, conforme já discutido.
[21] Note que se *m* está no conjunto A, *k* no conjunto B e *r* no conjunto A e no conjunto B ao mesmo tempo, fica claro que $r \leq k, r \leq m$.
[22] Note que somente em r é favorável tanto a A, quanto a B.

Da mesma forma se obtém $P(B/A)$, onde:

(II) $P(B/A) = \frac{P(A \cap B)}{P(A)}$

Evidentemente, caso B (ou A) seja um evento impossível, a equação I (ou II) perde seu significado.

Assim, imposta a condição de que os eventos acima não sejam impossíveis, cada uma das fórmulas (I) e (II) passa a equivaler ao chamado teorema da multiplicação, segundo o qual:

$$P(A \cap B) = P(A) \cdot P(B/A) = P(B) \cdot P(A/B)$$

Em outras palavras, a probabilidade da interseção de A e B é igual ao produto da probabilidade de A (ou B) pelo produto da probabilidade de B (ou A) dado que A (ou B) ocorreu.

Note que o teorema da multiplicação continua a fazer sentido se um dos eventos é impossível[23], uma vez que neste caso as equações ficam $P(A \cap B) = 0$ e $P(A\ /\ B) = 0$, quando $P(A) = 0$.

Dizemos que um evento A é independente de um evento B quando $P(A\ /\ B) = P(A)$. Em outras palavras, a ocorrência do evento B não afeta a probabilidade do evento A. Trata-se de tema de grande importância para o estudo de probabilidade, uma vez que os problemas são simplificados quando a independência se evidencia.

Com efeito, se o evento A é independente do evento B, o teorema da multiplicação $P(A)P(B/A) = P(B)P(A/B)$ pode ser simplificado para $P(B/A) = P(B)$ facilitando sobremaneira os cálculos.

[23] A restrição deve ser imposta para que se fale em equivalência com I e II.

O conceito de independência de eventos desempenha um papel crucial na teoria da probabilidade. A maioria dos resultados apresentados nos textos sobre o tema é derivado da hipótese de que os eventos considerados são independentes (Gnedenko, 1978).

Conforme argumenta Gnedenko (1978), usualmente não avaliamos se tais relações são atendidas para estabelecer a independência entre eventos. Na prática, partimos de noções intuitivas para essas considerações. Com efeito, quando estamos analisando lançamentos de moedas, partimos da noção intuitiva de que o fato de o lançamento de uma moeda resultar em *cara* não tem qualquer efeito sobre o lançamento de outra moeda (Gnedenko, 1978). O mesmo vale para um enorme número de outros problemas probabilísticos, como o nascimento de uma menina, que não tem qualquer impacto no sexo do filho nascido em outra família (Gnedenko, 1978).

Partindo-se do teorema da multiplicação, é possível notar que a fórmula relacionada à probabilidade de A e B também é especialmente simplificada no contexto de eventos independentes, uma vez que a ocorrência de A não exerce influência sobre B, e vice-versa (Gnedenko, 1978). A expressão matemática correspondente para os eventos A e B, quando considerados independentes, é apresentada a seguir:

$$P(A \cap B) = P(A) \cdot P(B)$$

A apresentação das expressões acima apenas com A e B tem uma finalidade exclusivamente didática, não havendo maiores dificuldade em ampliar o conceito para uma coleção de eventos. Embora não nos detenhamos neste aspecto, é fundamental ressaltar, como apontado por Gnedenko (1978), que, para que um conjunto de eventos seja considerado mutuamente independente, não é suficiente que esses eventos sejam independentes em pares[24].

A fórmula $P(A\ /\ B)=\frac{P(A\cap B)}{P(A)}$, que no caso da definição clássica é decorrência direta da definição dada à probabilidade condicional, seguirá a mesma estrutura para a definição axiomática de probabilidade. Desse modo, no caso geral onde P(A)>0 temos a definição

$P(B\ /\ A)=\frac{P(A\cap B)}{P(A)}$ [25]

Isso nos possibilita a aplicação automática do conceito geral de probabilidade a todas as definições e resultados previamente apresentados (Gnedenko, 1978).

Suponha que dado evento B ocorre em apenas um dos n eventos mutuamente exclusivos A1, A2, ..., An. Em outras palavras, vamos assumir que

$B=BA_1+BA_2+...+BA_n$

Onde os eventos BA_i e BA_j com subscritos distintos e i e j são mutuamente exclusivos. Pelo teorema da adição, temos

$P(B)=P(BA_1)+P(BA_2)+...+P(BA_n)$

Do teorema da multiplicação resulta

$$P(B) = \sum_{i=1}^{n} P(A_i)(B/A_i)$$

[24] Para um aprofundamento no assunto, sugerimos a leitura de Gnedenko (1978) e Serge N. Bernstein (1943).

[25] Para o caso de P(A)=0, a probabilidade condicional P(A/B) é indefinida.

Essa relação é conhecida como a fórmula da probabilidade total e exerce um papel crucial na teoria da probabilidade (Gnedenko, 1978), além de nos permitir derivar a relevante fórmula de Bayes, que é central para o presente estudo.

Com efeito, suponha, com já discutido, que tenhamos

$$B = \sum_{i=1}^{n} BA_i$$

Calcula-se a probabilidade de A_i supondo a ocorrência de B aplicando-se o teorema da multiplicação.

$P(A_i \cap B) = P(B)P(A_i/\mathrm{B}) = P(A_i)P(B/A_i)$

Logo:

$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{P(B)}$

Com a fórmula da probabilidade total, temos que:

$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{\sum_{j=1}^{n} P(A_j)P(B/A_j)}$

A expressão acima e sua simplificação, $P(A_i/B) = \frac{P(A_i)P(B/A_i)}{P(B)}$, representam as célebres fórmulas de Bayes.

Conforme esclarece Gnedenko (1978), o procedimento geral para a aplicação da fórmula de Bayes na resolução de problemas práticos pode ser delineado da seguinte maneira: considere que o evento B pode ocorrer sob diversas condições, definidas por n hipóteses A1, A2, ..., An. As probabilidades associadas a essas hipóteses $P(A_i)$ são previamente conhecidas.

Além disso, conforme estabelecido, a probabilidade condicional P(B/A_i) nos informa que a hipótese A_i atribui uma probabilidade condicional P(B/A_i) ao evento B. Assim, como argumenta Gnedenko (1978), após a realização de um experimento onde o evento B ocorre, é necessário reavaliar as probabilidades das hipóteses A_i, sendo a fórmula de Bayes o método quantitativo utilizado para resolver esse problema (Gnedenko, 1978).

Como se verá, o teorema de Bayes, embora simples, tem aplicações muito interessantes. Como afirma Silver (2012), a matemática do teorema de Bayes é tão simples que surpreende, mas seus fundamentos filosóficos são profundamente ricos. Sua forma mais simples envolve apenas quatro variáveis, três delas já conhecidas, mas essa fórmula simples pode nos conduzir a um vasto campo de previsões e à possibilidade de raciocinar melhor em contextos de incerteza.

1.5 TEOREMA DE BAYES

Como pudemos observar a partir de cálculos simples, é possível derivar a conhecida fórmula de Bayes a partir dos axiomas da probabilidade. Em sua formulação mais simples, essa fórmula é escrita como $\boldsymbol{P(A/B) = \frac{P(B/A).P(A)}{P(B)}}$. Isso demonstra claramente que a validade do teorema de Bayes é inquestionável. Na verdade, mesmo antes dos axiomas de Kolmogorov, o teorema de Bayes já era demonstrado de forma simples; os axiomas serviram apenas para formalizar o processo.

Com efeito, a origem da discussão acerca do teorema de Bayes remonta ao século XVIII. A opção por deslocar a discussão sobre Bayes para essa parte do texto teve dois motivos. O primeiro, não dificultar a compreensão da evolução dos conceitos centrais acerca da probabilidade. O segundo, deixar para o final da primeira parte a análise do ponto principal do texto.

Retornemos um pouco na história, para compreender o contexto do surgimento da teoria bayesiana e o motivo de tanta controvérsia sobre tema que, como vimos acima, matematicamente é irretocável.

Thomas Bayes (1702-1761), cuja fama está associada ao teorema que leva seu nome, nasceu em uma família abastada e recebeu educação em casa, possivelmente sob a tutela de Abraham de Moivre, o que explicaria sua habilidade matemática e interesse pela teoria da probabilidade (Tabak, 2016). Bayes frequentou a Universidade de Edimburgo, atuou como ministro da Igreja Presbiteriana e foi eleito membro da Royal Society (Tabak, 2016). Reconhecido por sua modéstia, ele é descrito como um matemático notável, embora se saiba pouco sobre sua vida pessoal e profissional (Tabak, 2016).

Após seu falecimento, seu amigo Richard Price descobriu e publicou um ensaio de Bayes sobre probabilidade, *intitulado An Essay towards Solving a Problem in the Doctrine of Chances* (Tabak, 2016). Inicialmente, o trabalho recebeu pouca atenção, mas nos séculos seguintes, suas ideias ganharam importância e geraram considerável controvérsia na teoria da probabilidade.

O ensaio de Bayes, publicado com o auxílio de Price, contém os fundamentos do que seria conhecido como teorema de Bayes. Neste trabalho, Bayes apresentou o problema da mesa de bilhar para ilustrar sua teoria (Bayes, 1763). McGrayne (2011) oferece uma descrição ilustrativa da aplicação prática do teorema de Bayes, utilizando uma analogia em que Bayes, de costas para a mesa, tenta inferir a posição de uma bola branca com base nas posições relativas de outras bolas lançadas. Bayes descobriu que, à medida que o número de lançamentos aumentava, ele poderia reduzir o intervalo provável da posição da bola branca. A genialidade de Bayes estava em sua habilidade de inferir a posição da bola com informações limitadas, aproximando-se cada vez mais da certeza à medida que novos dados eram coletados (McGrayne, 2011).

Embora o teorema de Bayes seja relativamente simples, ele enfrenta críticas devido às suas suposições subjetivas. Alguns pesquisadores temem que essas suposições possam introduzir vieses; no entanto, como será discutido mais adiante, muitos exemplos demonstram que a experiência pode ser uma aliada poderosa (Tabak, 2016). Embora Bayes não tenha abordado todos os aspectos do tema com sucesso, ele forneceu uma introdução natural ao trabalho subsequente de Pierre-Simon Laplace (1749-1827), que aprofundou o teorema e mitigou o problema do viés do pesquisador (Stigler, 1986).

No século XVIII, a teoria da probabilidade era constituída por resultados dispersos. No entanto, no início do século XIX, Laplace unificou esses conceitos. Seu trabalho influenciou gerações de matemáticos, sendo sua obra *Théorie Analytique des Probabilités* considerada revolucionária (Tabak, 2016).

Após a morte de Laplace, a formalização da matemática tornou-se um foco central, destacando-se figuras como Frege, Hilbert e Kolmogorov. Kolmogorov, como mencionado anteriormente, estabeleceu os axiomas da probabilidade que são amplamente aceitos até hoje (Kolmogorov, 1956). Como já destacado, através de cálculos simples, presentes na maioria dos livros introdutórios de probabilidade, é possível partir dos axiomas de Kolmogorov e chegar à conhecida fórmula de Bayes[26].

[26] No presente texto, tendo em vista sua destinação, centramos nossas discussões em probabilidades discretas. Todavia, importante ressaltar que a discussão pode ser ampliada para o caso contínuo apenas com uma boa base de cálculo diferencial e integral. Dado o parâmetro de interesse, digamos θ, a distribuição a priori seria dada por $p(\theta)$, a função verossimilhança dos dados D por $p(D|\theta)$ e a distribuição a posteriori por $p(\theta|D)$. A relação entre as funções segue dada pelo teorema de Bayes. Assim, $p(\theta|D) = \frac{p(D|\theta).p(\theta)}{p(D)}$, onde $p(D) = \int p(D|\theta)\, p(\theta)\, d\theta$. Se olharmos para a integral como uma soma, é possível notar que o denominador, apesar desse formato estranho, não passa da partição encontrada no teorema de Bayes em sua forma geral $P(Hi|B)=\frac{P(B|Hi).P(Hi)}{\Sigma_j P(B|Hj).P(Hj)}$. Para uma visão geral do método bayesiano em ambientes contínuos, consulte o capítulo 4 de Gnedenko et al. (1999); para uma análise mais detalhada, veja Casella e Berger (2011).

$$\boldsymbol{P(A|B) = \frac{P(B|A).P(A)}{P(B)}}.$$

Um exemplo amplamente conhecido da aplicação do teorema de Bayes é o famoso problema da escola de medicina de Harvard. Nesse problema, um teste diagnóstico para determinada doença (D) pode resultar apenas em positivo ou negativo, indicando a presença ou ausência da doença. Estima-se que a probabilidade de um falso negativo seja 0% e a probabilidade de um falso positivo seja 5%. A taxa de incidência da doença é baixa, com um levantamento indicando que ocorre um caso por mil habitantes na população. Se uma pessoa selecionada ao acaso na população for submetida ao teste e o resultado for positivo, qual é a probabilidade de essa pessoa ter a doença D?

A maioria das pessoas submetidas a esse problema respondem que a probabilidade é de 95%. No entanto, a resposta correta é inferior a 2%. Aplicando o teorema de Bayes, a solução seria a seguinte:

O que desejamos saber é a probabilidade de ter a doença (D), dado que o teste foi positivo (P).

P(D/P) =?

P(D/P) = $\frac{P(P/D).\ P(D)}{P(P)}$

Para a solução precisamos dos valores à direita da equação.

P(P/D) = 100%, uma vez que não temos resultado falso negativo. Se tem a doença, o resultado será positivo.

P(D)=1/1000, a ocorrência da doença na população é de um caso para cada mil habitantes.

P(P)= 1/1000 + 5%. (999/1000), tem a doença e deu positivo + não tem a doença e deu positivo

Portanto: P(D/P) = 1,96%

Note que a solução desse problema é incontroversa. Estamos lidando com um teorema, uma consequência direta dos axiomas de Kolmogorov, onde foram aplicados valores obtidos por meio de cálculo de frequência. Esse tipo de solução não é alvo de críticas dos frequencistas. A controvérsia sobre o teorema de Bayes reside, como já mencionado, em seu uso subjetivo, na escolha arbitrária da probabilidade prévia, o que não é o caso do exemplo acima, pois ela é dada como registro empírico da frequência da doença. Como discutiremos na segunda parte do texto, esse é também o grande problema de seu uso na filosofia.

Por esse motivo, antecipamos que, sendo rigorosos, não há uma equivalência exata entre os termos bayesianismo e probabilidade subjetiva, embora, como se verá mais adiante, o uso do bayesianismo, especialmente na filosofia, esteja certamente associado à probabilidade subjetiva.

Antes de avançarmos para outro exemplo, é importante mencionar que, na ciência forense, frequentemente há uma escassez de dados, tornando o cálculo da razão de verossimilhança muitas vezes desafiador (Aitken & Taroni, 2004). A abordagem bayesiana, em sua forma mais útil, como veremos na segunda parte deste texto, considera probabilidades como medidas de crença (probabilidade subjetiva). Isso permite a combinação de probabilidades objetivas (baseadas em dados) com probabilidades subjetivas, de modo que o conhecimento e a experiência forense possam auxiliar na obtenção de estimativas (Aitken & Taroni, 2004).

Um exemplo típico de aplicação forense da estatística bayesiana é o seguinte: imagine que, em uma investigação criminal, sejam encontradas duas evidências, uma impressão digital e um fio de cabelo, no local do crime. Para que o problema seja relevante, considerando a alta confiabilidade do método Vucetich e, especialmente, do exame de DNA, podemos supor que a impressão digital seja apenas parcial, dependendo da experiência do papiloscopista para discutir a probabilidade, e que o DNA no fio de cabelo esteja parcialmente degradado. Os cálculos poderiam ser realizados da seguinte forma:

Inicialmente os investigadores consideram a evidência da impressão digital encontrada no local do crime. A probabilidade prévia (*prior*) de que o suspeito seja culpado (P(Suspeito)) pode ser atualizada com base na probabilidade da impressão digital ser do suspeito (P(Impressão/Suspeito) e a probabilidade de encontrar aquela impressão digital (P(Impressão)).

$$\text{P(Suspeito/Impressão)} = \frac{P(Impressão/Suspeito).\ P(Suspeito)}{P(Impressão)}$$

Onde:

P(Suspeito/Impressão) é a probabilidade de que o suspeito seja o culpado, dado que a impressão digital foi encontrada.

P(Impressão/Suspeito) é a probabilidade de encontrar a impressão digital se o suspeito for culpado.

P(Suspeito) é a probabilidade prévia de o suspeito ser o culpado.

P(Suspeito/Impressão,Cabelo) é a probabilidade do culpado ser o suspeito dada a impressão digital e o fio de cabelo.

Após considerar a impressão digital, os investigadores adicionam a evidência do fio de cabelo. A probabilidade atualizada após a impressão digital (P(Suspeito/Impressão)) é novamente atualizada com a nova evidência (P(Cabelo/Suspeito) e P(Cabelo/Impressão)).

$$P(\text{Suspeito/Impressão,Cabelo}) = \frac{P(Cabelo/Suspeito,Impressão).\ P(Suspeito/Impressão)}{P(Cabelo/Impressão)}$$

Onde:

P(Suspeito/Impressão,Cabelo) é a probabilidade de que o suspeito seja o culpado, dado que a impressão digital e o fio de cabelo foram encontrados.

P(Cabelo/Suspeito,Impressão) é a probabilidade de encontrar o fio de cabelo se o suspeito for o culpado e a impressão digital for encontrada.

P(Suspeito/Impressão) é a probabilidade atualizada de que o suspeito é o culpado, dada a impressão digital.

P(Cabelo/Impressão) é a probabilidade de encontrar aquele fio de cabelo, dado que a impressão digital foi encontrada.

McGrayne (2011, pp. 242ss) discute um uso amplamente difundido da estatística bayesiana: os filtros de spam. Segundo a autora, a técnica de Bayes é utilizada por empresas como o Google para classificar spam, pornografia e identificar palavras, frases e documentos relacionados. A filtragem de spam em e-mails é um exemplo clássico da aplicação da estatística bayesiana com múltiplas evidências. Os métodos bayesianos analisam palavras e frases da mensagem para determinar a probabilidade de ser spam. Esse processo é reiterado com cada nova evidência, alcançando até 99,99% de certeza (McGrayne, 2011).

O sistema de filtragem de spam utiliza diversas evidências nos e-mails para determinar se são spam ou não. Essas evidências podem incluir a presença de determinadas palavras, o nome do remetente, a estrutura do e-mail, a existência de links, anexos, entre outros. A fórmula de Bayes combina essas evidências para calcular a probabilidade de um e-mail ser spam. A seguir, vejamos um exemplo para ilustrar melhor essa aplicação.

Suponha que S represente o fato de o e-mail ser spam; L, o fato de ser legítimo; e E1, E2, ... En, as evidências observadas.

A aplicação do teorema de Bayes pode ocorrer da seguinte maneira:

Antes analisarmos qualquer evidência é necessário partir de uma probabilidade inicial de que o e-mail seja um spam. Chamamos essa probabilidade de P(S). E de que o e-mail seja legítimo. Chamaremos esse segundo caso de P(L). Suponha que, com base em dados históricos, saibamos que em torno de 20% dos e-mails seja spam. Assim, P(S)=0,2 e P(L)=0,8.

Para cada evidência Ei, teremos a probabilidade condicional de observar Ei dado que o e-mail é spam, P(Ei/S), e a probabilidade condicional de observar Ei dado que o e-mail é legítimo, P(Ei/L). Essas probabilidades podem ser estimadas com base em um conjunto de dados de e-mails previamente classificados.

Suponha que no exemplo em estudo tenhamos:

E1: O remetente do e-mail contém a palavra *coaching*.

E2: O e-mail traz um link para um site cujo domínio é desconhecido.

E3: O e-mail foi enviado a um elevado número de destinatários.

Agora suponha que com base em dados históricos de e-mails seja possível definir probabilidades condicionais da seguinte maneira:

P(E1/S)=80%   P(E1/L)=10%[27]

P(E2/S)=60%   P(E2/L)=20%

---

[27] Alguns autores preferem exemplificar esse percentual de forma complementar. Assim, se a probabilidade de ser spam for de 60% a probabilidade de não ser será de 40%. Isso parece adequado, especialmente quando se está tratando de histórico. O que acontece é que às vezes não dá para identificar se o e-mail era realmente spam e esses casos acabam por prejudicar essa complementaridade.

P(E3/S)=40%   P(E3/L)=30%

Com base nesses valores, podemos calcular a probabilidade de que um determinado e-mail, que apresente as evidências destacadas, seja um spam. Para isso, utilizamos a fórmula da Bayes.

$$P(S/E1,E2,E3) = \frac{P(E1,E2,E3)/S).\ P(S)}{P(E1,E2,E3)}$$

Onde,

P(E1,E2,E3/S)=P(E1/S).P(E2/S).P(E3/S)

P(E1,E2,E3/L)=P(E1/L).P(E2/L).P(E3/L)

P(E1,E2,E3)=P(E1,E2,E3/S).P(S)+P(E1,E2,E3/L).P(L)

Admitindo que as evidências são condicionalmente independentes, temos:

P(E1,E2,E3/S) = 0,8.0,6.0,4=0,192

P(E1,E2,E3/L)=0,1.0,2.03=0,006

Disso resulta que:

P(E1,E2,E3)=(0,192.0,2)+(0,006.0,8)=0,0384+0,0048=0,0432

Assim, temos:

$$P(S/E1,E2,E3) = \frac{0{,}192.0{,}2}{0{,}0432} = \frac{0{,}0384}{0{,}0432} = 0{,}889$$

Logo, a probabilidade de que o e-mail em análise seja spam, dado que no remetente aparece a palavra *coaching*, o e-mail traz um *link* para um site cujo domínio é desconhecido e o e-mail foi enviado a um elevado número de destinatários é de aproximadamente 90%.

Durante as décadas de 1930 e 1940, o uso do teorema de Bayes quase desapareceu. Um nome crucial para sua sobrevivência nesse período foi Harold Jeffreys, que manteve a discussão do bayesianismo praticamente sozinho, em meio a intensos ataques (McGrayne, 2011). Um sucesso notável ocorreu durante a Segunda Guerra Mundial, quando Alan Turing utilizou o teorema de Bayes para decifrar a criptografia da máquina Enigma, contribuindo significativamente para a defesa da Grã-Bretanha e para o fim da guerra, além de impulsionar o desenvolvimento dos computadores modernos (McGrayne, 2011). No entanto, a aplicação de Bayes nos esforços de guerra não se tornou conhecida do grande público na época, devido ao sigilo imposto sobre esses esforços e sucessos de guerra (McGrayne, 2011).

Após o período de guerra, Winston Churchill, primeiro-ministro do Reino Unido durante a Segunda Grande Guerra, ordenou a destruição de todas as evidências de que a decodificação realizada com o auxílio do teorema de Bayes ajudou a vencer a guerra, adiando o reconhecimento do teorema até a década de 1970 (McGrayne, 2011). A história do projeto de Turing desenvolvido em Bletchley Park e seu sucesso com a máquina Enigma só foi revelada após a substituição das máquinas Lorenz[28] por novos criptossistemas (McGrayne, 2011).

Ronald Fisher (1890-1962) foi um grande crítico do bayesianismo, considerando-o um grave erro (Aylmer Fisher, 1950). Jeffreys, por outro lado, era um defensor do bayesianismo objetivo, entendendo a probabilidade como uma medida objetiva de crença (Gorroochurn,

[28] Embora a máquina Enigma seja a mais famosa da Segunda Grande Guerra, a máquina Lorenz, também conhecida como *Tunny*, era uma evolução da Enigma, utilizando doze rotores em vez dos três usados pela Enigma, resultando em um número muito maior de combinações possíveis (Roberts, 2017). Utilizada por Hitler, seu alto comando e pelos principais generais, a Lorenz era pouco conhecida fora deste círculo (Roberts, 2017). Enquanto a Enigma era empregada para comunicações aéreas, terrestres e marítimas desde 1923, a Lorenz foi utilizada apenas a partir de 1940 (Roberts, 2017). A Enigma foi decifrada por Turing em 1941, com essa conquista sendo desclassificada na década de 1970, enquanto a Lorenz foi decifrada por Bill Tutte, com a revelação desse feito ocorrendo apenas em 2002 (Roberts, 2017).

2016). A disputa entre Fisher e Jeffreys, dois renomados estatísticos, ocorreu na década de 1930 e se tornou amplamente conhecida no meio estatístico, especialmente pela intensidade das posições de Fisher (McGrayne, 2011). No final, os dois acabaram se tornando amigos, apesar das divergências teóricas (McGrayne, 2011).

A principal objeção de Fisher ao teorema de Bayes estava na noção de probabilidade prévia, que ele considerava subjetiva. Fisher não aceitava o uso de probabilidades que não decorressem de métodos matemáticos robustos. Para ele, não fazia sentido utilizar valores cuja correção não pudesse ser afirmada com certeza. É importante destacar que Silver (2012) argumenta que o teorema de Bayes permite que as crenças iniciais convirjam para a verdade à medida que mais evidências são apresentadas, tema que será explorado em maior detalhe na segunda parte deste texto. O bayesianismo ganhou força novamente na década de 1990, com o aumento de publicações e os avanços computacionais que facilitaram o cálculo de probabilidades posteriores (Corfield; Williamson, 2001).

O bayesianismo teve grande influência na inteligência artificial, com a adoção de técnicas probabilísticas e redes bayesianas (Corfield; Williamson, 2001). Nos últimos anos, a disputa entre bayesianos e frequencistas diminuiu, levando a uma aproximação entre grandes estatísticos e resultando em muitas aplicações práticas recentes que combinam métodos estatísticos bayesianos e frequentistas (McGrayne, 2011). Muitos frequencistas proeminentes, como Bradley Efron, tornaram-se mais moderados em suas posições, reconhecendo a importância do raciocínio bayesiano para o avanço da ciência (McGrayne, 2011).

Atualmente, como ilustrado pelos poucos exemplos apresentados, o teorema de Bayes está presente em inúmeras áreas relacionadas à análise de dados. Com o aumento do volume de dados e da capacidade computacional, a tendência é que sua aplicação se expanda ainda mais para diversas áreas (McGrayne, 2011).

No próximo capítulo, aprofundaremos a discussão sobre o bayesianismo, com um foco especial na filosofia, uma área onde o uso do bayesianismo tem crescido significativamente. Aproveitaremos para explicar melhor o impacto da probabilidade subjetiva, que, como mencionado anteriormente, foi o principal motivo das críticas acadêmicas ao bayesianismo.

2 ESTATÍSTICA E A FILOSOFIA DA CIÊNCIA

A relação entre a estatística e a filosofia da ciência é extremamente rica e produtiva, abrangendo uma ampla gama de questões e autores de grande relevância. As discussões percorrem um vasto campo filosófico, abordando questões epistemológicas, metodológicas e ontológicas. Ao longo dos séculos, a relação entre essas disciplinas evoluiu significativamente, sendo moldada pelas contribuições de figuras influentes e pelos contextos históricos que impulsionaram o desenvolvimento de métodos estatísticos e filosóficos.

O desenvolvimento dessa área integrada de conhecimento foi influenciado por grandes personalidades de diversas disciplinas. Pela sua importância histórica, podemos destacar Pierre-Simon Laplace, Thomas Bayes, Ronald A. Fisher, Jerzy Neyman, Egon Pearson, Leonard J. Savage, David Hume, Bruno de Finetti, Frank P. Ramsey e John Maynard Keynes, entre muitos outros. Cada um desses pensadores trouxe avanços significativos, seja na formulação de teorias probabilísticas, no desenvolvimento de métodos inferenciais ou na reflexão sobre a natureza do conhecimento científico.

Nas próximas seções, buscaremos esclarecer a influência de alguns desses pensadores no desenvolvimento e aplicação do bayesianismo na filosofia, destacando o quanto suas teorias contribuíram para a evolução da sua estrutura teórica.

2.1 A Probabilidade e a Filosofia da Ciência

Conforme já mencionado, a intersecção entre estatística e filosofia da ciência se apresenta como um campo rico em possibilidades de pesquisa. Tal contexto implica a necessidade de uma escolha criteriosa dos temas, mesmo para uma abordagem preliminar como a que se propõe neste trabalho. Central para este campo é o conceito de inferência científica, que explora as modalidades através das quais conclusões robustas podem ser extraídas de dados empíricos. A estatística, com seu conjunto de ferramentas e modelos matemáticos, é instrumental nesse processo, facilitando a análise e interpretação de dados para inferências científicas. Como mencionado, emergem predominantemente duas abordagens estatísticas na filosofia da ciência: a frequencista e a bayesiana.

O presente texto concentra-se nessas duas abordagens estatísticas predominantes dentro do discurso filosófico-científico, bem como em discussões complementares que enriquecem a compreensão do assunto. É importante salientar que este texto não tem a pretensão de ser exaustivo; as delimitações temáticas foram guiadas pela necessidade de proporcionar uma exposição clara e abrangente do tema. Naturalmente, muitos aspectos relevantes serão omitidos, o que pode servir como estímulo para pesquisas subsequentes.

À medida que avançamos na exploração da relação entre estatística e filosofia, torna-se essencial obter uma compreensão mais clara dos conceitos discutidos até o momento. Isso permitirá uma apreciação mais profunda dos fundamentos envolvidos no tema, evitando equívocos que poderiam obstruir o progresso das discussões teóricas.

A teoria da probabilidade abrange dois aspectos distintos: o matemático e o fundacional ou filosófico, que frequentemente apresentam contrastes marcantes (Gillies, 2000). Embora haja um grande consenso quanto à matemática subjacente à teoria da probabilidade, as divergências de opinião são substanciais no que diz respeito à sua filosofia (Gillies, 2000). Com algumas exceções, a maioria dos probabilistas aceita um conjunto comum de axiomas matemáticos, resultando em um acordo generalizado sobre os teoremas derivados (Gillies, 2000). No entanto, ao longo do século XX, surgiram quatro interpretações notavelmente distintas deste cálculo matemático, cada uma das quais continua a ter defensores na atualidade (Gillies, 2000).

De acordo com Gillies, as quatro principais interpretações da probabilidade são: (i) a teoria lógica, que identifica a probabilidade com o grau de crença racional; (ii) a teoria subjetiva, que define a probabilidade como o grau de crença de um indivíduo específico; (iii) a teoria da frequência, que associa a probabilidade a uma frequência limite após uma longa série de eventos semelhantes; e (iv) a teoria da propensão, que considera a probabilidade como uma tendência inerente a um conjunto de condições repetíveis.

Conforme já mencionado, neste estudo abordaremos a interpretação frequencista e a interpretação da propensão como espécies da conceito de probabilidade física, enquanto as demais serão tratadas como espécies de probabilidade epistêmica. Ressalta-se que não há consenso sobre os conceitos de probabilidade, e que a divisão proposta por Gillies, embora amplamente utilizada, não é a única. Nosso estudo começará com uma distinção ligeiramente diferente, conforme antecipado.

Assim, dividiremos a probabilidade em dois grupos: probabilidade física e probabilidade epistêmica. A probabilidade física, que expressa estados de coisas físicas e é comumente chamada simplesmente de probabilidade, é mais proeminente nas ciências empíricas (Romeijn, 2022). Esta se refere à probabilidade associada à propriedade de um tipo de evento, podendo ser entendida como a frequência ou a tendência com que essa propriedade se manifesta em uma série de eventos desse tipo (Romeijn, 2022).

Romeijn (2022) utiliza o exemplo clássico do lançamento de moedas para ilustrar a ideia de probabilidade física. Ele explica que a probabilidade de obter *cara* em um lançamento é 50%

se, em uma série de lançamentos similares, *cara* ocorrer em “metade das vezes”[29], conforme detalhado anteriormente no texto. Alternativamente, essa probabilidade é considerada 50% se houver uma propensão igual para ambos os resultados possíveis na configuração do lançamento da moeda.

Como visto, o conceito de probabilidade física inclui a probabilidade frequencista, entre outras. A probabilidade frequencista estrutura a chamada estatística clássica (ou frequencista), que não deve ser confundida com a probabilidade clássica, outra forma de probabilidade física[30] baseada na relação entre casos possíveis e favoráveis, conforme amplamente explorado na primeira parte do texto.

Na estatística clássica, a frequência dos eventos é o principal foco de análise. Segundo Romeijn (2022), a interpretação frequencista da probabilidade, proposta por von Mises, é central nessa abordagem estatística. Conforme já discutido, nessa interpretação, as probabilidades são entendidas como frequências ou proporções dentro de uma classe de eventos, onde essas proporções definem as chances. Não é exagero repetir que essa visão difere da teoria da propensão, na qual a probabilidade de um evento ou objeto individual é vista como uma tendência inerente à natureza. Essa distinção será relevante para o aprofundamento que faremos nas próximas seções.

No decorrer da primeira metade do século XX, a abordagem frequentista consolidou-se como a mais proeminente entre os estatísticos, conforme observado por Paulino et al. (2018). Essa perspectiva ainda se mantém relevante nos dias atuais; no entanto, a interpretação bayesiana tem se fortalecido e, atualmente, conquista um espaço de interesse equivalente à teoria frequencista no meio acadêmico.

---

[29] Na primeira parte, explicamos que não se trata de obter exatamente metade de caras (número de caras igual ao número de coroas), mas sim que a fração do número de caras sobre o total de lançamentos se aproxima cada vez mais de 50%.
[30] Como já discutido, há divergência no ponto.

Antes de avançarmos nas discussões sobre a estatística clássica, frequentemente referida como estatística frequencista, é fundamental compreender o conceito de probabilidade epistêmica, que se contrapõe à probabilidade física, incluindo a abordagem frequentista que fundamenta a estatística clássica. Esta sequência de ideias foi estabelecida para esclarecer quaisquer equívocos sobre o ponto da discussão em que nos encontramos, a partir do qual continuaremos a desenvolver o tema.

É relevante notar que a discussão sobre probabilidade epistêmica não é tão prevalente no contexto matemático e estatístico brasileiro quanto é na filosofia, em parte devido à crítica prolongada à probabilidade subjetiva dentro dessas disciplinas. Contudo, com o avanço do bayesianismo, as noções associadas à probabilidade subjetiva têm ganhado destaque globalmente, estimulando debates não apenas em grandes centros acadêmicos, mas em todo o mundo.

A concepção epistêmica da probabilidade emergiu no século XIX com pioneiros como De Morgan e Boole, e foi posteriormente expandida por Keynes, Ramsey e de Finetti (Romeijn, 2022). O tema foi enriquecido por teóricos da decisão, filósofos e lógicos indutivos, incluindo nomes como Carnap, Savage, Levi e Jeffrey, conforme apontado por Romeijn (2022). No âmbito estatístico, destacam-se figuras como Jeffreys, Edwards, Lindley, Good e Jaynes, além de filósofos bayesianos contemporâneos como Goldstein, Kadane, Berger e Dawid (Romeijn, 2022). Esses estudiosos defendem uma visão que coloca a probabilidade no domínio epistêmico, não como uma propriedade intrínseca ao mundo físico, como sugere a teoria da propensão, mas como um instrumento de modelagem de sistemas representativos, como a mente humana (Romeijn, 2022).

De acordo com Romeijn (2022), a probabilidade epistêmica, assim como a probabilidade física, pode ser dividida em diferentes tipos. Nesse contexto, a probabilidade epistêmica pode ser vista como representativa de uma atitude doxástica, destinada a especificar as crenças de um agente racional acerca de dados e hipóteses. Assim, a probabilidade estaria associada ao grau ou intensidade de crença que esse agente racional possui em relação à veracidade de uma determinada opinião (Romeijn, 2022).

Romeijn (2022) argumenta que a probabilidade epistêmica pode também ser interpretada com base na teoria da decisão, influenciando as predisposições de um agente racional para tomar decisões e agir com base em dados e hipóteses. Além disso, a probabilidade epistêmica pode ser vista sob uma perspectiva lógica, servindo como uma representação formal que estabelece a relação entre premissas e conclusão, proporcionando um padrão para o raciocínio sob incerteza.

Explorando mais a fundo as subdivisões, percebe-se que elas nem sempre são distintas e independentes. Por exemplo, a concepção doxástica, como vimos anteriormente, frequentemente se apoia na teoria da decisão para determinar o grau de crença e demonstrar sua representação pela teoria matemática da probabilidade, conforme discutido por Romeijn (2022). Um exemplo recorrente dessa relação é o uso do *livro holandês* como argumento de racionalidade (Romeijn, 2022).

Além disso, Romeijn (2022) sugere que dentro da concepção doxástica é possível diferenciar entre a doxástica subjetiva e a doxástica objetiva. Na doxástica subjetiva, os agentes são livres para manter as crenças que considerarem apropriadas, com a única restrição sendo os axiomas da probabilidade. Por outro lado, a doxástica objetiva impõe critérios adicionais de racionalidade que dependem da teoria em análise. Na verdade, mesmo a restrição imposta à doxástica subjetiva associada aos axiomas é, por vezes, amenizada por alguns autores, como discutido na primeira parte deste texto. Isso levanta questões sérias acerca da racionalidade dos resultados que daí decorrem.

Segundo De Finetti (2017), em determinados casos, a formulação de um problema ou a justificação de um raciocínio exige uma escolha entre uma interpretação subjetiva inquestionável e uma interpretação supostamente objetiva. A primeira se basearia nas opiniões ou atitudes de uma pessoa específica, enquanto a última resultaria de uma transposição confusa dessa opinião para o complexo indefinível de circunstâncias objetivas que poderiam ter contribuído para sua determinação. De Finetti sugere que a melhor escolha é a primeira. Ele expressa uma preocupação compartilhada por outros autores da época sobre a busca incessante pela objetividade, que pode impactar significativamente a aplicação teórica, embora os ganhos sejam, em alguns casos, bastante limitados. A opinião subjetiva, conhecida pelo indivíduo em questão, seria, nesse sentido, algo objetivo e poderia ser objeto de um estudo rigoroso (De Finetti, 2017).

Em relação à suposta contradição de tratar uma opinião subjetiva como um dado objetivo, De Finetti (2017) sustenta que as afirmações adquirem um significado objetivo quando podem ser avaliadas como verdadeiras ou falsas com base em uma observação claramente definida. Nesse contexto, é admitida uma considerável amplitude de variação, contanto que não ocorra qualquer forma de manipulação (De Finetti, 2017). Ao expandir o conceito de "objetivo" dessa forma, De Finetti parece refletir uma preocupação predominante na sua época, relacionada à busca pelo objetivismo. Como se verá, a evolução computacional e o uso de evidências posteriores, cujo uso o próprio De Finetti promoveu, contribuíram para ajustar essa opinião, mitigando a preocupação excessiva com a objetividade inicial. Parece que a questão está intrinsecamente ligada à necessidade de um suporte social, especialmente em virtude da aversão demonstrada pelos cientistas do período em relação à adoção de teorias subjetivas.

De Finetti (2017) proporciona uma análise perspicaz sobre como a teoria da decisão pode fornecer importantes percepções para questões dentro da estatística matemática, e sobre como aquela esclarece comparações entre as metodologias bayesiana e objetivista, com esta última sendo descrita por ele como atualmente mais popular. É importante notar que, neste contexto, o termo *bayesiano* é frequentemente usado como sinônimo de *subjetivo*. Embora isso seja comum, não é a abordagem que adotamos neste texto. A justificativa para essa escolha será detalhada posteriormente, juntamente com uma exploração mais aprofundada do bayesianismo, à medida que o tema for desenvolvido.

Dentre as numerosas contribuições de De Finetti (2017), destaca-se sua exploração sobre a maximização da utilidade esperada na tomada de decisões. Simplificadamente, ele propõe que uma decisão eficaz deve ser alcançada ao primeiro avaliar os ganhos individuais de utilidade associados às consequências de diferentes opções disponíveis e, posteriormente, equilibrar esses ganhos com as respectivas probabilidades (De Finetti 2017). Portanto, decisões devem ser fundamentadas em probabilidades posteriores, determinadas a partir do conjunto completo de informações disponíveis até aquele momento, o que se constitui como o aspecto central da análise segundo De Finetti (2017).

De acordo com De Finetti (2017), a tomada de decisão eficaz requer inicialmente uma fundamentação em teoria estatística que possa oferecer conclusões na forma de probabilidades posteriores, calculadas a partir de todas as informações disponíveis até então. Essa abordagem é aplicada com sucesso pela abordagem bayesiana, que se distingue das outras metodologias nesse aspecto, como será visto posteriormente no texto.

Segundo a interpretação lógica, conforme descrito por Romeijn (2022), os valores de probabilidade atribuídos a dados e hipóteses desempenham um papel análogo aos valores de verdade na lógica dedutiva. Esses valores têm a função de assegurar uma inferência válida, sem implicar que os números correspondam a elementos de destaque psicológico.

A concepção de probabilidade lógica, que depende da evidência disponível, remonta a mais de duzentos anos e foi sistematicamente articulada pela primeira vez por John Maynard Keynes (Hacking, 2001). Keynes, um renomado economista cujas teorias são frequentemente creditadas por ajudar a salvar o capitalismo durante a Grande Depressão de 1929, propôs essa ideia em sua obra *A Treatise on Probability* (1921). Sua abordagem sugeria que a probabilidade poderia ser vista como uma relação lógica entre proposições, baseada na evidência disponível, o que atraiu a atenção de lógicos e filósofos, como Rudolf Carnap, que viu nessa abordagem uma maneira de formalizar a inferência indutiva (Hacking, 2001). No entanto, a concepção de Keynes enfrentou críticas significativas, como as de Frank P. Ramsey, que argumentava que a probabilidade lógica não capturava adequadamente a incerteza subjetiva e a crença pessoal (Hacking, 2001). Essa divergência entre a visão lógica e a subjetiva da probabilidade desencadeou debates profundos que influenciaram o desenvolvimento da teoria da probabilidade e da estatística ao longo do século XX.

Keynes (1921) argumenta que quando temos duas proposições — uma como premissa e outra como conclusão — estabelece-se entre elas uma única relação de um tipo específico, denominada relação de probabilidade. Se essa relação possuir um grau ***a***, então, a partir de uma crença total na premissa, racionalmente deveríamos sustentar uma crença de mesmo grau ***a*** na conclusão. A interpretação lógica estabelece uma relação dedutiva entre um conjunto de proposições, consideradas como evidências, e uma proposição adicional, vista como hipótese. Essa relação é avaliada em termos do grau de implicação (ou confirmação, conforme proposto por Carnap) que as evidências oferecem à hipótese (Paulino et al., 2018). Os defensores dessa perspectiva argumentam que esse grau de implicação é único, racional e impessoal, promovendo, assim, um ideal de objetividade. No entanto, a dificuldade em quantificar tal grau de implicação frequentemente limita a aplicabilidade prática desta abordagem, exceto nos casos em que a probabilidade é claramente 0 ou 1 (Paulino et al., 2018).

Retomando a concepção de probabilidade como uma atitude doxástica, conforme abordado por Romeijn (2022), esta perspectiva trata de definir opiniões sobre dados e hipóteses a partir da visão de um agente racional idealizado, refletindo, assim, o grau de crença na validade de uma determinada opinião. Essa abordagem doxástica contrasta com a visão lógica ao enfatizar a subjetividade inerente à formação de crenças, onde a probabilidade é vista como uma medida do grau de crença individual, baseada em evidências disponíveis e adaptável conforme novas informações são adquiridas.

A análise da probabilidade lógica e da probabilidade doxástica ressalta a intricada tarefa de identificar um único modelo probabilístico que satisfaça tanto as exigências teóricas quanto as práticas. Enquanto a interpretação lógica almeja um padrão de implicação que seja unívoco e objetivo, a abordagem doxástica aceita a variabilidade e a subjetividade inerentes às crenças individuais.

A epistemologia bayesiana, por sua vez, foca em um tipo específico de atitude doxástica, também referida como grau de crença, grau de confiança ou credibilidade, conforme discutido por Titelbaum em *Bayesian Epistemology* (2022). Filósofos da mente de língua inglesa frequentemente empregam o termo *crença* para descrever a atitude adotada quando assumimos algo como verdadeiro (Schwitzgebel, 2024). A formação de crenças é um aspecto fundamental da mente humana e desempenha um papel central tanto na epistemologia quanto na filosofia da mente. Não é surpresa que grande parte da epistemologia esteja dedicada a explorar a justificação de nossas crenças e a construção do conhecimento a partir delas (Schwitzgebel, 2024).

De acordo com Schwitzgebel (2024), grande parte dos filósofos contemporâneos define crença como uma atitude proposicional. Isso é entendido como um estado mental que envolve uma atitude mental específica e, consequentemente, uma ação em relação a uma proposição ou circunstâncias nas quais essa proposição se mostra verdadeira. Schwitzgebel (2024) observa que não é difícil imaginar que uma pessoa possa ter uma crença com variados graus de confiança, onde uma maior confiança representaria uma maior disposição em confiar nessa crença para modelar ações.

Para que essa ideia possa ser mais facilmente transposta para a noção de probabilidade, como vimos na primeira parte do texto, é comum quantificá-la em uma escala que varia de zero a um. Nesse contexto, zero representa certeza quanto à falsidade da proposição, e um representa certeza quanto à sua veracidade. Os valores intermediários correspondem aos variados graus de confiança na veracidade ou falsidade da proposição. Quanto mais próximo de um, mais nos aproximamos da certeza de veracidade; quanto mais próximo de zero, mais nos aproximamos da certeza de falsidade. Essa quantificação da crença é essencial para a epistemologia bayesiana, que utiliza esses graus tanto inicialmente quanto para a atualização de crenças com base em novas evidências apresentadas, alinhando-se assim à abordagem probabilística e racionalista defendida por pensadores como De Finetti (Schwitzgebel, 2024).

De acordo com R. Jeffrey (1965), a probabilidade, quando vista de uma perspectiva subjetiva, representa uma crença parcial. Assim, afirmar que um agente atribui uma probabilidade de 0,7 à previsão de chuva para amanhã equivale a dizer que ele possui um grau de crença de 0,7 nessa proposição. Isso implica, aproximadamente, que ele estaria disposto a investir $0,70 para receber $1 caso realmente chova amanhã, e nada se não ocorrer chuva.

Relacionado a essa afirmação, é importante destacar que Keynes argumenta que nem sempre é possível quantificar essa probabilidade numericamente, o que complica a abordagem do tema sob a ótica de uma igualdade matemática. Ramsey (1926), por sua vez, sugere que seria mais adequado afirmar que existe uma correspondência unívoca entre as relações de probabilidade e os graus de crença justificados por elas.

Jeffrey (1965) reconhece que a caracterização numérica mencionada é aproximada, pois a interpretação mais precisa das declarações de probabilidade, segundo ele, seria fornecida pela teoria de Ramsey. No entanto, ele considera que essa abordagem, embora menos exata, é extremamente útil, especialmente porque serve para justificar as premissas da teoria elementar das probabilidades. Embora este texto não pretenda explorar esse tópico em profundidade, é relevante mencionar que a obra de Jeffrey (1965) apresenta uma discussão valiosa sobre o tema, introduzindo o conceito de *desejabilidade*. Esse conceito, juntamente com a probabilidade, constitui a fundação para a classificação das preferências de um agente, conforme a teoria desenvolvida por Ramsey.

Para encerrar esta discussão e destacar a relevância das ideias apresentadas por Jeffrey (1965), é pertinente mencionar uma reflexão de Schwitzgebel (2024), que critica o termo "grau de crença" por sua potencial ambiguidade. Ele exemplifica com o caso de uma pessoa racional que possui um bilhete de uma loteria justa; tal pessoa pode considerar uma boa possibilidade de ganho apesar da probabilidade extremamente baixa. Isso levanta questionamentos sobre a racionalidade do apostador que participa de uma loteria, onde a esperança[31] de ganho é muito inferior ao valor investido, mas ainda assim mantém a crença na possibilidade de sucesso (Schwitzgebel, 2024).

Essa reflexão sublinha a complexidade das atitudes doxásticas e a importância de compreender como os agentes racionais formam e ajustam suas crenças diante de incertezas e probabilidades, um tema central na epistemologia bayesiana e na teoria da decisão.

Parece aceitável, como argumenta Keynes, que o grau de uma relação de probabilidade corresponda ao grau de crença que ela justifica, assumindo, portanto, que tanto a relação de

---

[31] De forma simplificada, a esperança matemática, também conhecida como valor esperado, pode ser entendida como o valor médio que esperaríamos obter se repetíssemos a experiência (ou jogo) indefinidamente (Triola, 1998). A esperança matemática possui diversas aplicações e desempenha um papel crucial na teoria da decisão (Triola, 1998).

probabilidade quanto o grau de crença possam ser expressos numericamente de forma adequada. Assim, o número que quantifica a relação de probabilidade coincide com aquele que representa o grau de crença correspondente (Ramsey, 1926). Ramsey (1926) também destaca a importância da existência de hierarquias de maior e menor, de modo a alinhar as relações de probabilidade e os graus de crença de maneira semelhante ao que era defendido por Russell.

Para avançarmos na conclusão deste tópico, é essencial compreender mais claramente nossa posição atual e como a objetividade e a subjetividade se entrelaçam com o que discutimos até agora. Não será produtivo analisar cada questão que surja com todas as nuances complexas a que a probabilidade está sujeita. Portanto, vamos focar nossa discussão em duas categorias principais que denominaremos de probabilidade objetiva e probabilidade subjetiva. Esta simplificação é justificável pois representa a abordagem adotada pela maioria dos estudiosos e abrange a maior parte das discussões, resultando em uma perda teórica mínima.

A questão que pode surgir é como definir as interpretações que classificaremos como objetivas e as que consideraremos subjetivas. Para simplificar o estudo e alinhar com a maioria dos autores, adotaremos a classificação proposta por Hacking (2001). Nesse sentido, denominaremos como objetiva a probabilidade que se relaciona com a frequência, e como subjetiva aquela que se relaciona com a crença. Vale mencionar que a probabilidade subjetiva é, por vezes, referida como *epistêmica*, aqui com uma extensão ligeiramente diferente da que atribuímos ao termo, e a probabilidade objetiva é, por vezes, chamada de *aleatória*. No entanto, conforme aponta Hacking (2001), esses termos mais técnicos não se popularizaram, pois palavras simples em inglês tendem a ser mais facilmente lembradas do que termos sofisticados em grego ou latim[32][33].

[32] Certamente, estamos nos referindo às perspectivas de indivíduos oriundos de países de língua inglesa, reconhecidos como disseminadores proeminentes do conhecimento em questão.
[33] Rudolf Carnap designou a probabilidade relacionada à crença como probabilidade 1 e a probabilidade associada à frequência como probabilidade 2. Contudo, essa nomenclatura nunca se tornou amplamente aceita, uma vez que pode ser confuso recordar qual conceito corresponde a cada número (Hacking, 2001.

Ainda em relação às abordagens objetivista e subjetivista, que se concentram respectivamente nas noções de frequência e crença, existe um debate entre os teóricos. Alguns, conhecidos como dogmáticos, defendem que apenas uma dessas perspectivas é verdadeiramente relevante. Por outro lado, outros teóricos sustentam que ambas as definições são valiosas e contribuem para o entendimento completo do conceito de probabilidade.

Os dogmáticos frequencistas afirmam que todo raciocínio indutivo deve ser interpretado exclusivamente através da probabilidade do tipo frequência, argumentando que outras abordagens probabilísticas não oferecem suporte adequado ao raciocínio indutivo. Por outro lado, os dogmáticos que favorecem a abordagem baseada em crença sustentam que as probabilidades do tipo frequência são irrelevantes para a ciência. Eles argumentam que todo raciocínio indutivo deve ser compreendido em termos de probabilidades do tipo crença, rejeitando a validade da abordagem frequencista como incoerente (Hacking, 2001).

Contudo, como já mencionado, existem aqueles que sustentam a complementaridade entre essas abordagens, argumentando que tanto a probabilidade frequencista quanto a probabilidade subjetiva, fundamentada em crenças, possuem aplicações distintas que podem aprimorar nossa compreensão e utilização da teoria da probabilidade.

Conforme Hacking (2001) sugere, adotar uma posição eclética pode ser a abordagem mais sensata. A maioria dos dados e argumentos que um dogmático de frequência considera pode igualmente ser analisada por um dogmático de crenças, e o inverso também é verdadeiro. As diferenças entre estas duas escolas de pensamento geralmente emergem apenas em situações específicas, onde podem extrair inferências distintas a partir dos mesmos dados. Na maioria das vezes, as divergências estão mais relacionadas aos métodos empregados para a coleta de dados, planejamento e execução de experimentos, e extração de informações.

Como será abordado com maior profundidade a seguir, existem circunstâncias em que uma das abordagens teóricas se revela inadequada para a resolução de determinados problemas, em função de sua própria concepção. Compreender essa nuance é fundamental para a seleção da teoria mais apropriada ao enfrentamento de questões filosóficas.

2.2 INDUÇÃO E A FILOSOFIA DA CIÊNCIA

Antes de iniciarmos as discussões sobre estatística, é crucial abordar um tema de importância fundamental, com implicações significativas para a subsequente análise estatística: o problema da indução. Esse tema, amplamente estudado pela filosofia da ciência, possui grandes reflexos para diversas áreas, especialmente na estatística. A questão central envolve a capacidade da estatística em realizar inferências, ou seja, formular e testar hipóteses a partir de análises empíricas específicas. Compreender o problema da indução é essencial para fundamentar as metodologias indutivas empregadas na estatística, bem como suas limitações epistemológicas.

Dada a amplitude e a complexidade do tema, este texto não pretende ser exaustivo. Trata-se de uma introdução, destinada a facilitar as discussões que serão aprofundadas no próximo capítulo.

David Hume elaborou a problemática da indução, desafiando a fundamentação lógica das generalizações baseadas em processos indutivos. Segundo Hume, a indução, ao extrair conclusões gerais a partir de observações específicas, carece de um respaldo lógico sólido, uma vez que pressupõe, sem qualquer garantida, que o futuro se comportará de maneira análoga ao passado. Ele sustenta que todos os argumentos derivados da experiência estão alicerçados na similaridade que percebemos entre objetos naturais, a qual nos leva a antecipar a ocorrência de efeitos análogos àqueles que já observamos em relação a tais objetos.

Numerosos filósofos dedicaram esforços consideráveis à resolução do problema da indução, mas muitos chegaram à conclusão de que se trata de uma questão verdadeiramente insolúvel (Henderson, 2022). Alguns estudiosos argumentam que o raciocínio de Hume não necessariamente implica uma posição cética em relação à viabilidade da indução. Há aqueles que sustentam que essa não era a verdadeira intenção de Hume, enquanto outros afirmam que ele formulou o problema de maneira inadequada (Henderson, 2022). Contudo, essa perspectiva não parece ser a mais comum; a maioria dos pensadores considera o problema da indução como um dos desafios filosóficos mais intrincados e profundos, com vastas implicações para a estrutura da ciência, uma vez que questiona uma das formas mais fundamentais de aquisição de conhecimento (Henderson, 2022).

O argumento de Hume tem sido apresentado em várias versões, e há um grande debate sobre a verdadeira interpretação pretendida por Hume com seu argumento, sem que haja consenso sobre o ponto (Henderson, 2022). Para nosso estudo, é suficiente compreender que Hume argumenta que não é possível fazer generalizações lógicas a partir de informações particulares, o que é essencial para a inferência estatística.

Karl Popper propôs o falseacionismo como uma solução alternativa ao problema da indução (Popper, 2017). Popper considerava o problema da indução intransponível, mas entendia que a ciência não dependia de inferências indutivas para a construção do conhecimento (Henderson, 2022). Em vez de buscar a verificação das teorias científicas por meio da indução, Popper argumentou que a ciência deveria avançar por meio da falseabilidade. As teorias científicas deveriam ser formuladas de modo a poderem ser testadas e refutadas.

Embora haja divergência acerca da adequação do falseacionismo à estatística clássica, muitos acreditam que ele pode ser aplicado através da formulação de hipóteses estatísticas. O objetivo seria testar a hipótese e buscar evidências que possam refutá-la, evitando assim o problema de fazer generalizações a partir de dados particulares. A teoria de Popper, que propõe a falseabilidade empírica como critério de demarcação entre ciência e não-ciência, é amplamente reconhecida, embora também tenha numerosos críticos.

Bruno de Finetti, um dos principais defensores do subjetivismo na teoria da probabilidade, propôs, como vimos, a probabilidade como uma medida de crença e a estatística bayesiana como solução para o problema da indução. De acordo com De Finetti (2017), a probabilidade é uma medida subjetiva de crença, que varia conforme a informação e a experiência individual, não exigindo o uso da indução para a construção do conhecimento.

A obra de De Finetti tem implicações significativas para a estatística, permitindo que as inferências sejam realizadas de maneira mais flexível, ajustando-se a novas evidências e a contextos específicos. Isso contrasta com a abordagem frequentista, que tende a ser mais rígida e pode não ser tão eficaz em cenários com dados escassos ou variáveis. Nas seções seguintes, exploraremos esse tema com maior profundidade.

2.3 ESTATÍSTICA CLÁSSICA E A FILOSOFIA DA CIÊNCIA

Já mencionamos que existe uma diferença entre a teoria clássica da probabilidade, cujas origens remontam aos estudos de Pascal e Fermat no século XVII, e a teoria clássica da estatística, frequentemente chamada de frequencista devido à centralidade das frequências dos eventos nos procedimentos clássicos e à relevância da interpretação frequencista desenvolvida por von Mises para os estudos estatísticos tradicionais.

Embora a estatística tenha uma origem mais antiga e esteja historicamente entrelaçada com a história da probabilidade, o que geralmente chamamos de teoria clássica da estatística foi desenvolvido na primeira metade do século XX, com Ronald Aylmer Fisher, Jerzy Neyman e Karl Pearson entre os nomes mais proeminentes.

Voltando um pouco mais no tempo, vemos que a palavra *estatística* provém do latim *status*, que significa estado, devido à sua origem relacionada à compilação de dados e gráficos que descreviam vários aspectos de um estado ou país (Triola, 1998). No século XVII, John Graunt publicou relatórios estatísticos sobre nascimentos e mortes, juntamente com informações sobre o tamanho da população, rendas e taxas de desemprego, que eram usados por governos e empresas para definir estratégias de ação, como a contratação de empregados, níveis de produção e expansão para novos mercados (Triola, 1998).

Graunt utilizou para suas análises as listas de nascimentos e mortes compiladas semanalmente pelos escrivães das paróquias, conhecidas como *Bill of Mortality*, uma prática iniciada em 1592 e padronizada em 1603 (Tabak, 2016). Ele analisou dados de 1604 a 1661 e publicou suas considerações em *Natural and Political Observations*, um documento que se tornou referência para muitas análises posteriores devido ao seu estudo detalhado dos dados (Tabak, 2016). Graunt desenvolveu técnicas para usar os dados na estimativa de riscos, algo que se mostrou de notável importância nos anos seguintes. Suas observações foram tão relevantes que o próprio Rei Carlos II intercedeu para sua admissão na Royal Society, embora isso não tenha efetivamente acontecido (Tabak, 2016).

Inicialmente, o trabalho estatístico estava amplamente associado a essas enormes compilações de grandes volumes de dados. Entretanto, como nos informa Morettin (1981), o estatístico moderno passou a se dedicar a tarefas científicas muito mais complexas, contribuindo para o adequado planejamento de experimentos, correta interpretação de dados e apresentação adequada dos resultados para facilitar a tomada de decisões. Em vez de focar apenas na análise descritiva, os estatísticos passaram a usar matemática avançada para criar modelos que não só descrevem o passado, mas sugerem cenários futuros, aumentando sua conexão teórica com o problema da indução mencionado anteriormente (Morettin, 1981).

O termo *clássico* na estatística provavelmente deriva do papel preponderante que esta abordagem assumiu na primeira metade do século XX, impulsionado pelos trabalhos de seus fundadores: Karl Pearson, Ronald A. Fisher e Jerzy Neyman (Paulino, et al. 2018). De acordo com a metodologia desenvolvida por esses pioneiros, o objetivo principal da estatística é estabelecer generalizações sobre uma população com base em amostras coletadas dessa população (Paulino, et al. 2018). Em um sentido mais amplo, fazer inferências a partir de observações, seja por meio de métodos dedutivos ou indutivos, constitui a essência da ciência empírica (Morettin e Bussab, 2010).

Para compreender a estatística clássica, é fundamental entender que, em vários campos práticos e científicos, existem experimentos e observações que podem ser repetidos inúmeras vezes em condições semelhantes (Triola, 1998). Cada uma dessas ocasiões produz um resultado de observação expresso por certos aspectos característicos. A estatística investiga e desenvolve métodos adequados para avaliar hipóteses a partir dessas observações (Triola, 1998). Assim, um método pode ser chamado de estatístico quando relaciona fatos e hipóteses (Triola, 1998). Os fatos são codificados e estruturados em conjuntos de dados, e as hipóteses são formuladas em termos de distribuições de probabilidade ao longo de conjuntos de dados possíveis. Em um estudo científico, os dados são registros de observações ou eventos específicos. A partir dessa construção, é possível relacionar a situação fática a uma probabilidade e fazer inferências baseadas nessas informações (Triola, 1998).

Os dados obtidos são chamados de amostras e fazem parte de um conjunto maior conhecido como espaço amostral, que, como já discutimos, inclui todas as amostras possíveis de um estudo. As hipóteses são afirmações gerais sobre o sistema objeto do estudo científico, o que as associa ao problema da indução. O estudo estatístico se desenvolve principalmente a partir de hipóteses estatísticas, onde informações gerais estão relacionadas a determinadas distribuições de probabilidade. Isso permite a análise dos dados com base nas probabilidades associadas a essas distribuições.

Devido à centralidade da probabilidade frequentista, os métodos clássicos de estatística baseiam-se em amostragens repetidas, onde os comportamentos dos métodos estatísticos são avaliados após um grande número de repetições sob as mesmas condições (Paulino et al., 2018). Um aspecto fundamental dessa abordagem é justamente a interpretação frequentista da probabilidade, que, como vimos, se fundamenta no uso de frequências para mensurar incerteza (Paulino et al., 2018).

A prática estatística contemporânea não se restringe a um único procedimento clássico; ela é uma combinação dos conceitos e métodos de Neyman-Pearson e Fisher, que são os métodos frequentistas mais conhecidos. Não entraremos em detalhes sobre as diferenças entre esses métodos, pois isso seria pouco relevante para a linha de argumentação deste texto. Para uma compreensão mais detalhada dessas diferenças, recomenda-se a leitura de Royall (1999), onde os capítulos 2 e 3 abordam especificamente as teorias de Neyman-Pearson e Fisher, respectivamente.

Os métodos estatísticos fornecem meios matemáticos para avaliar hipóteses estatísticas a partir das amostras obtidas. A probabilidade desempenha um papel fundamental nesse estudo, permitindo discutir, com base em modelos matemáticos, até que ponto determinadas amostras apoiam nossas hipóteses, proporcionando uma avaliação mais robusta da confiabilidade dessas hipóteses. Por se tratar de uma discussão probabilística, alguns autores não associam a refutação de hipóteses, nesse caso, ao falseacionismo de Popper, onde, em tese, haveria uma refutação definitiva.

Segundo Neyman (1977), a fundamentação dos modelos frequencistas na análise de fenômenos naturais remonta ao trabalho do matemático Émile Borel. Sua obra seminal, publicada em 1909, estabeleceu um marco significativo ao discutir a construção de modelos estocásticos. Borel reconheceu a importância desses modelos não apenas como instrumentos matemáticos, mas também como uma abordagem essencial para a resolução de problemas complexos que surgem no campo da estatística.

Neste momento, já deve ter se tornado evidente a profunda conexão entre essa discussão e a problemática da indução. Como discutido anteriormente, a estatística pode ser concebida como uma forma de inferência generalizada (ou hipótese) derivada de dados, frequentemente oriundos de investigações empíricas. Assim, alguns argumentam que estamos, na verdade, ampliando o domínio de nossa inferência, transitando de um conjunto específico de informações para reflexões mais abrangentes ou preditivas que expandem o alcance dos dados originais.

O problema da indução é um dos mais importantes debates científicos abordados pela filosofia da estatística. Para Howson e Urbach (2006), as hipóteses científicas geralmente possuem um alcance mais amplo do que as observações empíricas que pretendem explicar, implicando fenômenos e eventos que estão além de qualquer evidência concreta existente. Eles usam como exemplo a teoria genética de Mendel, que se aplica a todas as características hereditárias em diversos tipos de flora e fauna, inclusive aquelas que ainda não foram observadas ou que possam vir a evoluir. Essa situação revela uma discrepância lógica entre as informações obtidas através de observações empíricas e o conteúdo das teorias científicas, levantando questões acerca da correção das teorias decorrentes (Howson e Urbach, 2006).

Ao confrontarmos a construção científica padrão com o problema da indução formulado por Hume, que questiona a validade de tirar conclusões gerais a partir de observações específicas devido à ausência de uma garantia lógica de que eventos futuros seguirão o mesmo padrão, nos deparamos com um problema que possui grandes implicações para o estudo da ciência.

Se aprofundarmos na discussão, chegaremos a um ponto intencionalmente ignorado na primeira parte do texto: a questão da convergência de uma série de eventos para um valor que chamamos de probabilidade (frequencista). De acordo com Hacking (2016), não há como inferir com base sólida em dados experimentais qualquer hipótese sobre a frequência a longo prazo. Isso difere de satisfazer os axiomas de Kolmogorov, que a frequência a longo prazo cumpre. A teoria formal baseada nesses axiomas permite provar proposições condicionais, como "se a frequência a longo prazo de algo é isso, então a de outra coisa é aquilo" (Hacking 2016).

Antes de prosseguir, é importante esclarecer algumas questões relevantes. Primeiramente, a discussão anterior não se refere à convergência de séries matemáticas[34]. Não estamos tratando de determinar se uma série específica, com dados iniciais dados, é convergente ou não, um problema comum na matemática que frequentemente possui soluções bastante simples.

O que Hacking (2016) argumenta é a inviabilidade de fazer suposições sobre a probabilidade com base na convergência dos resultados de uma série de eventos além dos observados, partindo de observações factuais. Esta questão está diretamente relacionada ao problema da indução e exige que discutamos a realidade factual em um espectro mais amplo do que o fornecido pelos dados.

A segunda questão relevante emerge da discussão iniciada na seção anterior e será explorada em maior profundidade adiante. É crucial destacar que a convergência em análise está vinculada à probabilidade frequencista, que fundamenta a estatística frequencista, embora

[34] Essa questão dificilmente se apresentaria dessa forma, tendo em vista estarmos falando de aleatoriedade.

não se confunda com ela. A conexão estabelecida com o falseacionismo de Popper refere-se à estatística frequencista e suas hipóteses. Nesse contexto, alguns autores argumentam que não estaríamos, de fato, lidando com um problema de indução. Isso se deve ao fato de que, ao se excluir a discussão probabilística subjacente, não se fazem afirmações gerais acerca das hipóteses, mas sim se discute a possibilidade de refutá-las. A questão, portanto, depende essencialmente das hipóteses em questão e do rigor na sua formulação. Esse raciocínio assemelha-se à ideia de corroboração proposta por Popper, embora possa ser interpretado como de menor intensidade, uma vez que se fundamenta em conceitos probabilísticos.

Certamente. Aqui está a reescrita em um nível mais acadêmico:

A inferência ampliativa e indutiva na estatística é frequentemente objeto de controvérsia, especialmente devido à sua fundamentação na probabilidade frequencista, que impacta as distribuições estatísticas subjacentes. Um exemplo elucidativo é a análise da proposta de Newton, segundo a qual a atração gravitacional entre corpos é proporcional às suas massas e inversamente proporcional ao quadrado da distância que os separa. Apesar da abundância de dados empíricos que apoiam esta tese, não se pode, de maneira justificada, inferir uma regra geral acerca da validade da lei da gravitação universal. A compreensão contemporânea revela que a teoria de Newton, embora prática em diversos contextos, carece da abrangência necessária para descrever fenômenos que envolvem campos gravitacionais intensos ou velocidades próximas à da luz. Nestes casos, a teoria da relatividade geral de Einstein se torna imprescindível, um fato bem estabelecido no conhecimento atual. Karl Popper fundamentou sua concepção de falseabilidade, em parte, na reflexão sobre essas limitações teóricas.

Conforme argumenta Romeijn (2022), os procedimentos clássicos fazem uso de dados para delimitar um conjunto de hipóteses, oferecendo, dessa forma, uma solução ao problema da indução. Segundo afirma, nesse contexto os dados operam indutivamente, promovendo a redução do conjunto inicial de hipóteses a um subconjunto mais restrito. A preocupação da filosofia estatística, nesse ponto, passa a ser a justificativa desse procedimento (Romeijn, 2022).

Este texto não pretende aprofundar-se no problema da indução. No entanto, é relevante mencionar que o problema da indução não parece ter sido realmente resolvido. Da forma como foi proposto por Hume, não há possibilidade lógica de fazer tal inferência. O falseacionismo, por exemplo, aborda a questão de uma maneira diferente, tentando resolver um problema da ciência, mas não a questão da indução. O próprio Popper (2017) foi claro ao afirmar que não há como uma teoria ser definitivamente provada com base em observações empíricas.

Popper (2017) propôs o falseacionismo como uma alternativa, sugerindo que as teorias científicas devem ser formuladas de maneira a poderem ser falseadas. Nesse contexto, não há realmente indução envolvida. No que diz respeito especificamente ao problema da indução, considerando a fundamentação lógica que embasa a argumento seguinte, não é exagero afirmar que qualquer teoria, por mais razoável que se apresente, que declare ter solucionado o problema de Hume conforme ele o formulou[35], estará, de maneira inevitável, equivocada.

2.4 CRÍTICAS À TEORIA FREQUENCISTA DA ESTATÍSTICA

Já se discutiu que a teoria clássica da estatística continua sendo a mais utilizada no campo da ciência empírica e que somente mais recentemente se começou a reconhecer a qualidade dos resultados obtidos com teorias de viés subjetivo. É indiscutível que o desenvolvimento das ciências se beneficiou significativamente do aperfeiçoamento das teorias estatísticas fundamentadas no frequencismo. Além disso, é amplamente reconhecido que

[35] Ver ressalvas sobre a formulação de Popper em comentário anterior.

muitas teorias científicas experimentais foram construídas com base na estatística frequentista. Trata-se de uma disciplina robusta, com aplicações em diversas áreas do conhecimento.

Apesar da grande importância da estatística clássica para as ciências, o tema não está isento de críticas. Esta seção do texto tem como objetivo apresentar algumas dessas limitações, de modo a aprofundar a compreensão da relação entre a estatística e a ciência.

2.4.1 Impraticabilidade de Sequências Infinitas

A questão das séries infinitas é uma das críticas mais significativas e amplamente discutidas tanto no campo da matemática quanto no da filosofia. Esta crítica aborda a questão essencial de como as probabilidades são definidas e aplicadas na prática, considerando a definição teórica de probabilidade como o limite de uma sequência infinita de eventos, conforme discutido anteriormente.

Como discutido na primeira parte deste texto, para von Mises, a probabilidade de um evento é definida como o limite da frequência relativa desse evento em uma sequência infinita de ensaios repetidos sob condições idênticas. Formalmente, se n é o número total de ensaios e n_E é o número de vezes que o evento E ocorre, a probabilidade P(E) é dada por:

$$P(E) = \lim_{n\to\infty}\left(\frac{n_E}{n}\right)$$ [36]

A definição mencionada pressupõe que é possível considerar uma sequência suficientemente extensa de eventos, grande o bastante para que as frequências relativas se estabilizem e se aproximem de um valor constante, correspondente à probabilidade do evento.

[36] $\lim_{x\to\infty} f(x) = L$ significa que para cada número real $\varepsilon > 0$, dado arbitrariamente, podemos encontrar $\delta > 0$ de modo que se tenha $|f(x) - L| < \varepsilon$ (Lima, 1995), em outras palavras, é sempre possível aproximar a função do valor L, bastando que se escolha um x suficientemente grande. Assim, o que a fórmula atribuída a von Mises nos diz é que para n suficientemente grande, podemos nos aproximar do resultado P(E) tanto quanto quisermos.

Na prática, contudo, é impossível realizar um número infinito de ensaios, o que é inquestionável. Mesmo ao relativizarmos a questão do infinito para um número muito grande de ensaios, surgem várias questões, como a limitação de recursos, a estabilidade das frequências relativas e a ocorrência de eventos raros.

A impraticabilidade de sequências infinitas tem importantes implicações epistemológicas. Se a definição teórica de probabilidade como limite de uma sequência infinita não pode ser aplicada na prática, em tese, não seria possível justificar o uso de probabilidades em inferências científicas, o que levantaria questionamentos acerca da base epistemológica da probabilidade frequencista.

A abordagem frequencista depende da inferência indutiva, onde generalizamos a partir de um número finito de observações para uma população mais ampla. Como já discutido, Hume argumentou que não há garantia lógica de que as regularidades observadas no passado se manterão no futuro, o que põe em dúvida a validade da indução. A dependência de sequências infinitas exacerba essa preocupação, pois a inferência é baseada em uma suposição teórica não realizável na prática.

Diversas soluções foram propostas para lidar com a inviabilidade prática das sequências infinitas. Alguns defensores da abordagem frequentista argumentam que os resultados permanecem confiáveis se os limites assintóticos puderem ser considerados aproximações úteis.

Métodos como a estimação pontual e a construção de intervalos de confiança são empregados para inferir probabilidades a partir de conjuntos de dados finitos. Esses métodos reconhecem explicitamente a variabilidade amostral, oferecendo uma margem de erro nas estimativas probabilísticas. Contudo, conforme argumentam Howson e Urbach (2006), um problema significativo desses métodos reside na forte dependência de pressupostos sobre a distribuição dos dados, os quais nem sempre podem ser estabelecidos com precisão.

A solução mais proposta para o problema apresentado, em grande parte devido à dicotomia teórica já referida, é a aplicação da abordagem bayesiana. A estatística bayesiana oferece uma alternativa que não depende de sequências infinitas. Nesta abordagem, as probabilidades são continuamente atualizadas com base em novas evidências por meio do teorema de Bayes. Isso permite a incorporação de conhecimento prévio (*prior*) e o ajuste das probabilidades conforme novas evidências são adquiridas. Savage (1972) argumenta que essa abordagem proporciona uma estrutura mais coerente para a tomada de decisões sob incerteza.

Entender a probabilidade e, consequentemente, a estatística como baseadas em um modelo que depende de séries infinitas pode ser problemático, especialmente em contextos empíricos onde é necessário que essas séries sejam ancoradas em dados observados ou em ampliações indutivas do que foi observado. Nesses casos, a possibilidade de utilizar evidências como um elemento adicional para ajustar a probabilidade surge como uma alternativa promissora.

Entretanto, essa questão é mais epistemológica do que metodológica. Do ponto de vista da estruturação do conhecimento, ter uma teoria que se baseia em uma probabilidade inicial, que pode ser atualizada com novas evidências, remove a fragilidade teórica presente na estruturação da ciência empírica segundo os argumentos probabilísticos de von Mises. Contudo, do ponto de vista prático, não há garantias de que o uso limitado de evidências seja suficiente para que os resultados obtidos com a estatística bayesiana sejam mais precisos. Essa questão abrange uma variedade de fatores, especialmente no que diz respeito à seleção da probabilidade prévia (prior), que representa o “calcanhar de Aquiles” da abordagem bayesiana, conforme será explorado posteriormente.

2.4.2 Inaplicabilidade a eventos únicos ou raros

Uma segunda crítica à abordagem da teoria clássica da estatística é resultado de sua limitação para aplicação a eventos únicos ou raros. Eventos únicos são aqueles que, por sua própria natureza, não podem ser submetidos a repetição ou replicação experimental. Exemplos típicos incluem eventos históricos, desastres naturais específicos e certas decisões únicas em processos empresariais ou políticos. A abordagem frequencista, que define a probabilidade como a frequência relativa de um evento em uma sequência infinita de experimentos, não pode ser diretamente aplicada a tais eventos, pois não há uma sequência infinita ou mesmo uma grande amostra de eventos similares para basear essa frequência.

De maneira análoga à crítica previamente mencionada, essa questão possui um impacto extremamente significativo. Na verdade, parece ser o maior obstáculo para a aplicação da estatística frequentista em diversos problemas práticos, especialmente em vários problemas que emergem no campo da filosofia.

McGrayne (2011) nos fornece numerosos exemplos de situações em que não seria possível a aplicação da teoria frequencista, tendo em vista se tratar de evento único ou raro. Apenas para citar alguns, temos a quebra do código Enigma[37] (p. 61-65), busca de submarinos russos[38] (p. 98-103), análise de risco nuclear[39] (p. 176-181) e busca de objetos perdidos[40] (p. 148-153).

---

[37] Durante a Segunda Guerra Mundial, como já mencionado, os aliados usaram o teorema de Bayes para ajudar a decifrar o código Enigma usado pelos alemães. Este foi um evento único, pois a quebra do código teve implicações diretas e imediatas para o desenrolar da guerra. A abordagem bayesiana permitiu aos criptoanalistas incorporarem informações previamente conhecidas e atualizar suas hipóteses conforme novas interceptações de mensagens eram decifradas (McGrayne, 2011).

[38] Segundo McGrayne (2011), durante a Guerra Fria, os Estados Unidos empregaram métodos bayesianos para localizar submarinos russos no oceano. Este é um exemplo de evento único onde a aplicação de técnicas frequencistas seria limitada devido à ausência de uma sequência de eventos repetidos. A abordagem bayesiana permitiu a integração de várias fontes de informação e a atualização contínua das probabilidades de localização dos submarinos com base em novas observações e dados de inteligência (McGrayne, 2011).

[39] A aplicação de métodos bayesianos na avaliação de risco nuclear é outro exemplo significativo de um evento único. Conforme discutiremos na última seção do texto, após o acidente de Three Mile Island, a análise bayesiana foi utilizada para avaliar a probabilidade de falhas adicionais em reatores nucleares (McGrayne, 2011). Este tipo de análise é essencialmente único, pois cada acidente nuclear possui características e contextos distintos que dificultam a aplicação de métodos frequencistas tradicionais.

[40] A busca pelo submarino USS Scorpion, que desapareceu em 1968, exemplifica o uso da abordagem bayesiana para eventos únicos. A Marinha

Em sua obra, McGrayne argumenta que, ao serem indagados sobre essa problemática, os frequencistas afirmavam que a probabilidade é aplicável exclusivamente a contextos que envolvem um elevado número de tentativas. Eles sustentavam que não era possível calcular a probabilidade de um evento singular, uma vez que este não se inseria em um processo que pudesse ser replicado.

Von Mises (1981), um grande expoente da teoria frequentista da probabilidade, afirmava que nada se podia dizer sobre a probabilidade de morte de uma pessoa específica, mesmo com informações detalhadas sobre sua saúde e hábitos de vida. Segundo ele, a expressão *probabilidade de morte* não teria qualquer significado quando aplicada a um indivíduo em particular (Von Mises, 1981). Essa é uma das mais importantes consequências de sua definição de probabilidade.

Karl Popper (2017) discutiu a importância dos eventos únicos em sua obra. Ele argumentou que a ciência frequentemente lida com hipóteses sobre eventos singulares e que a falseabilidade dessas hipóteses é crucial para o progresso científico. No entanto, a abordagem estatística clássica, devido à sua dependência na frequência, falha em fornecer um mecanismo adequado para lidar com eventos únicos. Motivado por essa problemática, Popper desenvolveu sua própria teoria probabilística, que será discutida em mais detalhes posteriormente neste texto.

Não parece haver uma solução viável dentro da estatística frequentista para resolver esse problema. A resposta mais comum é a aplicação da teoria bayesiana, seja isoladamente ou em conjunto com a teoria frequentista. Acredito que a estatística bayesiana seja realmente a melhor abordagem para solucionar o problema, e essa é, inclusive, a razão pela qual o bayesianismo parece ser a melhor escolha para aplicação na filosofia, dada a singularidade de

---

dos EUA empregou técnicas bayesianas para combinar várias pistas e dados incompletos na tentativa de localizar o submarino desaparecido (McGrayne, 2011). A capacidade de atualizar probabilidades com base em novas informações foi crucial para a eficácia da busca.

muitos de seus problemas. No entanto, como se verá, o bayesianismo também apresenta suas dificuldades, que serão discutidas posteriormente.

2.4.1 Distribuições e Convergência

A abordagem frequentista da estatística normalmente emprega distribuições específicas, como a distribuição normal, para realizar testes de hipóteses e construir intervalos de confiança. Essa prática se fundamenta em suposições sobre a forma da distribuição dos dados, que são essenciais para a correta execução das inferências estatísticas. No entanto, essa dependência pode ser problemática quando os dados reais não seguem a distribuição assumida, levando a inferências incorretas e, por conseguinte, a erros nas conclusões científicas.

Métodos frequencistas clássicos, como o teste t de Student[41], a análise de variância (ANOVA) e os intervalos de confiança baseados na distribuição normal, dependem da suposição de que os dados sejam normalmente distribuídos ou que possam ser aproximados por uma distribuição normal através de alguma função ou fundamentação teórica. Essas suposições são fundamentais para a obtenção de propriedades estatísticas desejáveis, como a consistência e a eficiência dos estimadores.

Howson e Urbach (2006) argumentam que, embora muitos testes assumam que os dados são normalmente distribuídos, na prática os dados raramente seguem perfeitamente uma distribuição normal. Quando os dados não aderem à distribuição teórica assumida, os testes frequencistas podem gerar resultados enganosos. Por exemplo, se os dados apresentam assimetria ou curtose excessiva, os testes que pressupõem normalidade podem subestimar ou superestimar a variabilidade dos dados.

[41] Especialmente para amostras pequenas.

Além disso, a confiança nos testes que dependem de distribuições teóricas pode levar a inferências incorretas sobre a significância estatística e os intervalos de confiança, resultando em uma alta taxa de falsos positivos ou falsos negativos, o que afeta a replicabilidade dos resultados (Howson & Urbach, 2006). Resultados baseados em suposições específicas de distribuição também podem não ser generalizáveis a outros contextos ou populações (Howson & Urbach, 2006). Os autores sugerem a abordagem bayesiana como solução, por não depender de suposições estritas sobre a forma da distribuição de dados.

É correto argumentar que, se os dados não podem ser modelados por uma determinada distribuição, seu uso pode resultar em distorções das informações e, consequentemente, em erros. Todavia, essa situação, que os autores fazem entender ser comum em algumas áreas de aplicação da estatística, parece resultar mais do desconhecimento das técnicas do que de uma falha inerente a elas. As técnicas estatísticas deixam claros os pressupostos necessários para a viabilidade dos testes. Utilizar dados que não apresentam normalidade, quando essa é uma premissa do teste, configura um erro de aplicação, não uma falha do teste em si. Embora se argumente que, em muitos casos, não se terá normalidade, isso não é necessariamente uma falha do teste, mas sim uma limitação de sua abrangência. Os pressupostos foram adequadamente estabelecidos e devem ser respeitados para que os testes sejam aplicados corretamente.

Howson e Urbach fazem diversas críticas aos métodos estatísticos clássicos, chegando a afirmar que essas teorias não são exemplos de sucesso, embora ainda mantenham influência entre filósofos e cientistas. Eles também sugerem que alguns testes, como o qui-quadrado, não deveriam mais ser utilizados (Gillies, 1997). Nesse contexto, Gillies (1997) argumenta que é pouco provável que testes tão amplamente utilizados e com tanto êxito deixem de ser empregados pela comunidade científica, e parece haver mérito em sua posição.

A estatística frequentista, embora amplamente utilizada, enfrenta críticas significativas, também em relação à convergência dos limites das frequências relativas e sua aplicabilidade a todas as possíveis sequências e repetições de um experimento. Ross (2010) destaca que essa suposição de convergência pode ser problemática, particularmente quando se considera a variabilidade intrínseca das sequências experimentais. A dificuldade reside em garantir que, independentemente da sequência específica de observações, a proporção de ocorrências de um evento se estabilize em um valor constante.

Suponha que realizamos um experimento repetido, como o lançamento de uma moeda. A questão central é determinar se a proporção de resultados *cara* obtidos nas primeiras n jogadas convergirá para um valor específico à medida que n tende ao infinito. Além disso, mesmo que haja convergência, surge a dúvida se, em uma segunda realização do experimento, a proporção limite de *cara* será a mesma. Ross (2010) ilustra essa preocupação ao questionar a constância da proporção limite de eventos em experimentos repetidos, sugerindo que a variabilidade das sequências pode comprometer a confiabilidade das inferências baseadas em frequências relativas.

Os defensores da teoria frequentista argumentam que a convergência para um valor limite constante é uma suposição fundamental ou axioma do sistema estatístico frequentista. No entanto, Ross (2010) critica essa posição, afirmando que tal suposição é excessivamente simplista e não é garantida a priori. Ele sugere que, embora seja esperado que a frequência limite exista, essa expectativa não é uma certeza evidente, exigindo uma fundamentação mais robusta para ser aceita.

De acordo com Ross (2010), uma abordagem mais robusta seria adotar um conjunto de axiomas simples e intuitivamente evidentes para a probabilidade e, a partir desses axiomas, demonstrar que a frequência limite constante existe sob determinadas condições. Essa metodologia reflete a abordagem axiomática moderna da teoria da probabilidade, que busca fundamentar as inferências probabilísticas em princípios bem definidos e logicamente consistentes. Ross (2010) argumenta que essa abordagem oferece uma base mais rigorosa e confiável para a estatística frequentista, minimizando as incertezas associadas às suposições de convergência das frequências relativas.

2.5 POPPER E A PROBABILIDADE

Karl Popper, como já mencionamos, eminentemente reconhecido como um dos filósofos mais influentes do século XX no campo da filosofia da ciência, também se destacou como filósofo social e político. Ele era um racionalista convicto e crítico ferrenho do ceticismo e do relativismo, tanto em questões científicas quanto em assuntos humanos em geral, defendendo fervorosamente o conceito de *Sociedade Aberta* (Thornton, 2023).

Importante destacar que a estatística clássica, discutida anteriormente, é frequentemente vista como associada ao falseacionismo de Popper. No entanto, trata-se, segundo Romeijn (2022), de uma associação enganosa, tendo em vista que nos procedimentos clássicos as hipóteses são descartadas quanto tornam a amostra muito improvável. Isso difere do procedimento adotado pelo falseacionismo, de descartar hipóteses que consideram a amostra observada impossível.

Embora a questão pareça ser realmente como apresentada por Romeijn (2022), um aprofundamento do tema mostra que descartar hipóteses que consideram a amostra observada impossível não é tão simples como possa parecer. Esse texto não tem a intenção de aprofundar sobre o tema, mas vale a referência ao próprio Popper (2017), que afirma que nunca se poderá efetivamente produzir a refutação concludente de uma teoria, uma vez que sempre será possível discutir a correção dos resultados experimentais ou da discrepância entre os resultados experimentais e a teoria. Para uma maior compreensão do tema, sugere-se uma leitura acerca do que passou a ser conhecido como falseacionismo ingênuo (Lakatos, 1979), em contraste com versões mais sofisticadas da metodologia científica popperiana.

A abordagem de Popper (2017) baseava-se no princípio da demarcação, segundo o qual a falseabilidade constitui o critério principal para uma teoria ser considerada científica, um princípio que influencia diretamente seus estudos em probabilidade. Popper argumentava que a indução não é um processo racional, uma vez que apenas a inferência dedutiva é correta e, portanto, essencial para a ciência, dada a capacidade dos dados empíricos de refutar teorias (Lipton, 1997). Para Popper, a indução seria irracional, mas a investigação científica não. Centrando seus argumentos na ideia de que as inferências científicas seriam exclusivamente dedutivas, Popper afirmava ser impossível mostrar a verdade de uma hipótese, havendo apenas a possibilidade de mostrar sua falsidade, com a exposição de um caso negativo[42] (Lipton, 1997).

Essa postura era considerada excessivamente cética em relação às possibilidades do conhecimento científico (Lipton, 1997). Lipton (1997) observa que Hume também era cético em relação à inferência indutiva, mas igualmente cético quanto à investigação científica. Popper, por sua vez, tentava conciliar o ceticismo de Hume sobre a indução com a afirmação de que a investigação científica é racional.

Essa visão de Popper abriu possibilidade para vários questionamentos. Um exemplo resulta do fato de os cientistas partirem de observações para a aceitação da verdade de certas proposições, mas essa não é uma forma válida de inferência segundo Popper[43] (Lipton, 1997). Porém, conforme aponta Lipton (1997), os críticos de Popper argumentam que essa posição inviabiliza o uso do mecanismo de refutação empírica, resultando no problema dos dados

[42] Segundo Popper (2017), nenhuma quantidade de corvos negros seria suficiente para provar a hipótese de que todos os corvos são negros, mas um único corvo branco seria suficiente para demonstrar sua falsidade.
[43] De acordo com Popper, apenas enunciados podem justificar outros enunciados. Isso implica que não é viável afirmar que uma hipótese é falsa com base em um enunciado observacional, pois este último poderia ser o falso (Lipton, 1997).

injustificáveis. Se não é possível saber se os dados estão corretos, também não se pode afirmar que as teorias são falsas. Na visão de Lipton (1997), não há um caminho seguro para o falseamento que não envolva a indução, pois os métodos indutivos são essenciais para gerar resultados positivos e, por extensão, para garantir a confiabilidade dos métodos negativos.

Em sua obra *A Lógica da Pesquisa Científica*, Popper dedicou considerável atenção à discussão sobre a probabilidade, refletindo seu profundo interesse pelo tema. Como destacado por David Miller a Gillies (1997), mais da metade dessa obra é dedicada à probabilidade.

Popper era um defensor do objetivismo e crítico do subjetivismo e do bayesianismo, promovendo uma interpretação própria da probabilidade baseada na teoria frequencista de von Mises. Ele criticava veementemente as concepções não frequencistas de probabilidade e via a teoria subjetiva como *sem esperança*, objetivando analisar e explicar probabilidades objetivas (Hammerton, 1968).

Naquele momento histórico, e nesse ponto específico, as ideias de Popper iam ao encontro das difundidas pelo Círculo de Viena, cujos membros em sua maioria seguiam a já explicada linha de pensamento científico da época, fortemente frequencista, mas Popper chegou posteriormente à conclusão de que havia se equivocado e desenvolveu a que ficou conhecida como teoria propensional (Gillies, 1997).

Segundo Gillies (2000) a distinção fundamental entre as concepções de propensão do Popper em suas fases inicial e posterior reside na abordagem da repetibilidade. Na fase inicial, Popper vincula as propensões a condições que são replicáveis. Por outro lado, em sua fase posterior, Popper propõe que as propensões físicas são propriedades intrínsecas da situação física e, por isso, sujeitas a alterações (Gillies, 2000).

A motivação de Karl Popper para desenvolver sua teoria da probabilidade emergiu de uma questão específica, nomeadamente o desafio de atribuir probabilidades a eventos singulares, um conceito que ele denominou como *probabilidades singulares* (Gillies, 1997). Essa noção enfrentou oposição de figuras como von Mises, que questionava a possibilidade de se estabelecer probabilidades singulares objetivas. Contudo, Popper persistiu na defesa dessa possibilidade, motivado em grande parte pela necessidade de aplicá-las em seus estudos sobre mecânica quântica, como reporta Gillies (1997).

O desenvolvimento dessa teoria por Popper levou ao que é conhecido como o problema da condicionalidade. Segundo Gillies (1997), esse problema surge da necessidade de considerar um conjunto de condições prévias que influenciam o valor probabilístico de um evento específico. Isso requer uma análise das condições que definem o evento antes de sua ocorrência para a determinação da probabilidade. Tal abordagem implica que a forma como um evento é percebido ou definido pode alterar significativamente a probabilidade resultante. Por exemplo, a probabilidade de uma pessoa chegar aos 80 anos pode variar se ela for considerada uma pessoa comum, um brasileiro, ou um brasileiro de boa condição financeira, esportista e com estilo de vida moderado.

Este aspecto da análise levou autores como Howson e Urbach a argumentarem que essas probabilidades singulares são, na verdade, subjetivas. Gillies (1997) concorda com esse argumento e pondera que realmente não parece viável estabelecer probabilidades singulares objetivas de maneira convincente, alinhando-se, assim, às críticas levantadas por Howson e Urbach nesse contexto.

Procederemos agora a uma análise mais detalhada da estatística bayesiana, um domínio onde a abordagem subjetiva encontra um ambiente propício para seu emprego de maneira mais robusta e abrangente. Pode-se afirmar que, embora seja possível identificar elementos de subjetividade na teoria clássica e traços de objetividade na teoria bayesiana, essas teorias foram historicamente e estruturalmente desenvolvidas para uma melhor adaptação às abordagens objetiva e subjetiva, respectivamente.

2.6 ESTATÍSTICA BAYESIANA E A FILOSOFIA DA CIÊNCIA

Já antecipamos que existe outra perspectiva para abordar as probabilidades nos métodos estatísticos, derivada da chamada teoria epistêmica. Neste contexto, referiremos a essa perspectiva como probabilidade subjetiva, embora já tenhamos explorado as nuances dessa definição anteriormente. As figuras proeminentes na probabilidade subjetiva são o matemático italiano Bruno de Finetti e o americano L. J. Savage, que defendiam que apenas a concepção de probabilidade como crença pessoal deveria ser entendida como válida (Hacking, 2001). De Finetti desenvolveu sua teoria simultaneamente a Ramsey, que reconhecia um espaço para um conceito de probabilidade como frequência, especialmente em contextos como a mecânica quântica (Hacking, 2001).

A probabilidade subjetiva é crucial para o estudo da inferência bayesiana, que emprega o teorema de Bayes para atualizar a probabilidade de uma hipótese à medida que novas evidências surgem, como discutido na primeira parte deste texto. Não revisitaremos a discussão matemática já apresentada.

Apesar de ser possível discutir a abordagem bayesiana a partir de probabilidades frequencistas, isso limitaria substancialmente o alcance do bayesianismo e introduziria problemas inerentes ao frequencismo nesta nova abordagem. A posição eclética tem suas vantagens, porém, a imposição da probabilidade frequencista na estatística bayesiana em busca de objetividade não parece ser a resposta adequada. Exploraremos este tema com mais detalhes nesta seção.

Conforme já discutido, Romeijn (2022) apresenta uma perspectiva alternativa sobre as probabilidades em métodos estatísticos, oriunda da teoria epistêmica, que seria dividida em probabilidades doxásticas, lógicas e de teoria da decisão.

As probabilidades doxásticas, por exemplo, seriam interpretadas como representações das opiniões de um agente racional idealizado sobre dados e hipóteses. Nesse contexto, a probabilidade pode quantificar o grau de crença na ocorrência de um evento, como a possibilidade de chuva no dia seguinte.

O método mais comum nas ciências empíricas continua sendo uma combinação dos métodos de Neyman-Pearson e de Fisher, geralmente denominados frequencistas, especialmente em discussões que os distinguem de um conjunto radicalmente diferente de conceitos e métodos conhecidos como bayesianos (Royall, 1999). A estatística bayesiana apresenta diferenças fundamentais em relação à estatística frequencista, e os pontos fortes e fracos de ambas as escolas têm sido objeto de intenso debate. Atualmente, a divisão mais estudada dentro da estatística não é entre as escolas Neyman-Pearson e Fisher, mas entre frequencistas e bayesianos (Royall, 1999). A estatística bayesiana baseia-se na atualização de crenças a priori com base em evidências subsequentes.

Conforme assinala Titelbaum (2022), uma epistemologia bayesiana pode ser compreendida como qualquer teoria que adota os seguintes princípios: 1) Os agentes possuem atitudes doxásticas que podem ser representadas por números associados a afirmações; 2) Os requisitos racionais referentes a essas atitudes doxásticas podem ser representados por restrições matemáticas nas atribuições de números reais, intimamente relacionadas ao cálculo de probabilidades. Note que, na visão de Titelbaum, a restrição racional está limitada ao cálculo de probabilidade, o que o inclui entre os doxásticos subjetivos na classificação de Romeijn (2022), por não exigir elementos adicionais de racionalidade.

Devido à forma como a estatística bayesiana é estruturada, ela parece modelar melhor certos problemas filosóficos, especialmente aqueles que envolvem eventos únicos ou raros que desafiam a abordagem frequencista. Muitos problemas filosóficos que exigem raciocínio probabilístico ocorrem em contextos onde a repetição de um grande número de eventos idênticos não é viável. Isso torna a metodologia de ajuste de crenças iniciais com base em novas evidências especialmente útil para lidar com questões filosóficas complexas.

Para ilustrar a relevância crescente do tema, faz-se referência ao bestseller de Kahneman (2011) que destaca o teorema de Bayes como um elemento fundamental para disciplinar o pensamento. Segundo o autor, pensar de acordo com as regras do teorema de Bayes ajuda a evitar erros de pensamento típicos do raciocínio rápido. A obra de Kahneman é especialmente relevante para a discussão deste texto, considerando que os vieses cognitivos têm uma relação direta com nossas crenças e, consequentemente, com um impacto significativo na probabilidade subjetiva. Também de grande relevância na discussão relacionada a vieses cognitivos e como elemento de normativo de racionalidade, está o probabilismo, tema que tem como precursores grandes nomes como Ramsey e De Finetti (Neiva, 2022).

O conceito de bayesianismo subjetivo é baseado na ideia de mínima restrição para o estabelecimento da probabilidade prévia, limitando-se apenas à coerência. Esta coerência é entendida por alguns como a adequação aos axiomas da probabilidade, embora haja variações nessa posição, conforme já discutido. Portanto, a controvérsia entre os adeptos do bayesianismo subjetivo não se concentra na probabilidade prévia em si, mas nos critérios necessários para garantir essa coerência (Lin, 2023).

A principal preocupação com o bayesianismo subjetivo é a percepção de que, nesta abordagem, quase tudo seria possível sem restrições significativas. Embora os teóricos não afirmem explicitamente que tudo é permitido, alguns críticos argumentam que essa liberdade excessiva pode negligenciar aspectos cruciais da objetividade científica (Lin, 2023). Para os proponentes dessa teoria, a coerência por si só é vista como suficiente para assegurar a objetividade científica necessária.

Há um argumento de que a ampla liberdade na escolha da probabilidade prévia, influenciada pela variabilidade das opiniões pessoais, poderia levar a escolhas divergentes. A menos que haja uma mudança uniforme de opinião, não seria possível afirmar que as escolhas foram objetivamente corretas (Lin, 2023). A objetividade, neste contexto, seria vista como um resultado da convergência das opiniões à medida que evidências compartilhadas se acumulam. Lin (2023) menciona que essa noção de convergência de opiniões como uma forma de objetividade científica tem suas raízes em Charles Sanders Peirce, embora ele tenha desenvolvido essa ideia no contexto da epistemologia de crenças binárias.

O problema que emerge é que a esperada convergência de opiniões pode não ocorrer, embora se acredite que isso aconteça a longo prazo. Surge então a questão: o que significa esse *longo prazo*? Pode ser necessário esperar um tempo excessivo para que tal convergência se realize.

Outra preocupação levantada por Lin (2023) decorre do fato de a prova do teorema da fusão de opiniões apresentada pelos seus defensores decorre da aditividade contável[44] como norma de credibilidade, o que é controverso[45].

[44] Para quaisquer proposições $A_1, A_2, \ldots A_n, \ldots$, mutuamente exclusivas, se alguém tiver crença nessas proposições e na sua disjunção $\bigcup_{n=1}^{\infty} An$, então a função credibilidade satisfaz a seguinte fórmula: $Cr(\bigcup_{n=1}^{\infty} An) = \sum_{n=1}^{\infty} Cr(\text{An})$ Lin (2023) .

[45] Contraexemplo de De Finetti (loteria infinita): Existe uma loteria justa com uma infinidade contável de bilhetes. Por ser justo, existe um e apenas um bilhete vencedor e todos os bilhetes têm a mesma probabilidade de ganhar. Para um agente que considera tudo isso garantido (ou seja, com total credibilidade), qual deveria ser sua credibilidade na proposição An que o enésimo bilhete vai ganhar? A resposta parece ser zero, mas esse resultado viola a aditividade contável: $Cr(\bigcup_{n=1}^{\infty} An) = 1, mas\ \sum_{n=1}^{\infty} Cr(\text{An}) = 0$ Lin (2023) .

Um aspecto importante que emerge é o conceito de coerência. Conforme Lin (2023) observa, para um bayesiano convencional, a coerência implica, no mínimo, aderência ao probabilismo. No entanto, o mesmo autor nota que há várias interpretações de probabilismo, sobre as quais os bayesianos têm opiniões divergentes quanto à mais adequada. Lin (2023) também aponta que existe um debate entre os bayesianos sobre se a coerência deveria englobar mais do que meramente o probabilismo. Essas divergências têm um impacto significativo nas aplicações práticas da epistemologia bayesiana.

2.6.1 Críticas ao enfoque bayesiano

O último capítulo deste texto será dedicado às críticas dirigidas ao paradigma bayesiano. Será conduzida uma análise das vantagens e desvantagens da abordagem subjetiva à teoria bayesiana. Propõe-se uma correlação entre a principal dificuldade de aplicar o bayesianismo no campo da filosofia e os desafios similares encontrados em sua implementação na matemática e estatística, questões estas que foram discutidas na primeira parte desta dissertação. Por meio desta abordagem, busca-se estabelecer um paralelo com os desenvolvimentos ocorridos nesse campo, facilitando assim uma melhor compreensão de como certas críticas podem ser efetivamente superadas.

2.6.1.1 Subjetividade na escolha da Probabilidade Prévia (*Priors)*

Analogamente à estatística frequencista, a abordagem bayesiana enfrenta várias críticas, sendo a questão da subjetividade da probabilidade prévia uma das mais persistentes e significativas desde o início de seu uso, como vimos. Os principais estudos realizados, tanto no campo da matemática e estatística quanto no campo da filosofia, identificam essa subjetividade como um problema central para inferência bayesiana. Além disso, diversas outras críticas surgem direta ou indiretamente desse problema. Por essa razão, este texto se concentrará nessa crítica específica e em algumas soluções apresentadas para mitigar esse problema, sem a pretensão de ser exaustivo. Essa escolha foi feita com o objetivo de proporcionar uma compreensão clara de um problema central, em vez de apresentar noções fragmentadas de vários problemas secundários.

A seleção da probabilidade prévia subjetiva está diretamente relacionada à interpretação da probabilidade subjetiva. Essa perspectiva difere da abordagem frequencista, resultando em uma clara perda inicial de objetividade. No entanto, oferece vantagens significativas em termos de flexibilidade e amplitude de aplicação da teoria. Embora alguns autores proponham restrições substanciais na escolha da probabilidade prévia para aumentar a objetividade, tais restrições podem limitar severamente a aplicabilidade do método.

A principal crítica é que a subjetividade pode introduzir viés nas inferências, particularmente em contextos em que os dados observacionais são escassos. Nestas circunstâncias, a probabilidade prévia pode predominar sobre a probabilidade posterior, influenciando significativamente os resultados e potencialmente desviando-se da objetividade científica frequentemente associada ao método frequencista.

Para resolver o problema da probabilidade prévia, uma abordagem adotada é a sua objetivação. Entre as técnicas aplicadas, destaca-se a desenvolvida por Edwin Jaynes na década de 1950. Conhecida como o princípio da máxima entropia, essa abordagem sugere uma maneira objetiva de selecionar a distribuição de probabilidade quando apenas informações limitadas estão disponíveis (Jaynes, 2003).

Autores como Berger (1985) argumentam que a escolha da probabilidade prévia pode ser extremamente influente, conduzindo a conclusões diversas conforme as crenças prévias do analista. Essa situação suscita preocupações acerca da reprodutibilidade e objetividade dos resultados bayesianos. Ademais, a discussão filosófica sobre a natureza subjetiva da probabilidade prévia continua a ser um tema de significativa controvérsia na comunidade estatística.

2.6.1.1.1 Princípio da Máxima entropia

O princípio da máxima entropia foi formulado por Edwin T. Jaynes em 1957. Inicialmente surgido no contexto da mecânica estatística, esse princípio evoluiu para se tornar uma metodologia amplamente aplicada em diversos campos científicos (Guiasu & Shenitzer, 1985). Essencialmente, o princípio da máxima entropia é utilizado para construir distribuições de probabilidade que abrangem todas as possibilidades viáveis, sujeitas a restrições conhecidas, tipicamente representadas pelos valores médios de certas variáveis aleatórias (Guiasu & Shenitzer, 1985).

De maneira mais específica, a teoria propõe o desenvolvimento de distribuições de probabilidade prévia, fundamentadas no conhecimento existente, visando uma abordagem mais objetiva e evitando suposições adicionais não justificadas (Jaynes, 2003). Conforme Guiasu e Shenitzer (1985), a formulação de entropia por Jaynes foi baseada na entropia discreta de C.E. Shannon e tem uma relação direta com a função H de Boltzmann, utilizada na mecânica estatística.

Segundo Guiasu e Shenitzer (1985), antes do experimento probabilístico, a entropia quantifica a incerteza dos resultados possíveis, o que contribui para a decisão acerca da distribuição prévia. Após o experimento, a entropia mensura as informações obtidas, auxiliando no ajuste posterior. É importante destacar que Shannon foi pioneiro no uso de funções matemáticas para mensurar incerteza de um experimento probabilístico, distanciando-se da medidas tradicionais da época (Guiasu & Shenitzer, 1985).

Interessante notar que o princípio da máxima entropia indica que a distribuição uniforme, onde todos os resultados são igualmente prováveis, maximiza a entropia quando a incerteza é máxima, isto é, quando não há informação adicional sobre qualquer um dos resultados[46]. Este entendimento está alinhado com o princípio da razão insuficiente de Laplace, discutido na primeira parte do texto, que propõe a igualdade das probabilidades como a abordagem mais adequada quando não temos motivos para diferenciá-las (Guiasu & Shenitzer, 1985).

Para Laplace, essa perspectiva era subjetiva, baseada em prudência e senso comum, sendo frequentemente aplicada na vida diária mesmo antes de se compreender o conceito de entropia. Dessa forma, a aceitação da entropia de Shannon como uma medida de incerteza reforça matematicamente o princípio da razão insuficiente, ao afirmar que a entropia é maximizada pela distribuição uniforme na ausência de restrições específicas (Guiasu & Shenitzer, 1985).

Em cenários onde restrições são aplicadas às distribuições de probabilidade, Jaynes propõe o princípio da máxima entropia como um critério de escolha natural, determinando que, dentre todas as distribuições de probabilidade que se ajustem aos valores médios conhecidos de uma ou mais variáveis aleatórias, deve-se optar pela que maximiza a entropia de Shannon (Guiasu & Shenitzer, 1985).

---

[46] Esse resultado pode ser obtido a partir da função entropia $H(P) = -\int P(x)\, log\, P(x)\, dx$ com a condição, dada a ausência de informações adicionais, de que $\int P(x) d(x) = 1$ (note que estamos diante de uma função densidade de probabilidade). A partir daí, basta maximizar H(P) com a restrição dada usando o método dos multiplicadores de Lagrange. Para o caso discreto ver a fórmula apresentada por Guiasu e Shenitzer (1985, p. 42).

2.6.1.1.2 Probabilidades prévias empíricas

Assim como no exemplo anterior, a ideia por trás das chamadas probabilidade prévias empíricas é reduzir a subjetividade na escolha dessas probabilidades prévias, baseando as distribuições diretamente nos dados observados. Diferentemente das probabilidades prévias subjetivas, que são definidas com base em crenças, as probabilidades prévias empiricas utilizam dados históricos ou amostras de dados para definir seus valores. Essa abordagem combina os métodos bayesianos com os métodos frequencistas.

Herbert Robbins, na década de 1950, introduziu a expressão *empirical Bayes* para descrever o método de estimar distribuições a priori a partir de dados observados. Robbins demonstrou que a utilização de dados agregados de amostras permite obter estimativas de probabilidades prévias objetivamente informadas pelos dados.

Os métodos *Empirical Bayes* têm uma longa história. Suas raízes podem ser rastreadas até os trabalhos de von Mises na década de 1940, mas o primeiro trabalho significativo deve ser atribuído a Herbert Robbins, embora sua formulação fosse um pouco diferente da utilizada atualmente (Casella, 1985).

A formulação de Robbins pode ser referida como Empirical Bayes não-paramétrico, enquanto a formulação atual é conhecida como Empirical Bayes paramétrico (Casella, 1985). A principal diferença é que a abordagem paramétrica especifica uma família paramétrica de distribuições prévias, enquanto a abordagem não-paramétrica deixa a distribuição prévia completamente não especificada (Casella, 1985). As probabilidades prévias empíricas são obtidas ajustando uma distribuição aos dados disponíveis antes de realizar a análise bayesiana. Isso reduz a subjetividade, pois as probabilidades prévias passa a refletir características observadas nos dados em vez de crenças subjetivas.

Os métodos *Empirical Bayes* têm se tornado cada vez mais populares e têm sido aplicados a muitos tipos de problemas. Alguns exemplos incluem probabilidades de alarmes de incêndio, compartilhamento de receita, garantia de qualidade e admissões em escolas de direito (Casella, 1985). Essa situação exemplifica o ponto discutido na primeira parte do texto, afirmando que o bayesianismo não é necessariamente sinônimo de subjetividade. Sempre que é possível obter a probabilidade prévia a partir dos dados, o usual é proceder dessa forma. No entanto, exigir que isso seja sempre feito reduz consideravelmente o escopo de aplicação da teoria. A grande vantagem do bayesianismo é seu amplo espectro de utilização, justamente devido à possibilidade de utilizar probabilidades prévias subjetivas.

Adiantando uma discussão da conclusão deste texto, esse é o motivo pelo qual bayesianismo se mostra mais adequando aos problemas filosóficos. Não acompanho a posição de que a estatística bayesiana seja superior à frequencista ou vice-versa; a questão, conforme defendido por Hacking (2001), é mais sobre o ambiente de aplicação. Uma abordagem pode ser mais útil para certas aplicações e a outra para outras.

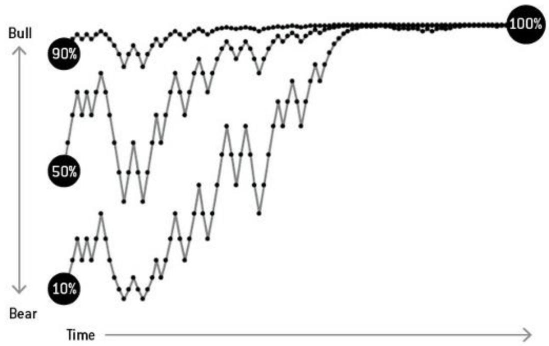
De qualquer forma, tratando-se de problemas filosóficos, que geralmente não são repetitivos, a estatística frequencista tem pouca aplicabilidade. Restringir a discussão das probabilidades prévias de maneira mais rígida certamente tem um grande potencial de prejudicar essa aplicação.

2.6.1.1.3 Convergência da probabilidade posterior

Em síntese, a proposta é que, mesmo iniciando com probabilidades subjetivas diversas, o processo de atualização de crenças baseado em novas evidências eventualmente levará à convergência das probabilidades posteriores ao longo do tempo. Esta ideia representa uma resposta chave às críticas dirigidas ao subjetivismo na inferência bayesiana. Um dos méritos desse argumento reside no fato de que ele evita a introdução de elementos que restrinjam excessivamente a seleção da probabilidade prévia. Embora as diversas abordagens para tornar as probabilidades prévias mais objetivas frequentemente se alinhem bem com a inferência estatística, essas mesmas abordagens complicam significativamente a aplicação da estatística no âmbito filosófico. De fato, já foi destacado que um dos principais problemas da estatística frequencista na filosofia é a sua inadequação para lidar com casos únicos ou raros, que são comuns na filosofia. Muitas das técnicas desenvolvidas para tornar as probabilidades prévias mais objetivas resultam em métodos estatísticos que restringem sua aplicação a eventos singulares, limitando assim o alcance da estatística bayesiana.

Quanto à convergência das probabilidades posteriores, é importante ressaltar que essa abordagem não modifica a metodologia de seleção das probabilidades prévias, mas apenas sublinha que, à medida que novas evidências são integradas, a probabilidade posterior se ajusta progressivamente ao resultado mais preciso. Neste contexto, Nathaniel Read Silver, conhecido como Nate Silver, ilustra essa dinâmica com o gráfico a seguir, referente ao mercado de ações. Nos Estados Unidos, um mercado onde há tendencia de alta para os ativos é conhecido como mercado touro (bull market). No caso de ciclo de baixa o mercado é conhecido como mercado urso (bear market). O gráfico mostra como a crença inicial de três investidores convergem ao final.

FIGURE 8-8: BAYESIAN CONVERGENCE



Nate Silver é um estatístico americano de renome, fundador do blog político Five ThirtyEight, associado ao New York Times (Albeck-Ripka, 2023), e foi nomeado uma das 100 pessoas mais influentes do mundo pela revista Time em 2009 (Time, 2009). Em seu livro, que se destina a um público não especializado em estatística, Silver (2012) argumenta que uma das propriedades do teorema de Bayes é que nossas crenças convergiriam umas com as outras e com a verdade à medida que recebemos novas evidências. Ele menciona que a ideia de consenso pode ser problemática, mas sugere que as opiniões da comunidade tendem a se alinhar com a verdade conforme as ideias são debatidas e novos indícios surgem.

Silver ganhou fama mundial por sua habilidade em prever resultados eleitorais, acertando a vitória de Barack Obama, assim como os resultados para o Senado e governos estaduais entre 2008 e 2012, como discutido em seu livro *O Sinal e o Ruído*. Um desafio mencionado é que Silver conta com um vasto volume de dados (*big data*), o que facilita a convergência para resultados mais precisos. Além disso, ele não parte de suposições completamente hipotéticas, pois utiliza informações de empresas e estatísticos que analisaram os dados antes dele, tornando sua abordagem menos subjetiva.

Segundo Silver (2012), suas análises alcançam um equilíbrio ao considerar uma variedade de fatores objetivos, incluindo seis sistemas de classificação ajustados para uma escala comum e cujas médias são calculadas (Silver, 2012). Isso inclui quatro conjuntos de classificações de poder e dois que envolvem julgamento humano. Não iremos nos aprofundar em sua pesquisa, os detalhes podem ser consultados na obra de Silver (2011).

Quando questionado sobre as divergências nas opiniões antecedentes, Savage respondeu que, à medida que as evidências aumentam, os subjetivistas tendem a convergir em suas opiniões (McGrayne, 2011). Os cientistas avançam em direção a um consenso com as evidências acumuladas (McGrayne, 2011).

A argumentação sobre a convergência das probabilidades enfrenta limitações, uma vez que tanto a velocidade quanto a extensão dessa convergência dependem da quantidade e qualidade das evidências coletadas. Em situações em que as evidências são escassas ou de baixa qualidade, a convergência pode ser lenta ou mesmo inadequada. Outro ponto relevante é a escolha da probabilidade prévia: valores extremos dificultam a convergência sem evidências de qualidade e em quantidade suficiente. Se dispomos de poucas evidências de boa qualidade, a probabilidade prévia passa a ter grande impacto na probabilidade posterior. Se há muitas evidências de qualidade, o efeito da probabilidade prévia diminui e a verossimilhança das evidências predomina, contribuindo para a convergência.

Finalmente, conforme já discutido, a seleção de probabilidades prévias objetivas pode restringir consideravelmente sua aplicabilidade, especialmente na filosofia, onde surgem frequentemente problemas singulares ou raros. Embora a convergência possa representar uma alternativa, sua eficácia depende da escolha apropriada da probabilidade prévia ou da disponibilidade de evidências suficientes—uma condição que nem sempre é viável em questões filosóficas, ao contrário do que ocorre em contextos de big data na estatística. Uma potencial solução para esse dilema será examinada na seção seguinte, que trata do conhecimento especializado.

2.6.1.1.4 *Conhecimento de especialistas*

Por tudo que se falou até o momento, parece não haver dúvidas de que a probabilidade prévia subjetiva é de grande importância para problemas singulares, o que é comum na filosofia. Entretanto, sua amplitude de aplicação tem o preço de exigir um cuidado maior com sua escolha ou necessidade de disponibilidade de um maior volume de evidências de qualidade.

Um ajuste adequado na probabilidade prévia, que passa a exigir uma variação menor para atingir valores satisfatórios, reduz a necessidade de grande volume de evidências. Essa escolha pode ser feita com ajuda de especialistas. Com efeito, quanto mais pessoas com conhecimento profundo da matéria pudermos acessar, menor a chance de escolhermos probabilidades prévias que destoem muito do valor real.

Para além da filosofia, esse tipo de problema também é comum em questões práticas que se apresentam à estatística. Isso pode ajudar a mostrar como a questão se desenvolve na estatística para exemplificar a importância que o conhecimento de especialistas pode ter para a filosofia, uma vez que estamos tratando do mesmo problema, o bayesianismo como forma de raciocínio que ajude na ampliação do conhecimento.

Assim, mostra-se de relevante importância apresentarmos grandes problemas singulares que foram resolvidos com inferência bayesiana e com auxílio de especialistas para que a solução apresentada fique mais evidente.

Um exemplo clássico apresentado por McGrayne (2011, pp. 176-181) refere-se ao acidente nuclear na usina de Three Mile Island, ocorrido na Pensilvânia no final da década de 1970. Após a inauguração da indústria de energia nuclear pelo presidente Eisenhower em 1953, passaram-se vinte anos sem que fossem realizados estudos abrangentes sobre os riscos à segurança pública ou ao meio ambiente. Mesmo assim, corporações privadas possuíam e operavam cinquenta usinas nucleares nos Estados Unidos. Quando o Congresso começou a debater a possibilidade de absolver os proprietários e operadores das usinas de toda a responsabilidade por acidentes, a Comissão de Energia Atômica dos Estados Unidos finalmente ordenou a realização de um estudo de segurança.

Quando Norman Carl Rasmussen foi nomeado para avaliar a segurança da indústria de energia nuclear, nunca tinha ocorrido um acidente em uma usina nuclear. Sem evidências diretas, a equipe de Rasmussen não teve outra escolha senão solicitar a opinião de especialistas. Analistas anteriores acreditavam que as chances de danos severos ao núcleo eram extremamente baixas e que os efeitos seriam catastróficos. No entanto, o Relatório Rasmussen concluiu o contrário: a probabilidade de danos ao núcleo era maior do que o esperado, mas as consequências nem sempre seriam catastróficas. O relatório também identificou dois problemas significativos: erro humano e liberação de radioatividade fora do prédio, além de identificar a sequência de eventos que estariam relacionados ao acidente.

Após o incidente, o relatório que antes estava destinado ao esquecimento ganhou nova relevância, como se tivesse antecipado os desdobramentos futuros, corroborando suas conclusões e evidenciando a eficácia da abordagem subjetivista na avaliação de riscos. Em sua análise, Rasmussen empregou métodos bayesianos, embora sem mencioná-los explicitamente, em virtude da controvérsia que cercava sua aplicação na época.

A história está repleta de exemplos bem-sucedidos do uso da estatística bayesiana para resolver problemas únicos ou raros, muitos dos quais foram solucionados com a ajuda de especialistas para a definição das probabilidades prévias ou para o fornecimento de evidências. Alguns desses casos são apresentados na obra de McGrayne (2011), que proporciona uma leitura agradável e fluente acerca da história do teorema de Bayes e suas aplicações.

3 CONCLUSÃO

Ao contrário de algumas décadas atrás, atualmente há muitos defensores tanto da abordagem estatística frequencista quanto da bayesiana. Ainda existem frequencistas dogmáticos que acreditam que todo raciocínio indutivo deve ser analisado exclusivamente pela probabilidade do tipo frequência e defensores dogmáticos da abordagem de crença que consideram que a probabilidade do tipo frequência não faz qualquer sentido. No entanto, parece que a melhor posição é a eclética, que valoriza os pontos fortes de ambas as teorias, como pondera Hacking (2001).

A estatística frequencista, apesar de suas limitações, já demonstrou seu valor ao longo do tempo. Ela é responsável por uma enorme quantidade de testes que contribuíram significativamente para o avanço das ciências empíricas. As críticas existentes não foram suficientes para diminuir sua importância e ela continua a ser a teoria mais aplicada no mundo, especialmente nas ciências empíricas. A estatística clássica conquistou seu espaço há muito tempo, graças a cientistas de renome como Fisher, Neyman e Pearson.

Da mesma forma, a estatística bayesiana, também com suas limitações, tem mostrado seu grande valor, especialmente, mais recentemente, com o uso de computação avançada. A possibilidade de utilizar dados com pouca informação e aproximar-se progressivamente do resultado correto com a apresentação de novas evidências torna essa abordagem especialmente importante para problemas complexos com pouca informação prévia.

Portanto, parece que, no contexto em que se exige objetividade e a análise de um grande volume de dados é possível, a estatística frequencista pode ser mais adequada, fornecendo ferramentas robustas para a análise de padrões, verificação de hipóteses e uso de um grande número de modelos matemáticos. Já em situações em que há limitação de conhecimento inicial e a atualização de crenças é determinante, a estatística bayesiana oferece uma abordagem mais rica e flexível, permitindo uma integração mais profunda de informações existentes e uma adaptação contínua a novas evidências.

Existe ainda a possibilidade de usar ambas as abordagens conjuntamente, o que pode contribuir significativamente para a solução de problemas complexos, onde além do volume de dados, a experiência se mostra de grande importância.

O problema epistemológico da probabilidade frequencista, conforme delineado por von Mises, parece não influenciar significativamente as práticas dos cientistas nas ciências empíricas. Mesmo que esse impacto ocorresse, tais questões estão intimamente ligadas ao dilema da indução, o qual, embora insuperável, não deve comprometer a validade da atividade científica, conforme sustentam Howson e Urbach (2006).

Por fim, no contexto da filosofia, a estatística bayesiana parece mais adequada devido à natureza dos problemas na área. Geralmente, não estamos lidando com questões que permitem o uso de um grande volume de dados para cálculo da probabilidade frequencista. Assim, para problemas únicos ou raros, ou onde a escolha subjetiva da probabilidade inicial contribui significativamente para o desenvolvimento do problema, a solução bayesiana geralmente se mostra mais acertada. O estudo dos limites e alcances da abordagem bayesiana em outros campos da Filosofia, como a filosofia da religião, por exemplo, é um desafio para um próxima etapa de estudos.

REFERENCIAS

AHMED, S. E.; REID, N. **Empirical Bayes and Likelihood Inference**. [S.l.]: Springer , 2001.

AITKEN, C. G. G.; TARONI, F. **Statistics and the Evaluation of Evidence for Forensic Scientists**. [S.l.]: John Wiley & Sons, Ltd, 2004.

ALBECK-RIPKA, L. Nate Silver, **FiveThirtyEight Founder, Expects to Depart ABC News Amid Layoffs**. The New York Times, 2023. Disponivel em: <https://www.nytimes.com/2023/04/25/business/media/nate-silver-abc-disney-fivethirtyeight.html?searchResultPosition=5>.

ALENCAR, M. S. D. **Teoria de conjuntos, medida e probabilidade**. São Paulo: Érica, 2014.

AYLMER FISHER, R. **Statistical Methods for Research Workers**. London: Oliver and Boyd, 1950.

BAYES, T. **An essay towards solving a problem in the doctrine of chances**. [S.l.]: Philosophical transactions of the Royal Society of London, 1973. p. p. 370–418.

BERGER, J. O. **Statistical Decision Theory and Bayesian Analysis**. [S.l.]: Springer, 1985.

BERNSTEIN, S. N. **Return to the questions of the accuracy of the Laplace limit formula**, 1943.

BOREL, É. **Elements de la Théorie des Probabilités**. Paris: Hermann 3ed., 1924.

CASELLA, G. **An Introduction to Empirical Bayes Data Analysis**. The American Statistician, Vol. 39, No. 2 Maio 1985. 83-87.

CASELLA, G.; BERGER, R. L. **Inferência Estatística**. [S.l.]: Cengage, 2011.

CHALMERS, A. F. **O que é ciência afinal?** [S.l.]: Editora brasiliense, 1993.

CORFIELD, D.; WILLIAMSON, J. **Foundations of Bayesianism**. [S.l.]: Springer-Science, 2001.

COURANT, R.; ROBBINS, H. **What is Mathematics?** [S.l.]: Oxford University Press, 1996.

CRAMER, H**. Metodos Matematicos de Estadistica**. Madrid: Aguilar, S.A. de Ediciones, 1963.

CRAMER, H. **The Elements of Probability Theory and some of its applications**. [S.l.]: Robert E. Krieger Publishing Company, 1973.

DALE, A. **A History of Inverse Probability - From Thomas Bayes to Karl Pearson.** New York: Springer-Verlag, 1999.

DE FINETTI, B. **Cambridge probability theorists**. Rivista di matematica per le scienze economiche e social, 1985. v.8, n.2, p. 79-91.

DE FINETTI, B. **Theory of Probability - A critical Introductory Treatment** (Translated by Antonio Machí and Adrian Smith). [S.l.]: Jihn Wiley & Sons Ltd, 2017.

DIETRICH, F.; LIST, C. **Reasons for (prior) belief in Bayesian epistemology**. Synthese , 2013. 787-808.

FELLER, W. **An Introduction to Probability Theory and its Applications**. 3ª. ed. [S.l.]: John Wiley & Sons, Inc., v. 1, 1968.

GARDING, L. **Encounter with Mathematics**. [S.l.]: Springer-Verlag, 1977.

GILLIES, D. **A Contribuição de Popper à Filosofia da Probabilidade**. In: (ORG.), A. O. Karl Popper: Filosofia e Problemas. São Paulo: Unesp, 1997. p. 125-145.

GILLIES, D. **Philosophical Theories of Probability**. London and New York: Routledge, 2000.

GNEDENKO, B. V. **The theory of probability**. Moscow: MIR PUBLISHERS, 1978.

GNEDENKO, B.; PAVLOV, I.; USHAKOV, I. **Statistical Reability Engineering**. [S.l.]: John Wiley & Sons, Inc., 1999.

GORROOCHURN, P. **Classic Topics on The History of Modern Mathematical Statistics - From Laplace to More Recent Times**. New York: John Wiley & Sons, Inc., 2016.

GUIASU, S.; SHENITZER, A. **The Principle of Maximum Entropy**. The Mathematical

Intelligencer Vol. 7 N. 1, 1985.

HACKING, I. **An Introduction to Probability and Inductive Logic**. [S.l.]: Cambridge University Press, 2001.

HACKING, I. **Logic of Statistical Inference**. [S.l.]: Cambridge University Press, 2016.

HÁJEK, A. **Interpretations of probability**. [S.l.]: [s.n.], 2002.

HAMMERTON, M. **Bayesian Statistics and Popper's Epistemology**. Mind, New Series, Vol. 77, No. 305, Janeiro 1968. 109-112.

HOWSON, C.; URBACH, P. **Scientific Reasoning: The Bayesian Approach**. [S.l.]: Open Court, 1989.

HOWSON, C.; URBACH, P. **Scientific Reasoning: The Bayesian Approach**. [S.l.]: Open Court Publishing, 2006.

HUME, D. **An enquiry Concerning Human Understanding**. [S.l.]: Charles River Editors, 2018.

JAYNES, E. T. **Information theory and statistical mechanics**. Physical review, 1957. v. 106, n. 4, p. 620.

JAYNES, E. T. **Probability Theory: The Logic of Science**. [S.l.]: Cambridge University Press, 2003.

JEFFREY, R. C. **The logic of decision**. Chicago: University of Chicago Press, 1983.

JEFFREYS, H. **Scientific Inference**. [S.l.]: Cambridge University Press, 1973.

KAHNEMAN, D. **Rápido e Devagar: Duas formas de pensar**. [S.l.]: Objetiva, 2011.

KEYNES, J. M. A **Treatise on Probability**. [S.l.]: Macmillan and Co., Limited, 1921.

KOLMOGOROV, A. N. **Foundations Theory of Probability**. [S.l.]: [s.n.], 1956.

LAKATOS, I. **O falseamento e a metodologia dos programas de pesquisa científica**. In: (ORG.), I. L.; (ORG.), A. M. **A crítica e o desenvolvimento do conhecimento**. São Paulo: Cultrix, 1979. p. 109-243.

LIMA, Elon Lages. **Curso de Análise vol. 1**. São Paulo: Projeto Euclides. IMPA, 1995.

LIN, H. **Bayesian Epistemology**. The Stanford Encyclopedia of Philosophy, 2023. Disponivel em: <https://plato.stanford.edu/archives/win2023/entries/epistemology-bayesian/>.

LIPTON, P. **Popper e o Confiabilismo**. In: (ORG.), A. O. Karl Popper: Filosofia e Problemas. São Paulo: Unesp, 1997. p. 41-55.

LOWELL COOLIDGE, J. **An introduction to mathematical probability**. [S.l.]: Oxford University Press, 1942.

M. STIGLER, S. **The History of Statistics: The Measurement of Uncertainty before 1900**. Cambridge and London: The Belknap Press of Harvard University Press, 1986.

MCGRAYNE, S. B. **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**. [S.l.]: Yale University Press, 2011.

MORETTIN, P. A. **Introdução à Estatística para Ciências Exatas**. [S.l.]: Atual, 1981.

MORETTIN, P. A.; BUSSAB, W. D. O. **Estatística Básica**. [S.l.]: Saraiva, 2010.

MORTARI, C. A. **Introdução à Lógica**. [S.l.]: Unesp, 2016.

NEIVA, A. **Probabilismo e bayesianismo em epistemologia**. Peri v.7, n.2, 2015. 45-69.

NEIVA, A. **Bayesianismo.** In: (ORG.), R. E. D. O., et al. Compêndio de Epistemologia. [S.l.]: Editora fi.org, 2022. p. 47-71.

NEYMAN, J. **Frequentist Probability and Frequentist Statistics**. [S.l.]: Synthese Vol. 36 n° 01, 1977.

PAULINO, C. D. et al. **Estatística Bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 2018.

PETTIGREW, R. **Epistemic Risk and the Demands of Rationality**. [S.l.]: Oxford University Press, 2022.

POPPER, K. **A Lógica da Pesquisa Científica**. São Paulo: Cultrix, 2017.

RAMSEY, F. P. **Truth and Probability**. [S.l.]: Springer International Publishing, 1926.

ROBERTS, J. **How Lorenz was different from Enigma**. The History Press, 2017. Disponivel em: <https://thehistorypress.co.uk/article/how-lorenz-was-different-from-enigma/>.

ROMEIJN, J.W. **Philosophy of Statistics**. The Stanford Encyclopedia of Philosophy, 2022. Disponivel em: <https://plato.stanford.edu/archives/fall2022/entries/statistics/>.

ROSS, S. **A first course in probability**. [S.l.]: Pearson, 2010.

ROYALL, R. **Statistical Evidence: A likelihood paradigm**. [S.l.]: Chapman & Hall/CRC, 1999.

SAVAGE, L. J. **The Foundations of Statistics**. New York: Dover Publications, Inc., 1972.

SCHUPBACH, J. N. **Bayesianism and scientific reasoning**. [S.l.]: Elements in the Philosophy of Science, 2022.

SCHWITZGEBEL, E. **Belief**. The Stanford Encyclopedia of Philosophy, 2024. Disponivel em: <https://plato.stanford.edu/archives/spr2024/entries/belief/>.

SILVER, N. **How We Made Our N.C.A.A. Picks**. The New York Times, 2011. Disponivel em: <https://archive.nytimes.com/fivethirtyeight.blogs.nytimes.com/2011/03/14/how-we-made-our-n-c-a-a-picks/>.

SILVER, N. **FiveThirtyEight Picks the N.C.A.A. Bracket**. The New York Times, 2012. Disponivel em: <https://archive.nytimes.com/fivethirtyeight.blogs.nytimes.com/2012/03/13/fivethirtyeight-picks-the-n-c-a-a-bracket/>.

SILVER, N. **The signal and the noise: the art and science of prediction**. [S.l.]: Penguin UK, 2012.

SKYRMS, B. **Choice and chance**. Belmont, CA: Wadsworth/Thompson, 2000.

TABAK, J. **Probability and Statistics - The Science of Uncertainty**. New York: Facts On File, Inc, 2016.

THORNTON, S. **The Stanford Encyclopedia of Philosophy**. https:

//plato.stanford.edu/archives/win2023/entries/popper/, 2023.

TIME. **The 2009 Time 100**. Time, 2009. Disponivel em: <https://content.time.com/time/specials/packages/completelist/0,29569,1894410,00.html>.

TITELBAUM, M. G. **Quitting certainties: A Bayesian framework modeling degrees of belief**. [S.l.]: Oxford University Press, 2013.

TITELBAUM, M. G. **Bayesian Epistemology**. Oxford: Oxford University Press, 2022.

TRIOLA, M. F. **Introdução à Estatística**. Rio de Janeiro: LTC - Livros Técnicos e Científicos Editora S.A., 1998.

VERBURGT, L. M. **Remarks on the idealist and empiricist interpretation of frequentism: Robert Leslie Ellis versus John Venn**. Journal of the British Society for the History of Mathematics, 29(3), 184-195., 29(3) 2014. 184-195.

VON MISES, R. **Probability, statistics, and truth**. [S.l.]: Curier Corporation, 1981.