UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS

MAÍRA BITTENCOURT

PESO DA EVIDÊNCIA: UMA CONTRIBUIÇÃO DOS CRIPTOANALISTAS ALAN TURING E JACK GOOD PARA A FILOSOFIA DA CIÊNCIA.

CAMPINAS
2021

MAÍRA BITTENCOURT

PESO DA EVIDÊNCIA: UMA CONTRIBUIÇÃO DOS CRIPTOANALISTAS ALAN TURING E JACK GOOD PARA A FILOSOFIA DA CIÊNCIA.

Dissertação apresentada ao Instituto de Filosofia e Ciências Humanas da Universidade Estadual de Campinas como parte dos requisitos para obtenção do título de Mestra em Filosofia.

Orientadora: PROFA. DRA. JULIANA BUENO-SOLER

ESTE TRABALHO CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO DEFENDIDA PELA ALUNA MAÍRA BITTENCOURT, E ORIENTADA PELA PROFA. DRA. JULIANA BUENO-SOLER.

CAMPINAS
2021

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Filosofia e Ciências Humanas
Cecília Maria Jorge Nicolau - CRB 8/3387

B548p

Bittencourt, Maíra, 1980-
Peso da evidência : uma contribuição dos criptoanalistas Alan Turing e Jack Good para a filosofia da ciência / Maíra Bittencourt. – Campinas, SP : [s.n.], 2021.

Orientador: Juliana Bueno-Soler.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Filosofia e Ciências Humanas.

1. Turing, Alan Mathison, 1912-1954. 2. Good, Irving John, 1916-2009. 3. Ciência - Filosofia. 4. Probabilidades. 5. Inferência bayesiana. 6. Evidência. I. Bueno, Juliana, 1976-. II. Universidade Estadual de Campinas. Instituto de Filosofia e Ciências Humanas. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Weight of evidence : a contribution from the cryptanalysts Alan Turing and Jack Good to the philosophy of science
Palavras-chave em inglês:
Science - Philosophy
Probabilities
Bayesian inference
Evidence
Área de concentração: Filosofia
Titulação: Mestra em Filosofia
Banca examinadora:
Juliana Bueno-Soler [Orientador]
Julio Michael Stern
Itala Maria Loffredo D'Ottaviano
Data de defesa: 25-02-2021
Programa de Pós-Graduação: Filosofia

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-9843-5402
- Currículo Lattes do autor: http://lattes.cnpq.br/8783719470096094

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS

A Comissão Julgadora dos trabalhos de Defesa de Dissertação de Mestrado, composta pelos Professores Doutores a seguir descritos, em sessão pública realizada em 25 de fevereiro de 2021, considerou a candidata Maíra Bittencourt aprovada.

Profa. Dra. Juliana Bueno (FT – Unicamp)

Prof. Dr. Julio Michael Stern (Instituto de Matemática e Estatística – USP)

Profa. Dra. Itala Maria Loffredo D'Ottaviano (Instituto de Filosofia e Ciências Humanas – Unicamp)

A Ata de Defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertações/Teses e na Secretaria do Programa de Pós-Graduação em Filosofia do Instituto de Filosofia e Ciências Humanas.

A meus pais

AGRADECIMENTOS

Gostaria de agradecer em primeiro lugar à minha orientadora, Profa. Dra. Juliana Bueno-Soler, pelas muitas observações e críticas relevantes que levaram a esta versão final, e por ler pacientemente as versões iniciais do texto da Dissertação.

Agradeço ao Prof. Dr. Walter A. Carnielli, por ter orientado, com a mesma diligência, uma parte da Dissertação nos primeiros anos da pesquisa; agradeço também pelas aulas ministradas por ele entre 2018 e 2019, período no qual pude definir o tema da pesquisa.

Aos membros da Banca de Defesa e Qualificação: Profa. Dra. Itala M. Loffredo D'Ottaviano, Prof. Dr. Evandro Luís Gomes e Prof. Dr. Julio M. Stern, que, além de me apresentarem novas perspectivas com as suas leituras da Dissertação, apontaram questões importantes a serem corrigidas e melhoradas.

Aos pesquisadores Dr. Fabio Bertato e ao Prof. Abílio Rodrigues, membros suplentes da banca de defesa, pelo aceite do convite e leitura do trabalho.

Agradeço à Dra. Aureci de Fátima da Costa, pela revisão da versão final da Dissertação.

Toda pesquisa necessita de uma grande estrutura para ser realizada e não posso deixar de mencionar muitas outras pessoas e instituições da Unicamp que colaboraram de alguma forma.

A todos os membros do Centro de Lógica, Epistemologia e História da Ciência (CLE): ao professor Prof. Dr. Marcelo Coniglio, que foi meu professor; aos colegas que participaram dos Seminários de Programa e *Colloquia Logicae*; a Abner de Mattos Brito, pela ajuda na disciplina de teoria de conjuntos; aos funcionários da Biblioteca Michel Debrun e aos funcionários do CLE, especialmente Rovilson Pereira e Augusto Fabiano Torres. Aos funcionários da Biblioteca Octavio Ianni (IFCH) e da Biblioteca do IMECC.

Ao Programa de Pós-Graduação em Filosofia da Unicamp, especialmente ao coordenador Prof. Dr. Márcio Augusto Damin Custódio e à secretária Daniela Paula Grigolletto, por toda a ajuda prestada.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil (CNPq), pelo suporte financeiro de 20 meses desta pesquisa (processo número 134078/2017-9).

Aos meus pais, que sempre incentivaram meus estudos, e me apoiaram em todos os momentos.

Às amizades que não posso deixar de citar pelo apoio nestes últimos quatro anos: Newton Bryan, Sandra Fantin, Juan Hinestroza, Andrea Laplane, Eliane Monteiro, Ana Moori, Luci Lira e Silva, Walter Valdevino, Sandra Varjabedian e Giselle Vianna. E a muitas outras amizades que vou agradecer pessoalmente. A todos os colegas de treino e professores da Gracie Barra Novo Cambuí, por todos os aprendizados, principalmente os de perseverança, disciplina, respeito e superação.

“Hume’s criticism of the customary forms of induction was correct. But still the basic idea of common sense thinking is vindicated: induction, if properly reformulated, can be shown to be valid by rational criteria.”

Rudolf Carnap

“But objectivity is precisely what a subjective theory of probability is for: its function is to introduce extra rationality, and therefore objectivity, into your degrees of belief.”

“Everyone is to some extent a Bayesian especially when using common sense."

Jack Good

“Bayesian statistics is not a branch of statistics. It is a way of looking at the whole of statistics.”

Dennis Lindley

RESUMO

A questão central desta Dissertação consiste na análise do conceito de *peso da evidência* e a contribuição desse conceito para a filosofia da ciência, particularmente para a relação entre hipótese e evidência. Esta Pesquisa iniciou-se com a constatação da existência de um dualismo na filosofia da ciência em relação à probabilidade. Existem duas visões antagônicas para o tipo de suporte probabilístico: Popper e Miller (1983) afirmam ter provado que o suporte é dedutivo, outros afirmam que o suporte é indutivo, a exemplo de Good (1985). O conceito de *peso da evidência* é apresentado por Good (1987) como alternativa à prova de impossibilidade da indução probabilística de Popper e Miller (1983), e também, como uma medida de grau de corroboração, uma medida utilizada para fazer inferências estatísticas. Vamos analisar esse conceito contextualizando-o no bayesianismo e em uma aplicação prática: na quebra do código do Enigma na Segunda Guerra Mundial.

Palavras-chave: Alan Mathison Turing, Irving John Good, filosofia da ciência, probabilidades, inferência bayesiana, evidência.

ABSTRACT

The central issue of this dissertation is the analysis of the concept of weight of evidence and the contribution of this concept to the philosophy of science, particularly for the relationship between hypothesis and evidence. This research started with the verification of the existence of a dualism in the philosophy of science in relation to probability. There are two antagonistic views for the type of probabilistic support: Popper and Miller (1983) claim to have proved that the support is deductive, others claim that the support is inductive, like Good (1985). The concept of weight of evidence is presented by Good (1987) as an alternative to the proof of impossibility of the probabilistic induction of Popper and Miller (1983), and also as a measure of degree of corroboration, a measure used to make statistical inferences. We will analyze this concept in the context of Bayesianism and in a practical application: in breaking the Enigma code in World War II.

Keywords: Alan Mathison Turing, Irving John Good, philosophy of science, probabilities, Bayesian inference, evidence.

SUMÁRIO

1. INTRODUÇÃO 13

2. EVIDÊNCIA, INDUÇÃO E DEDUÇÃO, PROBABILIDADE. 17

2.1 Introdução 17

2.2 Lógica indutiva na argumentação: relação evidencial 19

2.3 Lógica indutiva pura 22

2.4 Evidência e verdade no contexto da lógica 24

3. PROBABILIDADE CONDICIONAL, TEOREMA DE BAYES, PROBABILIDADE SUBJETIVA E ESTATÍSTICA BAYESIANA 26

3.1 Introdução 26

3.2 Algumas considerações gerais sobre o bayesianismo 26

3.3 Probabilidade condicional e teorema de Bayes 33

3.4 Estatística bayesiana e probabilidade subjetiva 41

3.4.1 Exemplo comparativo: inferência frequentista e inferência bayesiana 46

4. PESO DA EVIDÊNCIA E SUPORTE PROBABILÍSTICO 50

4.1 Introdução 50

4.2 Peso da evidência: suporte probabilístico indutivo de acordo com Good 54

4.2.1 Exemplo de cálculo de peso da evidência 61

4.2.2 Variações da fórmula do peso da evidência 63

4.2.3 Peso da evidência e quantidade de informação 64

4.3 A relação entre indução, corroboração e probabilidade de acordo com Good 66

4.4 Visão antagônica de Popper e Miller: o suporte probabilístico dedutivo 69

4.5 Outras visões sobre medidas de corroboração 72

5. A QUEBRA DA CRIPTOGRAFIA DO ENIGMA 77

5.1 Introdução 77

5.2 Algumas características básicas da máquina Enigma 81

5.3 Descrição dos métodos desenvolvidos por Turing de acordo com Good 83

6. CONCLUSÃO 89

REFERÊNCIAS 92

APÊNDICE 99

A. Axiomas da probabilidade: probabilidade absoluta e probabilidade relativa 99

B. Transformação algébrica: Razão bayesiana de verossimilhança em Fator Bayes-Jeffreys-Turing 104

1. INTRODUÇÃO

Esta Dissertação foi pautada, inicialmente, pela pesquisa em fundamentos da probabilidade, tanto do ponto de vista axiomático quanto filosófico. Um texto em particular nos chamou a atenção: "*A Proof of the Impossibility of Inductive Probability*", de Popper e Miller (1983). Este curto texto, mais precisamente uma carta, foi publicado pela revista *Nature*, ocupando apenas três páginas. Apesar de ser um texto sucinto, e além disso não apresentar uma matemática avançada, não podemos presumir que o assunto seja simples, que não apresente complexidade e que sua contribuição seja pequena. Pelo contrário: já nas primeiras linhas, observamos que o assunto não é trivial. Os autores afirmam que:

> Provas da impossibilidade da indução veem tornando-se 'natimorta[s] do prelo'[1] desde que a primeira delas (no Tratado da Natureza Humana de David Hume) apareceu em 1739. Um de nós (K.P.) vem produzindo-as por mais de 50 anos. A presente prova pareceu bela a nós dois. (POPPER; MILLER, 1983, p. 687, tradução nossa).

Ou seja, as provas da impossibilidade da indução, de acordo com os autores, iniciaram-se com Hume. Popper produziu-as por 50 anos até a publicação acima citada, e essa discussão apresentada em 1983 por Popper e Miller continuou a ocupar os próprios autores e outros filósofos, por muito anos. Nossa posição a esse respeito, como será demonstrado ao longo desta Dissertação, é que esse debate ainda está em aberto e que a prova de Popper e Miller (1983) não está totalmente consolidada. Em contraposição à visão desses autores, vamos apresentar a proposta de Good (1987) para o suporte probabilístico.[2] Good (1985) afirma logo no título de um dos seus artigos que a "Indução probabilística é inevitável" e continua a afirmar o mesmo em diversas publicações, por exemplo, em Good (1990).

Temas que permeiam a discussão de Popper e Miller (1983) serão abordados tanto para contextualizar como para expandir o problema filosófico em debate. Consideramos importante, em primeiro lugar, apontar qual é a diferença entre processos indutivos e processos dedutivos. Portanto, no capítulo 2, apresentamos uma breve contextualização histórica sobre o tema da indução, a partir de Hacking (2006). Em sequência, elaboramos uma seção sobre a

[1] Para a tradução dessa expressão utilizada por Hume, nos beneficiamos da tradução de Marques, encontrada em Hume (2004, p. 11).
[2] Good apresenta seus argumentos mais detalhadamente no artigo de 1987, porém também vamos analisar outras publicações a esse respeito.

lógica indutiva no campo da argumentação, com base em Skyrms (1966). Embora não seja diretamente aplicável ao nosso trabalho, também citamos outra vertente da lógica indutiva que segue a tradição de Carnap: a lógica indutiva pura (matemática), vertente desenvolvida por Paris e Vencovská (2015). Por fim, indicamos uma forma de tratar a evidência no campo da lógica, um trabalho recentemente desenvolvido por Rodrigues, Carnielli e Bueno-Soler (2020).

A visão de Good sobre o suporte probabilístico que será apresentada em contraposição à de Popper e Miller é uma visão bayesiana da relação entre evidência e hipótese. Esta visão engloba, além de conceitos de fundamentos da probabilidade, alguns conceitos de estatística e de filosofia da ciência (no caso desta última, destaca-se particularmente o conceito de corroboração). A partir deste momento nos direcionamos, paulatinamente, à filosofia da estatística. Elaboramos então o capítulo 3 que discorre sobre o bayesianismo: após uma breve introdução, apresentamos o conceito de probabilidade condicional, o teorema de Bayes, a probabilidade subjetiva e alguns conceitos básicos sobre a estatística bayesiana.

No capítulo 4 apresentamos, finalmente, o conceito de *peso da evidência*. Este conceito é fundamental na argumentação de Good (1987) contra a proposta de Popper e Miller (1983) para a relação entre evidência e hipótese. De acordo com Good (1983, p. xi), "O conceito de *peso da evidência* captura completamente [o conceito] do grau no qual a evidência corrobora a hipótese". Portanto, a discussão que se inicia com o debate entre as visões de suporte probabilístico (dedutivo ou indutivo) se expande para outras questões, como o grau de corroboração. Neste capítulo, apresentamos um exemplo de cálculo de *peso da evidência*, as possíveis variações da fórmula original, sua relação com a quantidade de informação, algumas visões antagônicas derivadas da filosofia da ciência de Popper e outras fórmulas bayesianas existentes na literatura.

Historicamente, o conceito de *peso da evidência* encontra-se pela primeira vez registrado na literatura no final do século XIX, em Pierce (1878, 1956)[3], e em sequência, foi desenvolvido por Jeffreys (1936) e Turing[4]. Good teve contato com esse conceito por intermédio de Turing em Bletchley Park. Lá, ele trabalhou como o principal assistente estatístico de Turing. O primeiro livro de Good, intitulado *Probability and the Weighing of Evidence*, foi publicado em 1950 – após esta data, o autor ainda publicou pelo menos 40 artigos sobre o assunto.

[3] Good faz ressalvas a esse respeito, dizendo que o conceito de Pierce tinha erros e era diferente do elaborado posteriormente por Jeffreys e Turing.
[4] Ver Good (1979, 1985a, 1993, 2000).

No capítulo 5, analisamos mais especificamente o conceito de *peso da evidência* de maneira aplicada. Vamos tratar sobre o que Good (2000) chama de antecipação do método "Bayes empírico" desenvolvido por Turing e a relação do conceito de *peso da evidência* com a quebra da criptografia do Enigma na Segunda Guerra Mundial. Também veremos como uma expressão de Shannon (1948) para informação mútua é um caso especial da expectativa do *peso da evidência*.

Tradicionalmente o conceito de *peso da evidência* não foi muito trabalhado pelos filósofos da ciência. Um exemplo é o livro *Probability and Evidence*, de Horwich (1982). Good (1984c) faz um *review* do livro e o avalia positivamente, porém, observa que o autor sequer menciona o *peso da evidência*, além de não citar nomes importantes para o bayesianismo como Jeffreys e Laplace. Outra publicação que não menciona o *peso da evidência* é o livro *Scientific reasoning: the Bayesian approach*, de Howson e Urbach (1989). Os autores escrevem uma seção dedicada justamente a discutir as respostas bayesianas à proposta de Popper e Miller (1983) que nega o suporte probabilístico indutivo. Mais uma vez, o *peso da evidência* de Good não é mencionado. Esperamos, portanto, que esta Dissertação colabore para ampliar a discussão filosófica sobre o tema *peso da evidência*.

Seguimos os preceitos gerais da epistemologia probabilística, apresentados em Galavotti (2014, p. 77): "A epistemologia probabilística afirma que a probabilidade é um ingrediente essencial da ciência e do conhecimento humano de maneira ampla, e que a indução é um constituinte necessário do método científico."[5] Além disso, nossa escolha de apresentar o conceito de *peso da evidência*, utilizado por Turing em Bletchley Park, como uma proposta mais adequada da relação entre evidência e hipótese é compatível com a seguinte visão a respeito do conhecimento científico:

> A convicção de que a probabilidade faz parte de todos os estágios de uma análise extensiva do conhecimento científico vai de encontro com a abordagem construtivista, de acordo com a qual tal análise deve começar a partir de uma metodologia experimental e mover-se para cima até a formulação de hipóteses, modelos e teorias. (GALAVOTTI, 2014, p. 78, tradução nossa)

Defendemos, portanto, que conceitos probabilísticos de aplicação prática, desde axiomas e teoremas da probabilidade, interpretações da probabilidade, passando por conceitos intermediários até os conceitos estatísticos mais avançados, devem, no mínimo, servir como

[5] A autora cita três exemplos de seguidores dessa linha da epistemologia probabilística: Suppes, Jeffrey e Skyrms.

ponto de partida para uma reflexão filosófica e, no caso desta Pesquisa, mais especificamente, para tentar esclarecer a relação entre evidência e hipótese, além do conceito de indução, neste contexto.

2. EVIDÊNCIA, INDUÇÃO E DEDUÇÃO, PROBABILIDADE.

2.1 Introdução

Em termos gerais, de acordo com Hacking (2006, p. 31), "alguns filósofos modernos" definem a probabilidade como a "relação entre evidência e hipótese". Se seguirmos essa linha de raciocínio, a probabilidade só pode ser definida após o conceito de evidência também ser definido. Isso ocorre no século XVII, e o primeiro registro que o autor cita é uma publicação de 1661, feita por Glanvill.[6] A evidência à qual Hacking se refere é a *evidência indutiva*, e esse conceito também é necessário para fundamentar o problema da indução[7]. Ainda de acordo com o autor, a evidência indutiva é ambígua porque pode assumir dois significados: a de uma "generalização ou (...) lei da natureza, adquirida de uma observação particular ou experimento" e a "indução de particular para particular. Hume, aliás, considera principalmente a última"[8].

Após o surgimento do conceito de evidência indutiva, podemos citar em sequência uma outra contribuição importante para a lógica indutiva. Segundo Hacking, "Leibniz antecipa o programa filosófico de J. M. Keynes, Harold Jeffreys e Rudolf Carnap"[9]. Leibniz afirmava que a probabilidade poderia ser um "novo tipo de lógica" (LEIBNIZ apud HACKING, 2006, p. 134). O programa filosófico de Leibniz ao qual Hacking faz referência era denominado *Characteristica universalis*, e tinha seis princípios. Aqui vamos parafrasear os princípios enumerados por Hacking:

1) Existe evidência não-dedutiva. Há razões para crer em *p* que não implicam materialmente *p*.

2) "Ser uma boa razão para" é uma relação entre proposições.

3) A relação entre as proposições pode ser formalizada numa relação entre sentenças.

[6] Republicado em 1970, ver referências.
[7] Ibid., p. 32.
[8] Ibid., p. 37.
[9] Ibid., p. 134.

4) Existe uma ordem de razões (*r*), de boas para ruins. E existe uma medida para o grau em que *r* é uma razão para *p*.

5) Essa medida do grau em que *r* é uma razão para *p* é autônoma (é uma medida objetiva, independentemente de opiniões pessoais).

6) Essa medida é global, válida para quaisquer pares de proposições (*r*, *p*).

Como podemos observar nesse conjunto de princípios, Leibniz trata de uma relação entre proposições ou sentenças que pode ser medida. Hacking afirma que Leibniz foi a única pessoa "de seu tempo (...) a declarar o caráter relacional das avaliações de probabilidade (*probability judgments*)"[10]. A ideia de que a relação entre proposições ou sentenças pode ser uma medida de probabilidade é uma ideia fundamental para a lógica indutiva. Ainda de acordo com o autor, "as probabilidades indutivas são relativas à evidência – nós não podemos falar sobre a probabilidade *p*; nós podemos apenas nos referir à probabilidade *p* relativa a *r*"[11].

Por outro lado, também temos o problema clássico do filósofo cético Hume em relação ao pensamento indutivo. Ele coloca o problema em 1739, nas "Investigações sobre o entendimento humano".[12] Para o ele, as inferências em relação ao futuro baseadas em fatos do passado são uma questão de hábito:

> O problema cético a respeito do futuro, frequentemente chamado de problema da indução, foi publicado pela primeira vez em 1739, no Tratado sobre a natureza humana de David Hume. [O Tratado] coloca em dúvida que qualquer fato conhecido a respeito de objetos ou eventos do passado dê qualquer razão para crenças a respeito de objetos ou eventos futuros (...). Este pão irá me nutrir? Hume argumenta que nenhuma coleção de observações passadas sobre alimentação dá qualquer razão para se pensar que o próximo pedaço de pão será nutritivo. Nossas expectativas são formadas pelo costume de pelo hábito, mas falta a elas a justificação. (HACKING, 2006, p. 176, tradução nossa).

Nesta Dissertação, seguimos a argumentação do programa filosófico que, de acordo com Hacking, começa com Leibniz e é representado por Keynes, Jeffreys e Carnap. O programa

[10] Ibid., p. 135.
[11] Ibid., p. 135.
[12] Para uma tradução para o português, ver Hume (2004).

filosófico de Hume é seguido por Popper, Miller, Gilles, entre outros. Esta é uma dualidade clássica e importante na filosofia. Nosso trabalho é um segmento deste debate e está particularmente concentrado nos temas de suporte probabilístico de uma hipótese pela evidência, de corroboração e de estatística.

Popper, por exemplo, defende sua visão na "Lógica da pesquisa científica". O autor sustentou que a indução não tem lugar na lógica da ciência. A ciência, na sua opinião, é um processo dedutivo no qual os cientistas formulam hipóteses e teorias que são testadas derivando consequências observáveis particulares. As teorias científicas não são confirmadas ou verificadas. Elas podem ser falsificadas e rejeitadas ou aceitas provisoriamente, se corroboradas na ausência de falsificação pelos tipos adequados de testes científicos. Popper considerou o problema da indução intransponível, mas argumentou que a ciência não se baseia de fato em métodos indutivos: a teoria da indução é, para Popper, supérflua.

Para adentrar neste debate, vamos primeiramente discorrer mais a respeito da lógica indutiva na argumentação, mostrando como ela difere de um processo dedutivo. Vamos apresentar a lógica indutiva como um processo de relação evidencial entre argumentos. Também citaremos outra vertente, da lógica indutiva pura, e, por fim, mostraremos como o conceito de evidência pode ser tratado na lógica da inconsistência formal $\mathrm{LET_F}$.

2.2 Lógica indutiva na argumentação: relação evidencial

Nesta seção, será definida a chamada lógica indutiva, comparando-a à lógica dedutiva no campo da *argumentação*. Nesse contexto, a lógica indutiva tem caráter aplicado. Vamos seguir algumas definições de Skyrms (1966, p. 2-9), autor que apresenta quatro definições para diferenciar esses dois tipos de argumentos. São elas:

Definição 1: "Um argumento é uma lista de afirmações, uma delas designada como conclusão e as restantes como premissas"

Definição 2: "A *Lógica* é o estudo da força da conexão evidencial entre as premissas e as conclusões dos argumentos"

Definição 3: "Um argumento é *dedutivamente válido* se, e somente se, é *impossível* que a sua conclusão seja falsa enquanto suas premissas são verdadeiras"

Definição 4: “um argumento é indutivamente forte se, e somente se, ele satisfaz duas condições:

i. É improvável que a sua conclusão seja falsa, considerando que suas premissas sejam verdadeiras.
ii. Ele [o argumento] não é dedutivamente válido.”

Vamos seguir e explicar o raciocínio de Skyrms (1966), que culmina nas definições acima expostas, utilizando apenas alguns exemplos diferentes daqueles utilizados pelo autor. Na lógica indutiva, pode ser provável que uma conclusão seja falsa considerando que (dado que) as premissas sejam verdadeiras e que o argumento seja fraco do ponto de vista indutivo. Ou seja, pode ser que ocorra o caso de uma conclusão ser falsa dados que as premissas são verdadeiras. Já no caso da lógica dedutiva, se as premissas são verdadeiras, a conclusão também precisa ser verdadeira: como vimos na Definição 3 acima de Skyrms[13], é impossível que as premissas sejam verdadeiras e a conclusão falsa. Na lógica indutiva, se considerarmos as premissas verdadeiras, pode ocorrer que a conclusão seja falsa, sendo assim o argumento fraco indutivamente. Vejamos um exemplo:

Premissa: Existe um monstro no Lago Ness. (A)
Conclusão: Existe um monstro no Lago Ness que fala. (B)

Vamos supor que exista um monstro do Lago Ness, um animal vivo e pré-histórico remanescente e que raramente é visto, mas existe. Mesmo considerando essa premissa verdadeira, é bastante provável que a conclusão seja falsa (ou seja, um condicional entre a sentenças A e B: qual seria a probabilidade de B *dado* A). Se o monstro realmente existir, é muito improvável que ele fale. O argumento é indutivamente fraco. Na lógica dedutiva, não é possível que as premissas sejam verdadeiras e que a conclusão seja falsa. Vejamos um exemplo:

Premissa 1: Todo *sommelier* já provou ao menos um vinho *Bordeaux*.
Premissa 2: Carla nunca provou um vinho *Bordeaux*.
Conclusão: Carla não é *sommelier*.

[13] Ibid., p. 7.

Neste caso, é impossível não concluir que “Carla não é *sommelier*”, ou seja: é impossível que as duas premissas sejam verdadeiras e a conclusão seja falsa. Já na lógica indutiva, não é apenas avaliado o valor verdade das premissas e da conclusão, como também a *relação evidencial* entre as premissas e a conclusão. O autor esclarece melhor este ponto a respeito da relação evidencial da seguinte maneira:

> O tipo de probabilidade que classifica a força indutiva dos argumentos — nós iremos chamá-la de *probabilidade indutiva* — não depende das premissas individualmente ou das conclusões individualmente, mas da *relação evidencial* entre as premissas e a conclusão. (SKYRMS, 1966, p. 10, grifos do autor, tradução nossa).

O argumento do monstro do Lago Ness, considerando a sua relação evidencial, é um argumento indutivamente fraco. Neste caso, analisamos a probabilidade condicional (indutiva) do argumento. Se formos considerar a conjunção da sentença A e da sentença B, ao invés do condicional que acabamos de mencionar, o argumento é muito pouco provável porque sua premissa é pouco provável.

Há também o caso de um argumento ter uma conclusão bastante provável e ao mesmo tempo não ser indutivamente forte. Isto ocorre quando a conclusão é em si, isoladamente, bastante provável: seria então muito pouco provável que a conclusão fosse falsa e as premissas fossem verdadeiras. Porém, se formos considerar as premissas como verdadeiras, a conclusão torna-se pouco provável. O argumento, então, *não é* indutivamente forte. Vejamos um exemplo de Skyrms (1966, p. 11) para ilustrar este caso:

Premissa: Existe um homem em Cleveland que tem 1999 anos e 11 meses de idade e está em boas condições de saúde.

Conclusão: Nenhum homem irá viver até os 2000 anos de idade.

Todos podem concordar que a probabilidade de um homem viver por tanto tempo é muito baixa, portanto, a conclusão é muito provável, é bastante razoável concordar com ela. Porém, se a premissa for verdadeira, a conclusão passa a ser improvável: se existir um homem de determinada cidade que tem 1999 anos e 11 meses de idade, é bem provável que ele possa viver ao menos mais um mês de vida e complete 2000 anos de idade.

2.3 Lógica indutiva pura

Encontramos uma referência bibliográfica durante este estudo que num primeiro momento nos pareceu bastante importante, tanto pelo título quanto pelo fato de seguir a lógica indutiva de Carnap. A chamada lógica indutiva pura faz parte da lógica matemática. Ela é o estudo da probabilidade racional nesse contexto, cooperando para o debate a respeito da racionalidade (ou raciocínio) sobre a incerteza. O livro de Paris e Vencovská (2015) reúne os principais resultados de estudos da área publicados nos últimos 70 e poucos anos, desde a primeira publicação (em 1932)[14] sobre lógica indutiva pura que os autores consideram mais importante. O livro segue a tradição de Carnap, no sentido de que os princípios racionais fornecem a base para associar probabilidades a sentenças de uma linguagem lógica.

O principal objetivo do trabalho de Paris e Vencovská (2015, p. 3-4) é utilizar uma formulação puramente matemática para definir como é feita a atribuição de probabilidades a qualquer evento ou sentença. Essas inferências são consideradas inferências *lógicas* ou *racionais*, em oposição às inferências *estatísticas*. A diferença deve-se ao fato de que, ao invés de depender de conceitos como frequência, por exemplo, as inferências lógicas dependem de fatores como *simetria*, *informação relevante* e *informação irrelevante*, *analogia*.

A questão chave à qual a chamada lógica indutiva pura se dedica é: na ausência de conhecimento prévio, qual é o grau de crença, com bases racionais, que alguém deve ter em uma sentença para que ela seja verdadeira em uma dada estrutura? A crença é identificada com probabilidade e sua racionalidade deve seguir os princípios básicos que citamos anteriormente:

1) Simetria: se duas sentenças A e B são equivalentes, atribua probabilidades iguais a elas.

2) Irrelevância e relevância: o ato de condicionar uma sentença A a uma B (após o recebimento de B como nova informação) não deve alterar a probabilidade atribuída para A no caso de B ser irrelevante para A, e não deve diminuir a probabilidade atribuída para A, caso contrário.

3) Analogia: Quanto mais uma sentença A é "parecida" com uma sentença B, tanto mais a condicionalização de A em B aumenta a probabilidade de A.

[14] Ver JOHNSON, W.E. Probability: the deductive and inductive problems. *Mind*, vol. 41 (1932), p. 409-423.

A primeira parte do livro é dedicada a definir noções básicas e a provar resultados básicos da configuração matemática subjacente. A segunda parte do livro é dedicada à lógica indutiva pura unária. Já a terceira parte do livro generaliza a estrutura para cobrir predicados de outras aridades. O *review* de Martin (2016) adverte que:

> o leitor interessado em obter conhecimentos gerais sobre a área e obter uma compreensão das principais noções e resultados irá ler a parte 1 e a maioria, senão toda, a parte 2, mas provavelmente não se aventurará na parte 3; é melhor deixar esta parte para especialistas ou aqueles determinados a se tornarem especialistas (MARTIN, 2016).

Isto esclarece que não se trata de um livro introdutório. Os capítulos, por exemplo, não podem nem ser lidos de forma independente um do outro. Também não encontramos referências a Popper ou a outras visões antagônicas. Estes dois motivos, a saber, ser um livro bastante avançado do ponto de vista formal e um livro que não trata de um debate mais filosófico, foram fatores que nos levaram à conclusão de que não seria produtivo entrar em maiores detalhes a esse respeito, pois está distante do nosso debate sobre o *peso da evidência*. No entanto, achamos importante mencioná-lo, pelo menos brevemente: existe uma linha de uma lógica indutiva pura, que segue os escritos de Carnap, e a indução pode ter um tratamento formal.

Finalmente, após discorrer sobre a lógica indutiva no campo da argumentação e sobre a lógica indutiva pura, gostaríamos de mencionar que, apesar de diferentes, a característica que as une é a sua diferença em relação a qualquer processo dedutivo. Embora a inferência indutiva não seja facilmente caracterizada, temos uma marca clara do que seja a indução: as inferências indutivas são contingentes, as inferências dedutivas são necessárias. A inferência dedutiva nunca pode apoiar julgamentos contingentes, como previsões meteorológicas, nem a dedução por si pode explicar sozinha o defeito do carro de alguém, descobrir o genótipo de um novo vírus ou reconstruir rotas comerciais do século XIV.

Por outro lado, a inferência indutiva pode fazer essas coisas com mais ou menos sucesso porque, na frase de Peirce (1992), as induções são ampliativas. A indução pode amplificar e generalizar nossa experiência, ampliar e aprofundar nosso conhecimento empírico. A dedução, por outro lado, é explicativa. A dedução ordena e reorganiza nosso conhecimento

sem aumentar seu conteúdo, como dizem alguns. Contudo, essa posição é muito debatível e conduz para uma nova área que escapa ao objetivo desta Dissertação.

2.4 Evidência e verdade no contexto da lógica

O artigo de Rodrigues, Carnielli e Bueno-Soler (2020) apresenta a lógica da evidência e da verdade LET_F como uma extensão da lógica FDE de quatro valores de Belnap-Dunn. A lógica LET_F é uma lógica de inconsistência formal e indeterminação equipada com um operador de não-classicidade ●, e seu dual ○ que expressa a noção de classicidade. O operador ○ em LET_F recupera a lógica clássica para proposições em seu escopo, onde a evidência é considerada uma noção mais fraca do que a verdade no sentido de que pode haver evidência para uma proposição α mesmo se α não for verdadeira.

O artigo foi concebido para ser um desenvolvimento adicional da abordagem à paraconsistência como preservação das evidências apresentadas em Carnielli e Rodrigues (2017, 2019), onde uma interpretação das contradições em termos de evidências não-conclusivas foi proposta. A suposição subjacente é que não existem contradições verdadeiras, mas sim, contextos argumentativos em que evidências, bem como a ausência de quaisquer evidências, podem ocorrer. A avaliação semântica é capaz de expressar apenas que há ou não evidências de uma proposição α, enquanto a semântica probabilística apresentada no artigo pretende expressar o grau de evidência desfrutado por uma dada proposição.

LET_F é um sistema lógico capaz de expressar preservação de evidências e preservação da verdade. O objetivo principal do artigo é propor uma semântica probabilística para LET_F onde as afirmações P (α) e P (○α) expressam, respectivamente, a quantidade de evidência disponível para α e o grau em que se espera que a evidência para α se comporte classicamente – ou não classicamente para P (● α).

A fim de capturar essa ideia de preservação de graus de evidência uma noção não-clássica de probabilidade é empregada. A semântica probabilística proposta em termos de uma nova noção de probabilidade paraconsistente segue as ideias apresentadas em Bueno-Soler e Carnielli (2016) e Carnielli e Bueno-Soler (2017).

Um cenário probabilístico é paracompleto quando P (α) + P (¬α) <1, e paraconsistente quando P (α) + P (¬α)> 1, e em ambos os casos, P (○α) <1.

Se P (○ α) = 1, ou P (● α) = 0, a probabilidade clássica é recuperada para α, isto é, α segue as leis da probabilidade clássica.

As propriedades da lógica permitem obter novas versões de resultados conhecidos da teoria de probabilidade padrão. Uma extensão dessas ideias no sentido de "peso paraconsistente de evidência" é bastante natural, bastando que se defina da forma esperada as noções de probabilidade condicional paraconsistente e de *peso da evidência*.[15] O resultado pode ser usado em cenários investigativos e bancos de dados com incerteza a respeito de graus de evidência atribuídos a proposições envolvendo contradições. Essa linha de investigação, contudo, será deixada para trabalhos futuros.

[15] O conceito de peso da evidência será exposto de maneira detalhada no capítulo 4 desta Dissertação.

3. PROBABILIDADE CONDICIONAL, TEOREMA DE BAYES, PROBABILIDADE SUBJETIVA E ESTATÍSTICA BAYESIANA

3.1 Introdução

Neste capítulo, fazemos uma exposição de conceitos fundamentais para a compreensão do conceito de *peso da evidência*, uma exposição especialmente de conceitos de probabilidade e estatística. Desta forma, este capítulo pode ser considerado uma introdução conceitual aos próximos capítulos. Em primeiro lugar, discorremos sobre algumas questões do bayesianismo de modo mais geral: falamos sobre a origem do teorema de Bayes, quais foram as primeiras publicações bayesianas e a relação entre o bayesianismo e a corrente filosófica de Carnap. Após esta breve introdução, apresentaremos detalhes mais técnicos a respeito do teorema de Bayes, além do conceito de probabilidade condicional, de probabilidade subjetiva e inferência bayesiana, com a inclusão de alguns exemplos.

3.2 Algumas considerações gerais sobre o bayesianismo

Selecionamos, de início, os tópicos mais importantes para a compreensão da característica indutiva da probabilidade e da estatística bayesiana. Foi necessário fazer um recorte do bayesianismo. Um artigo de Good é bastante citado quando se discute a variedade e amplitude do bayesianismo: "46.656 *varieties of bayesians*".[16] Este número corresponde a onze facetas do bayesianismo e aos diferentes posicionamentos que podem ser tomados em cada uma delas. Alguns bayesianos, em relação às probabilidades físicas, por exemplo, podem assumir que elas existem, já outros afirmam que não existem, e há ainda a possibilidade de fazer uso das probabilidades físicas, porém sem o que Good chama de "comprometimento filosófico". Também há variações em relação aos axiomas, em relação às utilidades, à intuição da probabilidade – a probabilidade subjetiva pode ser considerada primária ou a probabilidade lógica (ou a credibilidade) pode ser primária, etc.

Esta diversidade não é exclusiva do bayesianismo. A matemática, que no início do século XX podia ser dividida em cerca de 12 áreas ou segmentos, hoje pode ser dividida em cerca de 70 áreas de pesquisa. A expansão foi tão significativa (a partir principalmente da década de 1980), que, de acordo com Devlin,

---

[16] Good (1983, p. 20-21).

> Existem tantas técnicas matemáticas diferentes, com novas delas sendo desenvolvidas a todo o tempo, que é impossível cobrir todas elas numa educação K-16.[17] No momento em que os graduados entram na força de trabalho, muitas das técnicas específicas aprendidas nos quatro anos de graduação provavelmente não serão tão importantes, enquanto novas técnicas estão à todo vapor. O foco educacional deve ser em aprender a como aprender. (DEVLIN, 2012, p. 6, tradução nossa).

Não devemos, portanto, dada a complexidade da matemática e mais especificamente da probabilidade e estatística, ter a intenção de cobrir de maneira ampla o bayesianismo, mas apenas fazer uma seleção dos tópicos mais relevantes ao problema filosófico que levantamos a respeito da relação entre evidência e hipótese.

Apesar da grande variedade de bayesianos, algumas considerações históricas e conceituais merecem ser mencionadas brevemente, de maneira geral, antes de adentrarmos nos detalhes mais técnicos do teorema de Bayes, da estatística bayesiana e da probabilidade subjetiva. McGrayne (2011) faz um excelente relato histórico sobre o teorema de Bayes, desde a sua descoberta até as mais recentes aplicações. Estima-se que Bayes descobriu o teorema no final da década de 1740. O teorema só foi publicado em 1763, postumamente – Bayes faleceu em 1761. Seu amigo Reverendo Richard Price (1723-91) descobriu o teorema nos escritos de Bayes e, após correções e revisões, publicou o artigo intitulado "*An essay towards solving a problem in the doctrine of chances*". McGrayne (2011) relata que Price só teve interesse em publicar o artigo porque achava que poderia ser uma resposta a Hume:

> Organizando os artigos de Bayes, Price encontrou 'uma solução imperfeita de um dos mais difíceis problemas na doutrina das chances'. Era o artigo de Bayes sobre a probabilidade das causas, em mover-se das observações sobre o mundo real até a sua mais provável causa. Em um primeiro momento Price não viu razão em dedicar muito trabalho ao artigo. (...) Mas assim que Price decidiu que o artigo de Bayes era a resposta ao ataque de Hume sobre a causação, começou a prepará-lo para publicação. (MCGRAYNE, 2011, p. 10-11, tradução nossa).

[17] A educação K-16 à qual Devlin refere-se corresponde, no Brasil, ao período estudantil até o término de um curso de graduação em matemática.

Price foi responsável pela primeira contribuição conhecida para a teoria da decisão bayesiana: a resposta que Price deu ao argumento cético de Hume contra os milagres atribuídos ao cristianismo. O nome de Price sobrevive na "Tabela Price", também chamado de sistema francês de amortização, um método usado em amortização de empréstimo. O método foi apresentado em 1771 por Price em sua obra "*Observations on Reversionary Payments*"[18].

O teorema de Bayes mais tarde se tornaria de grande valia na estatística, mas Bayes não viveu para vê-lo ganhar destaque. Quando Bayes morreu, em 1761, e Price foi convidado a examinar seus papéis e decidir se havia algo digno de publicação, Price também acrescentou uma introdução e exemplos, já que Bayes aparentemente não estava interessado em aplicações. Foi Laplace quem formulou a versão moderna de "probabilidade inversa" do teorema de Bayes, onde as probabilidades anteriores são combinadas com a função de verossimilhança observada para obter a distribuição condicional posterior.

O interesse de Price na fórmula não foi motivado puramente pela matemática. Ele viu o teorema de Bayes como um meio de provar a existência de Deus. Em 1748, Hume publicou "*Of Miracles*"[19]. No ensaio, Hume argumenta que quando alguém afirma ter visto um milagre, essa é uma evidência muito pobre de que ele realmente aconteceu, já que vai contra o que vemos todos os dias. Hume escreve, em seus "*Of Miracles*": "Nenhum testemunho é suficiente para estabelecer um milagre, a menos que o testemunho seja de tal tipo, que sua falsidade seja mais milagrosa do que o fato que se esforça para estabelecer." Embora Hume nunca tenha admitido isso diretamente, seu ensaio é geralmente entendido como um argumento contra os relatos milagrosos do cristianismo.

O reverendo Price discordava. Ele acreditava que a fórmula de Bayes provava que Hume estava errado. Em um ensaio de 1767, Price mostra que, mesmo se uma pessoa observar que a maré subiu um milhão de vezes, em bases estatísticas, ela não pode razoavelmente dizer que ela nunca parará de subir. Usando o teorema de Bayes, com base em milhões de observações, Price calculou que há uma probabilidade de a maré não subir um dia, e este valor está entre 1 em 600.000 e 1 em 3 milhões. Portanto, argumentou ele, não é possível eliminar a chance de um milagre com base em um grande número de observações negativas.

Embora Hume (1751) nunca tenha negado seu argumento, ele levou a sério a réplica de Price. "Admito que a Luz, na qual você colocou esta controvérsia, é nova, plausível e

[18] Ver Price (1771).
[19] Cf. Hume (1751). A primeira edição é de 1748; a partir de 1758 o título mudou para *An Enquiry Concerning the Human Understanding*. O texto "*Of Miracles*" aparece como a Seção X.

engenhosa, e talvez sólida", escreveu Hume a Price[20]. Embora não tenha provado a existência de Deus, o teorema de Bayes provou ser útil de outras maneiras. A fórmula agora é usada para testar novos medicamentos, avaliar a previsão do tempo e melhorar a recepção do telefone celular, sendo ainda fundamental para os mais recentes desenvolvimentos em inteligência artificial. Na seção 3.4 apresentaremos algumas dessas aplicações.

Por todas as contribuições de Price para o teorema de Bayes, concordamos com McGrayne (2011) quando ela afirma que em padrões contemporâneos o teorema deveria chamar-se "teorema de Bayes-Price", por toda a influência de Price na publicação, não só pela descoberta, mas pelas correções, edições e debate. Poderíamos também incluir o nome de Laplace no teorema, já que ele o descobriu de forma independente em 1774. Até a década de 1950, no entanto, o teorema não era conhecido como teorema de Bayes, foi só a partir desse momento que o seu nome ficou consolidado.

Na década de 1950, são feitas as três primeiras publicações de maior impacto e alcance que são consideradas bayesianas. A primeira delas é o livro "*Probability and the weighing of evidence*", de Good (1950). O livro já estava pronto para ser publicado em 1946, porém, sua divulgação teve de ser adiada em grande parte devido ao sigilo que vigorava no período da Guerra Fria. A segunda publicação é a de Lindley (1953), um artigo intitulado "*Statistical inference (with discussion).*". A terceira é uma publicação de Savage (1954), "*Foundations of statistics*", que McGrayne (2011, p. 103) classifica como um "livro revolucionário".

Good (1975) faz uma representação do posicionamento dos estatísticos num espaço abstrato nos anos de 1950 e em 1973. A figura a seguir mostra em 1(a) que praticamente não havia estatísticos bayesianos em 1950. Já em 1973, vemos um grupo formado por Ramsey, de Finetti, Savage e Lindley, entre outros. Good está ao centro, pois afirma que ele representa um "compromisso Bayes/não Bayes" (GOOD, 1975, p. 39). À direita, tanto em 1(a) como em 1(b) vemos agrupados os estatísticos da interpretação frequentista: os que seguem a linha de Fisher e os que seguem a linha Neyman-Pearson. Em 1950, podemos ver os estatísticos que seguem a linha da interpretação lógica da probabilidade espalhados entre os bayesianos: Keynes, Johnson, Jeffreys e Carnap. Em 1973, podemos ver uma mudança de posicionamento de Carnap, que se encontra bem mais próximo de Good.

[20] O trecho encontra-se em uma publicação das correspondências de Price. Ver Thomas e Peach (1983, p. 45-47).

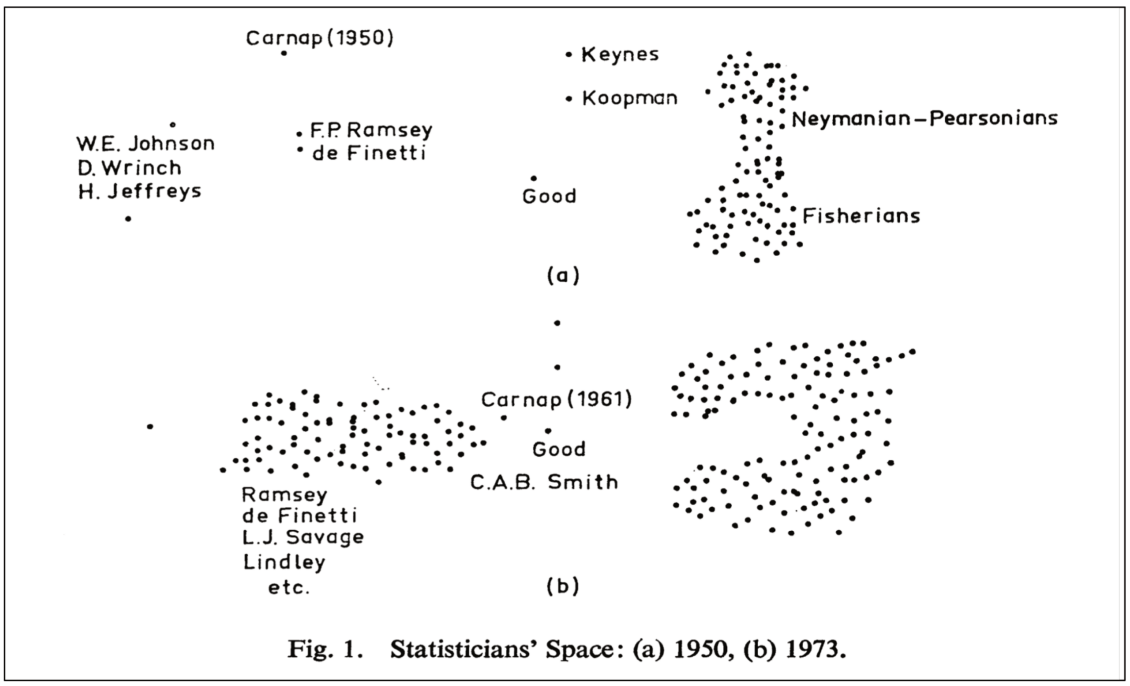


Fig. 1. Statisticians' Space: (a) 1950, (b) 1973.

Figura 1 - Espaço de estatísticos: (a) 1950, (b) 1973. FONTE: Good (1975).

É importante mencionarmos e analisarmos essa aproximação do posicionamento de Carnap em direção ao posicionamento de Good. Como vimos na seção 2.3, Carnap foi o grande responsável pelo desenvolvimento do que é classificado hoje como lógica indutiva pura. Galavotti (2005, p. 164-178) resume a trajetória de Carnap no campo da lógica indutiva. Carnap foi aluno de Frege, tendo estudado física, filosofia e matemática. Ele influenciou as diversas fases do empirismo lógico, acompanhando as mudanças que ocorreram no movimento filosófico. Seus estudos sobre probabilidade foram iniciados na década de 1940 e culminaram na publicação de *Logical foundations of probability:*

> Após a publicação, em 1950, de *Logical foundations of probability*, uma pedra fundamental da literatura da filosofia da probabilidade, Carnap dedicou sua atenção quase que exclusivamente à probabilidade nos últimos 20 anos de sua vida, publicando muitos trabalhos importantes. Apesar da sua perspectiva neste longo período ter sofrido mudanças significativas, Carnap nunca abandonou o programa de desenvolver uma lógica indutiva. (GALAVOTTI, 2005, p. 165, tradução nossa).

A mudança principal no pensamento de Carnap ocorre na década de 1960, por isso podemos ver a transição de posicionamento apontada por Good na Figura 1. No prefácio à

segunda edição de *Logical foundations of probability*, que foi escrito em 1962, Carnap expõe com clareza qual é o seu conceito de probabilidade elencando as suas afinidades com outros estudiosos da probabilidade e estatística:

> Minha concepção de probabilidade lógica (denominada 'probabilidade$_1$' neste livro) tem algumas características em comum com aquelas de outros autores, e.g., John Maynard Keynes, Frank P. Ramsey, Harold Jeffreys, Bruno De Finetti, B. O. Koopman, Georg Henrik von Wright, I. J. Good, e Leonard J. Savage, para mencionar apenas os nomes mais conhecidos. Todas essas concepções partilham das seguintes características. Elas são diferentes da concepção frequentista ('probabilidade$_2$' neste livro). Elas enfatizam a relatividade da probabilidade em relação à evidência. (Por esta razão, alguns autores chamam a sua concepção de 'subjetiva'; no entanto, esse termo não parece muito apropriado para a probabilidade lógica (...)). Ademais, a probabilidade numérica de um evento desconhecido possível pode ser considerada como um quociente de uma aposta coerente [*fair betting quotient*]. E, finalmente, se as relações lógicas (e.g., implicação lógica ou incompatibilidade) mantêm-se entre proposições dadas, então as suas probabilidades precisam, de acordo com estas concepções, satisfazer determinadas condições (usualmente baseadas em axiomas) para garantir a racionalidade das crenças e das ações, e.g., apostas, baseadas nessas probabilidades. Eu tenho a impressão de que o número daqueles que pensam e trabalham na direção indicada tem crescido. Este certamente é o caso entre os filósofos. Mas parece que também entre aqueles que trabalham com a estatística matemática mais e mais começam a considerar o uso exclusivo consuetudinário do conceito frequentista da probabilidade como insatisfatório e estão procurando por outro conceito. (CARNAP, 1962, p. xiv, tradução nossa).

Carnap então passa a considerar a lógica indutiva como uma teoria da decisão, utilizando o conceito de *coerência*[21], característico da probabilidade subjetiva. De acordo com Galavotti:

> nos seus escritos tardios, Carnap considera a lógica indutiva como uma teoria da decisão. Contextualmente, estes escritos incorporam uma justificação dos princípios básicos da lógica indutiva em termos de coerência, típica da abordagem subjetivista de Ramsey e de Finetti. (GALAVOTTI, 2005, p. 175, tradução nossa).

[21] O conceito de coerência será apresentado na seção 3.4, "Estatística bayesiana e probabilidade subjetiva".

Não é apenas o conceito de coerência que Carnap assume como parte fundamental da sua lógica indutiva. Existem pelo menos mais dois conceitos fundamentais: *credence* e *credibility*. *Credence* é a crença de um agente racional num determinado momento e está conectado com a ideia de lógica indutiva como teoria da decisão, que lida com decisões racionais de um agente. Para que um determinado agente aja racionalmente, ele precisa tomar decisões racionais baseadas em funções *credence*. Já *credibility* "expressa a disposição permanente dele [do agente] para formar e mudar as suas crenças à luz de informações"[22].

Good (1975, p. 41) relata que Carnap afirmou que o que Good chama de "probabilidade subjetiva" deveria ser chamado de "credibilidade racional" (*rational credibility*); mas se Good fosse escolher uma definição que expressasse melhor a sua própria visão, ele a definiria como *lógico-subjetivista*. Podemos constatar como os dois autores se aproximam em suas visões, como a lógica se aproxima da probabilidade, ou como as interpretações lógica e subjetivista da probabilidade podem se encontrar. Good[23] cita três tipos diferentes de probabilidade, e afirma que há uma gradação contínua entre elas: a probabilidade psicológica[24], a probabilidade subjetiva e a probabilidade lógica que, segundo ele, se ela "existe ou não é um ideal para manter na mente como uma verdade absoluta". Em uma publicação póstuma de 1980, Carnap mostra-se bastante próximo do subjetivismo: ele classifica a sua posição como um "ponto de vista subjetivista (modificado)" (CARNAP, 1980 apud GALAVOTTI, 2005, p. 175).

Neste capítulo, portanto, vamos analisar melhor os conceitos básicos que formam o bayesianismo, que incorpora tanto a visão subjetivista quanto alguns aspectos da lógica indutiva desenvolvida por Carnap. Nas palavras de Galavotti:

> Os métodos de Carnap pertencem à ampla família dos métodos Bayesianos, a discussão a respeito deles até um certo ponto misturou-se com aqueles sobre confirmação Bayesiana, que representa a principal tendência da literatura sobre confirmação probabilística. Essa direção, por exemplo, foi tomada por Richard Jeffrey que, após estudar sob orientação de Carnap em Chicago e mais tarde ter publicado os seus últimos manuscritos, chegou à convicção de que a concepção da lógica de confirmação de Carnap como uma mistura de um componente puramente lógico e um componente puramente empírico deveria ser substituída por uma abordagem mais eclética, mais próxima do Bayesianismo

[22] Ibid., p. 176.
[23] Ibid., p. 41.
[24] De acordo com Good (Ibid., p. 40), a probabilidade psicológica consiste em um "julgamento momentâneo puro de um grau de crença ou uma intensidade de convicção, sem recorrer à coerência, (...) do tipo que foi investigada, por exemplo, por John Cohen em crianças."

> subjetivista de Bruno de Finetti. (GALAVOTTI, 2005, p. 178, tradução nossa).

3.3 Probabilidade condicional e teorema de Bayes

O teorema de Bayes é uma consequência algébrica da *probabilidade condicional*. Para que esses conceitos sejam definidos, é necessário definir primeiramente os axiomas a partir dos quais podemos fazer o cálculo de probabilidades. Antes disso, precisamos introduzir brevemente algumas definições mais elementares: a definição de probabilidade, evento, espaço amostral e estatística.

Imagine um experimento, prático ou hipotético, o qual é possível determinar com antecipação todos os resultados possíveis. O conjunto de todos os resultados possíveis é chamado de *espaço amostral*, e é representado pela letra grega Ω. Um subconjunto qualquer do espaço amostral é chamado de *evento*, e representamos os eventos por letras romanas maiúsculas. Para cada evento, podemos atribuir uma *probabilidade*, que representamos, por exemplo, por $P(A)$, que é um número que indica a probabilidade de que do evento A vai ocorrer.

Existem diferentes interpretações para a probabilidade, porém, a teoria matemática da probabilidade não depende destas interpretações. Podemos definir a probabilidade como a quantificação da incerteza, a quantificação de um evento incerto. Vejamos os problemas com os quais a teoria matemática da probabilidade lida:

> Quase todo o trabalho na teoria matemática da probabilidade, dos livros didáticos mais elementares até a pesquisa mais avançada, tem sido relacionado com os dois seguintes problemas: (i) métodos para determinar as probabilidades de certos eventos a partir das probabilidades especificadas de cada resultado possível de um experimento e (ii) métodos para revisar as probabilidades de eventos quando informações relevantes adicionais são obtidas. (DEGROOT; SCHERVISH, 2012, p. 6, tradução nossa).

Wackerly et al. (2008. p. 1-2) fazem uma exposição de várias definições diferentes do conceito de *estatística* e apontam as características em comum das definições: a estatística é um ramo da matemática (ou um ramo do método científico) que coleta dados para fazer inferências; a estatística recolhe subconjuntos de dados (*amostra*) de uma coleção ampla de dados para que se possa inferir características do conjunto completo (*população*). Os autores

concluem que todas as definições de *estatística* que eles selecionaram implicam que "a estatística é uma teoria da informação, com a realização de inferências como seu objetivo"[25].

Para que as inferências estatísticas sejam feitas, existem vários métodos de coleta e análise de dados que envolvem medidas numéricas descritivas como média, variância e desvio padrão. E a relação entre probabilidade e estatística pode ser resumida da seguinte maneira: "Básico à inferência é o problema de calcular a probabilidade de uma amostra observada. Como resultado, a probabilidade é o mecanismo utilizado para fazer inferências estatísticas"[26]. Em outras palavras, "Uma *inferência estatística* é um procedimento que produz uma afirmação probabilística a respeito de algumas ou todas as partes de um modelo estatístico." (DEGROOT; SCHERVISH, 2012, p. 378, grifos dos autores).

Feitas estas definições elementares, vamos considerar os seguintes axiomas da probabilidade[27]:

1) Axioma da não-negatividade, $P(A) \geq 0$ (a probabilidade de um evento qualquer é maior ou igual a zero);
2) Axioma da normalização, $P(\Omega) = 1$ (ou seja, o espaço amostral é igual a um);
3) Axioma da adição infinita:
   para toda sequência infinita de eventos dois a dois disjuntos A_1, A_2, ...,
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Vamos definir o conceito de probabilidade condicional utilizando um exemplo. Suponha que desejamos calcular a probabilidade de um dado apresentar o número 6 após um único lançamento. O dado é não viesado e possui 6 lados. Portanto, a probabilidade de apresentar qualquer lado após um lançamento é igual a $1/6$:

Espaço amostral (Ω) = 1, 2, 3, 4, 5, 6.

Evento A = dado apresentar o número 6.

---

[25] Ibid., p. 2.
[26] Ibid., p. 14.
[27] Para mais detalhes sobre a nossa escolha de axiomas e alguns teoremas importantes da probabilidade, ver item A do Apêndice, "Axiomas da probabilidade: probabilidade absoluta e probabilidade relativa". As operações entre as probabilidades seguem as regras das operações da teoria de conjuntos, para esclarecimentos a esse respeito, pode-se consultar HRBACEK, K.; JECH, T. *Introduction to set theory*. 3rd ed., rev. and expanded. New York: Marcel Dekker, 1999.

$P(A) = 1/6$.

Agora, suponha que, após o lançamento do dado, alguém nos informe que o número apresentado é par, sendo que não foi informado qual é o número exato. Qual é a probabilidade, após essa nova informação, do dado ter apresentado o número 6? A resposta para essa questão é o que definimos por probabilidade condicional. Neste caso do dado, ocorreu uma *atualização de informações*. Num primeiro momento, desejávamos calcular a probabilidade de o dado apresentar o número 6, porém, após sermos informados que o número apresentado foi um número par, nossa probabilidade mudou:

Espaço amostral (Ω) = 1, 2, 3, 4, 5, 6.
Evento A = dado apresentar o número 6.
Evento B = dado foi lançado e apresentou um número par qualquer.
$P(A|B) = 1/3$.

A probabilidade condicional, portanto, representa uma forma de atualização de informações. Se temos uma probabilidade qualquer $P(A)$, por exemplo, podemos atualizar essa probabilidade com a informação de que o evento B ocorreu (ou pode ocorrer). Então temos a probabilidade condicional, $P(A|B)$. A probabilidade $P(A)$ pode ser considerada nesse sentido uma probabilidade *a priori*, e a probabilidade condicional $P(A|B)$ uma probabilidade *a posteriori*. Esta atualização de informações acontece quando a ocorrência do evento B muda a probabilidade inicial $P(A)$.

Existe uma regra para calcular a probabilidade condicional. Sejam A e B dois eventos, $P(B) \neq 0$ e $P(A \cap B)$ a probabilidade da intersecção dos eventos A e B. Ilustramos a seguir os conjuntos (ou eventos) A, B e $A \cap B$ com um diagrama de Venn para maior clareza do conceito:[28]

[28] Lembrando que as operações entre as probabilidades de eventos respeitam as operações da teoria de conjuntos.

Figure 2.1 The outcomes in the event B that also belong to the event A.

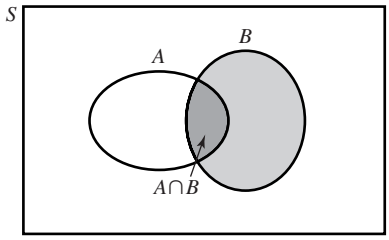


Figura 2 - Eventos A, B e A∩B. FONTE: DeGroot e Schervish (2012).

Definimos, então, a probabilidade condicional $P(A|B)$, ou seja, a probabilidade do evento A ocorrer *dado* que o evento B ocorreu:

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

Se o evento B não mudasse a probabilidade inicial $P(A)$, os eventos A e B seriam *independentes*. Teríamos $P(A) = P(A|B)$, e então $P(A \cap B) = P(A).P(B)$. Já no caso do evento B mudar a probabilidade inicial $P(A)$, classificamos os eventos A e B como *dependentes*. Nosso exemplo do lançamento de um dado foi um exemplo de probabilidade condicional de eventos dependentes. Agora veremos um exemplo de probabilidade condicional de eventos independentes.

Consideremos a seguinte questão: qual a probabilidade de uma criança nascer com alguma determinada característica genética heterozigota, ou seja, qual o valor da probabilidade $P(D \cap d)$, dado que tanto o pai quanto a mãe também são heterozigotos?[29] Vejamos o cálculo dessa probabilidade nesta árvore de probabilidades. Os alelos da mãe e do pai são recessivos (d) ou dominantes (D). A criança pode ser homozigota possuindo a característica dominante (D∩D), ou homozigota recessiva (d∩d) ou heteorozigota (D∩d) ou (d∩D).

[29] Exemplo retirado do curso online *Statistics in medicine (Stanford Online)*, criado pela Universidade de Stanford, disponibilizado pelo sistema *Lagunita Stanford Online* e ministrado por Kristin Sainani. O curso foi retirado da plataforma em março de 2020 e não se encontra mais disponível.

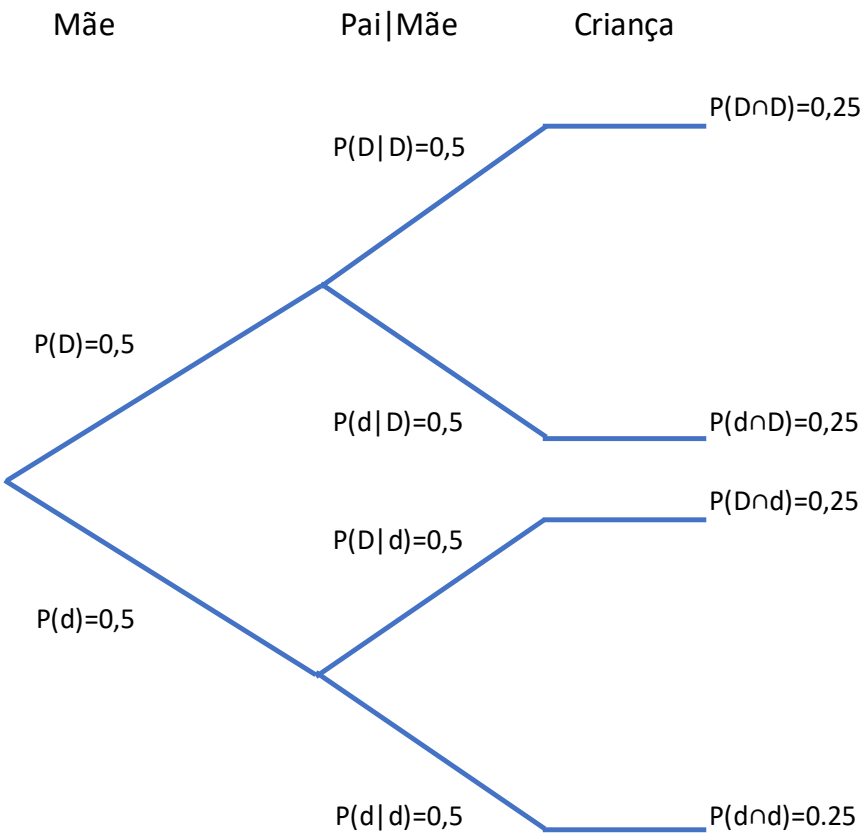


Figura 3 - Diagrama em árvore: eventos condicionais e independentes

A probabilidade de a criança ser heterozigota é igual a $P(D \cap d)$ somada a $P(d \cap D)$, ou seja, 0,5. Podemos constatar também que a probabilidade do alelo paterno transmitir a característica dominante ou recessiva é um evento independente do alelo materno transmitir a característica dominante ou recessiva, pois em qualquer ramo da árvore as probabilidades condicionais do alelo paterno dado o alelo materno $P(♂D|♀D)$ e $P(♂D|♀d)$ e a probabilidade $P(♂D)$ têm o mesmo valor (0,5). Vemos, portanto, que os eventos são condicionais, porém independentes.

Com algumas transformações algébricas, a partir da probabilidade condicional chegamos ao teorema de Bayes:

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

$$P(A \cap B) = P(A|B).P(B)$$

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

$$P(A \cap B) = P(B|A).P(A)$$

$$P(A|B).P(B) = P(A \cap B) = P(B|A).P(A)$$

$$P(A|B).P(B) = P(B|A).P(A)$$

O que nos leva finalmente ao teorema de Bayes:

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

Com o teorema de Bayes podemos calcular o que é chamado de "probabilidade inversa", ou seja, se temos a informação do valor de $P(A|B)$, calculamos o valor de $P(B|A)$, sendo o contrário também verdadeiro. Foi por este motivo que o teorema de Bayes ficou conhecido por cerca de 200 anos como *probabilidade das causas* ou *probabilidade inversa*[30], e só a partir da década de 1950 que recebeu a denominação de Bayes.

Há outra forma de representar o teorema de Bayes, considerando a lei da probabilidade total. Suponha que o evento *B* ocorra num espaço amostral Ω, e que esse mesmo espaço amostral esteja dividido em partições contáveis, em eventos mutuamente exclusivos: $\{A_n : n = 1, 2, 3 \ldots\}$; $\{A_1, A_2, \ldots, A_n\}$ é a partição do espaço amostral dado que $A_i \cap A_j = 0$, sempre que $i \neq j$; $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$. Suponha também que as probabilidades $P(A_i)$ sejam conhecidas. Então, temos a segunda forma de representar o teorema de Bayes:

$$P(A_i|B) = \frac{P(B|A_i).P(A_i)}{\sum_{j=1}^{n} P\left(B\middle|A_j\right).P(A_j)}$$

Particularizando o teorema de Bayes para o caso em que o espaço amostral foi particionado em A e $\overline{A}$, e dado que a probabilidade de eventos mutuamente exclusivos soma 1, ou seja, que $P(A) + P\left(\overline{A}\right) = 1$, podemos simplificar o teorema de Bayes da seguinte maneira:

[30] Cf. McGrayne (2011, p. 12).

$$P(A|B) = \frac{P(B|A).P(A)}{P(B|A).P(A) + P\left(B\middle|\overline{A}\right).P(\overline{A})}$$

Finalmente, para consolidar os conceitos trabalhados nesta seção, vamos apresentar um exemplo [31] com a construção de uma árvore de probabilidades com probabilidades condicionais de eventos dependentes, e calcular uma probabilidade inversa utilizando o teorema de Bayes.

Suponha que uma empresa tenha contaminado 15% de uma região com um composto químico que chamaremos de Z. Sabemos que Z é cancerígeno, que a exposição humana a ele aumenta a probabilidade de desenvolver câncer (PC). Um advogado defende um cliente que está com câncer de pâncreas, e o cliente não sabe se foi exposto ou não ao composto químico Z. O advogado está processando a empresa e precisa quantificar a evidência de que seu cliente tenha desenvolvido câncer por um suposto contato com o composto químico Z, calculando da probabilidade $P(Z|PC)$. Possuímos as seguintes informações:

$P(Z) = 0{,}15$ — Probabilidade de contaminação da área, exposição a Z.

$P(PC|Z) = 0{,}0004$ — Probabilidade de desenvolver câncer de pâncreas, dado a exposição a Z.

$P\left(PC\middle|\overline{Z}\right) = 0{,}0001$ — Probabilidade de desenvolver câncer de pâncreas, dado que não houve exposição a Z.

$P(Z|PC) = ?$ — Probabilidade de ter sido exposto a Z, dado que se tem câncer de pâncreas. Esta é a probabilidade que queremos calcular.

A partir destes dados, podemos calcular $P(Z|PC)$ com o teorema de Bayes considerando a lei da probabilidade total:

$$P(Z|PC) = \frac{P(PC|Z).P(Z)}{P(PC|Z).P(Z) + P\left(PC\middle|\overline{Z}\right).P(\overline{Z})}$$

$$P(Z|PC) = \frac{0{,}0004.0{,}15}{0{,}0004.0{,}15 + 0.0001.0{,}85}$$

[31] Este exemplo também foi retirado do curso *Statistics in Medicine* citado anteriormente.

$$P(Z|PC) = \frac{0{,}0004.0{,}15}{0{,}000145}$$

$$P(Z|PC) = 41{,}4\%$$

Também podemos calcular a probabilidade $P(Z|PC)$ com o teorema de Bayes na sua primeira forma utilizando os dados de um diagrama da árvore de probabilidades, ilustrado na figura a seguir:

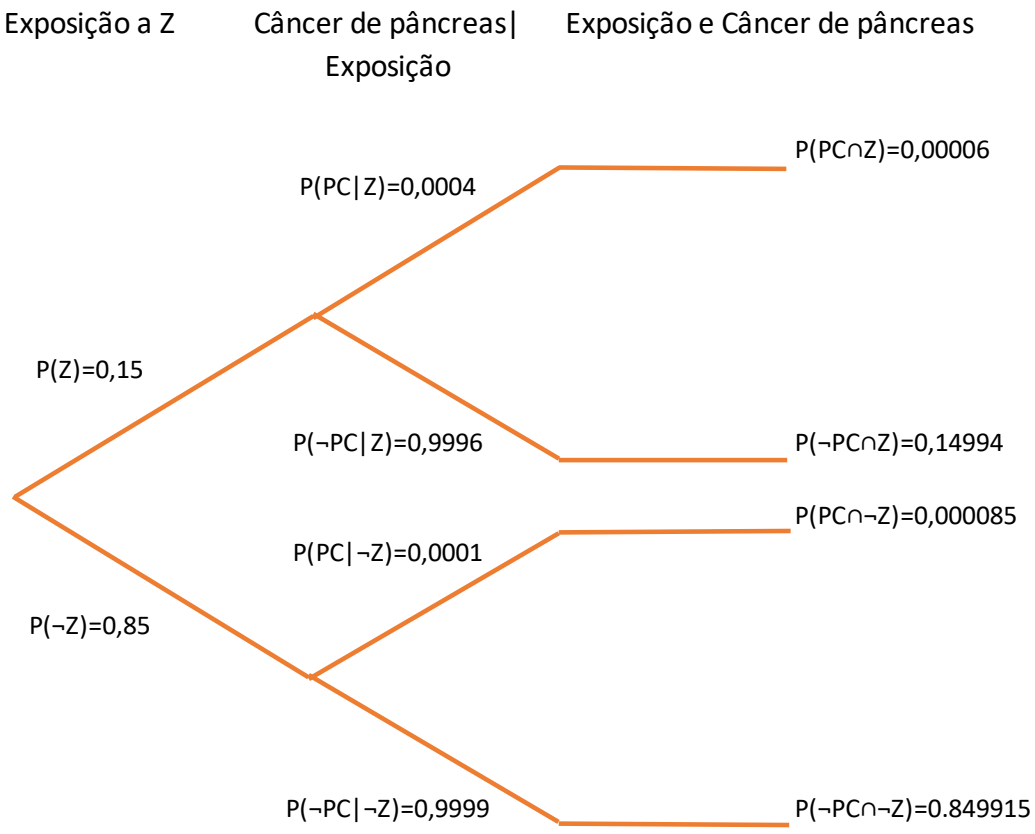


Figura 4 - Diagrama em árvore: eventos condicionais e dependentes

Obtemos o mesmo valor calculado anteriormente:

$$P(Z|PC) = \frac{P(PC|Z).P(Z)}{P(PC)}$$

$$P(PC|Z).P(Z) = P(PC \cap Z)$$

$$P(PC|Z).P(Z) = 0{,}00006$$

$$P(PC) = P(PC \cap Z) + P(PC \cap \overline{Z})$$
$$P(PC) = 0{,}00006 + 0{,}000085 = 0{,}000145$$

$$P(Z|PC) = \frac{0{,}00006}{0{,}000145}$$
$$P(Z|PC) = 41{,}4\%$$

Concluindo, neste caso, portanto, o advogado pode afirmar que há uma probabilidade de 41,4% do seu cliente ter sido exposto ao composto químico Z. Calculamos a probabilidade inversa, pois a partir de $P(PC|Z)$ chegamos ao valor de $P(Z|PC)$. Mostramos que os eventos são dependentes, pois quando sabemos que o evento Z ocorreu, há uma atualização de informações, e assim a probabilidade condicional $P(PC|Z)$ apresenta um valor diferente da probabilidade $P(PC)$.

3.4 Estatística bayesiana e probabilidade subjetiva

Nesta seção, vamos descrever basicamente em que consiste a estatística bayesiana, utilizando alguns exemplos. Nossas principais fontes serão O'Hagan e Forster (2004) e conhecimentos adquiridos em um curso online, "*Bayesian statistics: from concept to data analysis*".[32] Como primeiro passo para compreender a estatística bayesiana, apresentaremos uma maneira mais "bayesiana" de enxergar o teorema de Bayes (seguiremos de perto o raciocínio de O'Hagan e Forster, 2004, p. 2-6).

Retomando o teorema de Bayes na sua forma geral, sejam $A_1, A_2, \dots, A_n$ eventos mutuamente exclusivos e coletivamente exaustivos, então

$$P(A_r|B) = \frac{P(A_r).P(B|A_r)}{P(B)}$$
$$= \frac{P(A_r).P(B|A_r)}{\sum_r P(B|A_r).P(A_r)}$$

[32] Curso oferecido pela plataforma *Coursera*, ministrado pela UCSC (*University of California Santa Cruz*), professor Herbert Lee (*Applied Mathematics and Statistics*). Curso concluído em junho de 2019.

O conjunto dos eventos A_r pode ser considerado um conjunto de hipóteses, sendo que apenas uma delas é verdadeira. A observação do evento B muda a probabilidade inicial $P(A_r)$ para a probabilidade posterior $P(A_r|B)$. Todas as probabilidades posteriores somadas têm valor 1, dado que apenas uma das hipóteses é verdadeira. O denominador $P(B) = \sum_r P(B|A_r).P(A_r)$ pode ser visto como a média ponderada das probabilidades $P(B|A_r)$, onde o peso é dado pelas probabilidades $P(A_r)$. A ocorrência de B aumentará a probabilidade de A_r quando a probabilidade $P(B|A_r)$ for maior que a média ponderada das probabilidades $P(B|A_r)$. E o que significa isso tudo em termos de hipóteses? Significa que a probabilidade da hipótese A_r que mais aumentará com a ocorrência de B é a que tem o maior valor de $P(B|A_r)$.

A probabilidade $P(B|A_r)$ é chamada de *verossimilhança*.[33] No caso, ela representa a verossimilhança de A_r por B, ou a "verossimilhança de A_r dado B". Temos então que o teorema de Bayes combina duas fontes de informação: a informação inicial representada pelas probabilidades "a priori" $P(A_r)$ e a nova informação representada pelas verossimilhanças $P(B|A_r)$, resultando na probabilidade posterior $P(A_r|B)$ que depende dessas duas fontes. Podemos voltar a pensar em como o teorema de Bayes está relacionado com diferentes hipóteses e informação:

> O Teorema de Bayes está em completo acordo com o raciocínio natural. As probabilidades posteriores das várias hipóteses são proporcionais aos produtos das suas probabilidades iniciais e das suas verossimilhanças. (O'HAGAN, FORSTER, 2004, p. 3, tradução nossa).

A noção de verossimilhança não é intuitiva: verossimilhança é a característica daquilo que é verossímil, que aparenta ser ou é tido como verdadeiro, ou semelhante à verdade. Se já parece pouco natural medir a veracidade ou não de alguma coisa, medir o grau de algo ser verossimilhante parece ainda mais complicado.

A verossimilhança não é uma probabilidade, mas é proporcional a uma probabilidade. A verossimilhança de uma hipótese H dados alguns dados D é proporcional à probabilidade de se obter D, dado que H seja verdadeira, multiplicado por uma constante positiva arbitrária K. Em outras palavras, se L representa a verossimilhança (*likelihood*), então $L(H \mid D) = K \cdot P(D \mid H)$. Uma vez que a verossimilhança não é realmente uma

[33] A teoria frequentista faz uso de um *estimador de máxima verossimilhança*, mencionaremos este conceito em um exemplo de inferência frequentista logo adiante, na seção 3.4.1.

probabilidade, ela não obedece necessariamente às leis da probabilidade. Por exemplo, a verossimilhança total não precisa somar 1.

Uma diferença crítica entre verossimilhança e probabilidade está na interpretação do que é fixo e do que pode variar. No caso da probabilidade condicional, $P\ (D \mid H)$, a hipótese é fixa e os dados podem variar. A verossimilhança, no entanto, é o oposto. A verossimilhança de uma hipótese, $L\ (H \mid D)$ condiciona os dados como se fossem fixos, mas permite que as hipóteses variem.

Autores como Edwards (1992) argumentam que a base axiomática apropriada para inferência indutiva não é a de probabilidade, com seu axioma de adição, mas a de probabilidade, conceito introduzido por Ronald Fisher como uma medida de suporte relativo entre diferentes hipóteses[34]. Edwards (1992, p. 30) define o Axioma da Verossimilhança como uma combinação natural da Lei da Verossimilhança e o Princípio da Verossimilhança. A Lei da Verossimilhança afirma que "no âmbito de um modelo estatístico, um determinado conjunto de dados apoia uma hipótese estatística melhor do que outra se a probabilidade da primeira hipótese, nos dados, exceder a probabilidade da segunda hipótese".

Em outras palavras, há evidência para H_1 com relação a H_2 se, e somente se, a probabilidade dos dados em H_1 for maior do que a probabilidade dos dados em H_2. Ou seja, D é evidência de H_1 sobre H_2 se $\mathrm{P}(D|H_1) > \mathrm{P}(D|H_2)$. Se essas duas probabilidades são equivalentes, então não há evidência para nenhuma das hipóteses sobre a outra. Além disso, a força da evidência estatística para H_1 sobre H_2 é quantificada pela razão de suas medidas de verossimilhança: $L\ (H_1|D)/L(H_2\ |D)$ (que novamente é proporcional a $P\ (D|H_1)/P(D|H_2)$). Esse quociente é bastante análogo à noção de *peso da evidência* proposto por Good (1985a), que avaliaremos na seção 4.2.

Já o Princípio da Verossimilhança afirma que a função de verossimilhança contém todas as informações relevantes para a avaliação da evidência estatística. Outras facetas dos dados que não estão presentes na função de verossimilhança são irrelevantes para a avaliação da força da evidência estatística. De toda forma, a noção de verossimilhança tem uma forte conexão com o conceito de evidência, e é usada de forma essencial na teoria da decisão.

Quando mencionamos o teorema de Bayes, falamos até agora das diferentes hipóteses A_r, porém, geralmente as questões estatísticas tratam, ao invés de hipóteses, de uma determinada quantidade. Quando a quantidade é desconhecida e queremos estimar o seu valor,

[34] Cf. Stigler (2007).

fazemos uma inferência estatística e ao invés de hipóteses temos um *parâmetro*, que nada mais é do que a quantidade desconhecida que se quer estimar. O parâmetro é representado por θ.[35]

A inferência estatística preocupa-se com quais afirmações podem ser feitas a respeito do parâmetro considerando os dados que foram coletados. Quando temos um parâmetro θ e dados x, como variáveis aleatórias, assumimos que a sua distribuição possui uma função densidade de probabilidade, e reescrevemos o teorema de Bayes da seguinte forma:

$$f(\theta|x) = \frac{f(\theta).f(x|\theta)}{f(x)}$$
$$= \frac{f(\theta).f(x|\theta)}{\int f(\theta).f(x|\theta)d\theta}$$

A função acima é aplicada para funções contínuas, já no caso de dados discretos, a integral do denominador é substituída por uma soma. A interpretação do teorema de Bayes é mantida da mesma forma: a densidade posterior $f(\theta|x)$ é o resultado da combinação de duas informações, a densidade inicial $f(\theta)$ multiplicada pela verossimilhança $f(x|\theta)$.

Finalmente, podemos fazer a inferência estatística:

> Tendo obtido a densidade posterior $f(x|\theta)$, o passo final para o método Bayesiano é derivar a partir dela algumas afirmações inferenciais adequadas. A questão inferencial mais usual é: após ver os dados x, o que nós sabemos a respeito do parâmetro θ? A única resposta para esta questão é apresentar toda a distribuição posterior. (O'HAGAN, FORSTER, 2004, p. 5, tradução nossa).

A diferença entre a estatística bayesiana e a frequentista (ou clássica) é, então, justamente a apresentação da distribuição posterior:

> A inferência básica bayesiana a respeito de uma hipótese é a sua probabilidade posterior, em contraste com a inferência clássica na qual as hipóteses são ou rejeitadas ou aceitas (ou 'não rejeitadas'). Aceitar ou rejeitar uma hipótese não é tanto uma inferência quanto uma ação ou uma *decisão*. A probabilidade posterior será um fator para tomar a decisão, mas nós também desejaremos levar em consideração as consequências das decisões corretas ou incorretas.

[35] As letras gregas são utilizadas para quantidades desconhecidas, as letras romanas maiúsculas para as variáveis aleatórias e as letras romanas minúsculas para os valores possíveis que as variáveis aleatórias podem assumir.

> (O'HAGAN, FORSTER, 2004, p. 6, grifos do autor, tradução nossa).

Como já mencionamos anteriormente, neste capítulo só podemos cobrir alguns conceitos essenciais do bayesianismo; em virtude da amplitude do tema precisamos nos limitar às ideias mais gerais e apresentar exemplos pontuais. Gostaríamos de mencionar mais alguns aspectos relevantes do bayesianismo.

O primeiro deles é que o bayesianismo aceita outras interpretações da probabilidade: a interpretação subjetiva e a interpretação lógica. A semelhança dessas duas interpretações é que a probabilidade é considerada um grau de crença em uma determinada proposição, e o grau de crença é formado de acordo com as informações disponíveis. Apesar de ambas as interpretações serem aceitas pelo bayesianismo, a interpretação lógica da probabilidade não tem uma aplicação prática, então a grande maioria dos bayesianos adota a interpretação subjetiva. Para os subjetivistas, todas as distribuições de probabilidade que mencionamos são subjetivas: $f(\theta)$, $f(x|\theta)$ e $f(\theta|x)$. Resumidamente, a estatística bayesiana pode fazer uso de probabilidades subjetivas, lógicas ou objetivas. Não é o uso da probabilidade subjetiva que define a estatística bayesiana, mas ela é uma estatística que permite um uso mais amplo de diferentes probabilidades, calculadas a partir de diferentes interpretações.

Uma maneira simples e clara de definir a probabilidade subjetiva é pelo método de aposta coerente. Isso significa que uma pessoa pode fazer uma aposta em um determinado evento: se o evento ocorrer, ela ganha uma certa quantia, se o evento não ocorrer, ela perde uma determinada quantia. A aposta é considerada coerente quando o retorno esperado, que é a soma dessas quantias multiplicado pela probabilidade dos eventos que ocorrem, é igual a zero. Ou seja, a pessoa não pode fazer uma aposta contra ela mesma, na qual ela tem maiores chances de perder a aposta, e também a aposta não pode ocorrer quando a chance de ganhar é maior que a chance de perder, caso contrário a aposta não se realiza.

Suponha que uma pessoa queira apostar se vai ou não vai chover no dia seguinte. Esta aposta pode ser representada de duas formas, uma inversa à outra, e ambas devem ter o retorno esperado igual a zero:[36]

[36] Exemplo retirado de Coursera (2019).

	vai chover	não vai chover
aposta 1	4	-1
aposta 2	-4	1

Quadro 1 - Aposta coerente

Em termos de chances (*odds*), a aposta é representada pela chance 4:1. Isso significa que uma pessoa aposta quatro unidades monetárias na possibilidade de chover no dia seguinte, e aposta menos uma unidade na possibilidade de não chover. As chances são apenas uma forma diferente de representar a probabilidade. Se um evento B tem uma chance $a{:}\,b$ de ocorrer, a probabilidade $P(B)$ é igual a $a/(a+b)$. Assim podemos calcular a probabilidade de não chover, $P(B) = 4/(4+1) = 4/5$, e a probabilidade de chover, que é complementar, $P(A) = 1/5$.

As duas formas de aposta têm retorno esperado igual a zero:

Retorno esperado da aposta 1: $a.P(A) - b.P(B) = 4(1/5) - 1(4/5) = 0$
Retorno esperado da aposta 2: $-a.P(A) + b.P(B) = -4(1/5) + 1(4/5) = 0$

3.4.1 Exemplo comparativo: inferência frequentista e inferência bayesiana

Para ilustrar como a inferência bayesiana é feita, vamos apresentar um exemplo de uma variável com distribuição binomial, e comparar as inferências que podem ser feitas com o paradigma frequentista e em seguida com o paradigma bayesiano.[37]

Dizemos que uma variável aleatória X tem distribuição binomial, com número de lançamentos n, número de sucessos x e probabilidade p de ocorrer cada sucesso:

$$X \sim Bin(n, p)$$

Seguimos com a sua função densidade de probabilidade,

$$P(X = x|p) = f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}$$

[37] Exemplo retirado de Coursera (2019).

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!} \text{ para todo } x \in \{0,1,\dots,n\}$$

E temos a expectativa[38] e variância[39] da distribuição,

$$E[X] = np$$
$$Var(X) = np(1-p)$$

Suponha que alguém vá participar de um jogo que envolva o lançamento de uma moeda e algum tipo de aposta que dependa do resultado. Antes de participar do jogo, existe a possibilidade de fazer 5 lançamentos da moeda e, a partir do resultado dos lançamentos, avaliar se a moeda é normal (equilibrada) ou enviesada. A moeda apresenta cara 2 vezes (sucesso) e coroa 3 vezes (fracasso). Sabemos que uma moeda normal tem probabilidade de 0,5 de apresentar a face cara (ou coroa), e que se ela for enviesada tem probabilidade de apresentar cara 0,7 e coroa 0,3. O que a inferência frequentista pode nos dizer a respeito da moeda? Ela é normal ou enviesada?

Nossa dúvida diz respeito ao parâmetro θ, e seguimos com os cálculos da inferência frequentista:

Parâmetro: $\theta = \{\text{normal}, \text{enviesada}\}$

Distribuição: $X \sim Bin(5, ?)$

Função: $f(x|\theta) = \begin{cases} \binom{5}{x}(0{,}5)^5, & \text{se } \theta = \text{normal} \\ \binom{5}{x}(0{,}7)^x(0{,}3)^{5-x}, & \text{se } \theta = \text{enviesada} \end{cases}$

Se $X = 2$, então $f(\theta|X=2) = \begin{cases} 0{,}3125, & \text{se } \theta = \text{normal} \\ 0{,}1323, & \text{se } \theta = \text{enviesada} \end{cases}$

EMV $\hat{\theta} = \text{normal}$.

O estimador de máxima verossimilhança (EMV) será o valor mais alto, que maximiza a função. Ou seja, de acordo com os dados, conclui-se que a moeda é normal. O resultado frequentista, no entanto, só permite duas possibilidades quando perguntamos qual o

[38] A expectativa é a média ponderada de todos os valores que a variável X pode tomar. Se X é uma variável discreta, $E(X) = \sum_x x.P(X=x) = \sum_x x.f(x)$. Se X for contínua, $E(X) = \int_{-\infty}^{+\infty} x.f(x)dx$.

[39] A variância é o quanto os valores estão "espalhados", é o valor esperado do desvio ao quadrado de X em relação à sua média. Tanto para casos discretos quanto contínuos, a variância pode ser calculada da seguinte forma: $Var(X) = E[X^2] - (E[X])^2$.

grau de certeza de que a moeda é normal ou enviesada. A probabilidade de a moeda ser normal, dado que apresentou a face cara duas vezes é igual a probabilidade de a moeda ser normal, e essa probabilidade é igual a zero ou um:

$$P(\theta = \text{normal}|X = 2) = P(\theta = \text{normal}) \in \{0,1\}$$

A inferência bayesiana tem uma abordagem diferente para trabalhar com o mesmo problema que foi colocado. Uma diferença relevante é que a inferência bayesiana pode levar em consideração uma probabilidade inicial subjetiva, no caso de possuirmos alguma informação antes de analisar os dados. Se existe algum motivo para suspeitar da pessoa que trouxe a moeda para o jogo, por exemplo, se há alguma desconfiança em relação à honestidade dessa pessoa, podemos atribuir uma probabilidade maior da moeda apresentar o lado cara, por exemplo, com maior frequência, ou seja, há desconfiança de que a moeda seja enviesada. Fazendo uma aposta coerente, suponha que se tenha chegado ao valor $P(\text{enviesada}) = 0{,}6$. Vejamos agora como prosseguir com o cálculo da probabilidade posterior neste caso:

Probabilidade inicial: $P(\text{enviesada}) = 0{,}6$

Função:

$$f(\theta|x) = \frac{f(\theta)f(x|\theta)}{\sum_\theta f(\theta)f(x|\theta)}$$

$$= \frac{\binom{5}{x}\left[(0{,}5)^5(0{,}4)I_{\{\theta=\text{normal}\}} + (0{,}7)^x(0{,}3)^{5-x}(0{,}6)I_{\{\theta=\text{enviesada}\}}\right]}{\binom{5}{x}\left[(0{,}5)^5(0{,}4) + (0{,}7)^x(0{,}3)^{5-x}(0{,}6)\right]}$$

$$f(\theta|X = 2) = \frac{0{,}0125I_{\{\theta=\text{normal}\}} + 0{,}0079I_{\{\theta=\text{enviesada}\}}}{0{,}0125 + 0{,}0079}$$

$$= 0{,}612I_{\{\theta=\text{normal}\}} + 0{,}388I_{\{\theta=\text{enviesada}\}}$$

A estatística bayesiana chega desta forma à probabilidade posterior $P(\theta = \text{enviesada}|X = 2)$:

$$P(\theta = \text{enviesada}|X = 2) = 0{,}388$$

Ou seja, podemos dizer que a probabilidade de a moeda ser enviesada, dado que lançamos a moeda 5 vezes e ela apresentou a face cara 2 vezes, é de 0,388. O resultado seria diferente dependendo da escolha da probabilidade inicial $P(\theta)$. Se fôssemos escolher uma probabilidade inicial assumindo que não temos informação alguma para achar que a moeda é enviesada ou não, atribuiríamos a probabilidade $P(\text{enviesada}) = 0{,}5$ (essa probabilidade pode até ser considerada uma probabilidade objetiva ou invés de subjetiva). A probabilidade posterior seria então $P(\theta = \text{enviesada}|X = 2) = 0{,}297$. De qualquer forma, os resultados bayesianos são diferentes da abordagem frequentista.

O nosso exemplo foi um exemplo simples, de uma variável aleatória discreta. No entanto, a maioria dos casos é de variáveis aleatórias contínuas, e outros casos podem ser muito mais complexos. No entanto, não cabe, neste trabalho, entrar em maiores detalhes estatísticos, porque nosso enfoque é mais filosófico. Em parte, este capítulo, inclusive, foi escrito com o objetivo de introduzir alguns conceitos necessários para a compreensão dos próximos capítulos.

Para finalizá-lo, gostaríamos de ressaltar que a diferença do método de inferência, frequentista ou bayesiano, não é uma questão de mera escolha, sem grandes consequências. Para exemplificar[40], podemos citar uma pesquisa que foi feita para testar a eficácia de uma vacina que tinha como objetivo a imunização contra o HIV.[41] O resultado dessa pesquisa, com o método frequentista, foi que a vacina era eficaz em 31,2% dos casos, com o $p - valor$ de 4%, ou seja, abaixo do nível de significância de 5%, o valor mais comumente utilizado para decisão em relação à aceitação ou rejeição da hipótese nula. Um estudo bayesiano, utilizando os mesmos dados, chegou à conclusão de que a vacina tinha uma probabilidade apenas de 22% de eficácia.[42] Ou seja, são resultados bem diferentes, com os mesmos dados. Esse exemplo revela a importância de haver debate em relação aos métodos de inferência utilizados, de discutir os métodos científicos.

[40] Exemplo retirado do curso online *Statistics in medicine (Stanford Online)*.
[41] RERKS-NGARM, S.; PITISUTTITHUM, P.; NITAYAPHAN, S.; et al. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N Engl J Med*. 2009 Dec 3;361(23):2209-20.
[42] GILBERT, P. B.; BERGER, J. O.; STABLEIN, D.; et al. Statistical interpretation of the RV144 HIV vaccine efficacy trial in Thailand: a case study for statistical issues in efficacy trials. *J Infect Dis*. 2011 Apr 1;203(7):969-75.

4. PESO DA EVIDÊNCIA E SUPORTE PROBABILÍSTICO

> O conceito de peso da evidência, embora não com esse nome, era básico na abordagem de Turing à criptoanálise do Enigma (...). Prevejo que em cinquenta anos a partir de agora [em 2034], se a civilização for restaurada, $W(H:E)$ será aceito pela maioria dos filósofos como a melhor explicação de longe para peso da evidência, e que após mais cinquenta anos [em 2084] terá uma grande influência na educação elementar, no diagnóstico médico diferencial, e no direito comum. (GOOD, 1984a, p. 163).

4.1 Introdução

Se fôssemos ilustrar o *peso da evidência* a favor ou contra uma hipótese, poderíamos pensar na deusa grega da justiça Themis, que, representada geralmente carregando uma balança, pesa as evidências a favor ou contra um argumento. Assim, o lado (ou o argumento) mais pesado, o que tiver maior evidência, é escolhido em detrimento do outro lado, do outro argumento, ou de outra alternativa que costuma ser o seu oposto. O mesmo processo de pesar as evidências ocorre num tribunal de júri que decidirá se as evidências pesam mais a favor ou contra a inocência de uma pessoa acusada de algum delito; com um detetive, que decidirá se deve ou não levar um caso à lei; com um médico, que pesará as evidências ao fazer um diagnóstico diferencial entre duas doenças possíveis para escolher um tratamento adequado ao paciente.[43]

O conceito de *peso da evidência* que apresentamos nesta Dissertação está diretamente relacionado não só à probabilidade, mas também à estatística. E, finalmente, o *peso da evidência* também apresenta interesse filosófico, pelo fato de se tratar de uma medida de evidência que equivale, segundo seus defensores, ao grau de corroboração de uma hipótese pela evidência. Desta forma, tem um interesse particular para a filosofia da ciência.

O *peso da evidência* é a resposta de Good (1987) para duas questões propostas pela linha da filosofia da ciência de Popper: ele é tanto uma medida de corroboração de uma hipótese pela evidência como também uma forma de suporte probabilístico indutivo. Nossa Pesquisa foi

[43] Nesta Pesquisa apresentamos um exemplo didático de *peso da evidência* na seção 4.2.1 e, no próximo capítulo, o exemplo da primeira aplicação prática de *peso da evidência* na quebra da criptografia da máquina Enigma. Para discussões a respeito da aplicação do conceito na área do direito, consultar Good (1986). Para o caso da aplicação do conceito ao diagnóstico diferencial, na área da medicina, consultar Good e Card (1971), Card e Good (1974) e Spiegelhalter e Knill-Jones (1984).

iniciada com o artigo de Popper e Miller (1983), artigo no qual os autores apresentam uma prova de que o suporte probabilístico não pode ser indutivo, e sim dedutivo. A partir do primeiro comentário de Good (1984b) a respeito do artigo de Popper e Miller, estudamos os argumentos do autor, que são baseados no seu próprio ponto de vista bayesiano e discordam da visão de Popper e Miller.

A proposta alternativa de Good não é apenas uma prova algébrica. O *peso da evidência* teve uma aplicação prática: a fórmula foi aplicada num procedimento que visava a quebra da criptografia da máquina Enigma na Segunda Guerra Mundial. Neste capítulo, vamos apresentar a argumentação de Good (1960, 1968, 1975, 1979, 1984b, 1984c, 1985a, 1985b, 1987, 1988), e, no próximo capítulo, vamos descrever a aplicação prática do *peso da evidência* na quebra da criptografia.

Apesar de termos iniciado nossa Pesquisa com o artigo de Popper e Miller, de 1983, foi necessário analisar uma série de publicações de Good muito anteriores a esta data. Na sequência de artigos de Good que continuam o debate, um deles, de 1987, discute outras medidas de suporte probabilístico. Uma delas é a de Gilles (1985), que faz uma defesa do argumento de Popper e Miller, e a outra de Redhead (1985), que tenta refutá-los. É neste momento que Good apresenta o conceito de *peso da evidência* como a sua proposta de medida de suporte probabilístico, refutando tanto Gilles quanto Redhead. A menção a este novo conceito de suporte probabilístico nos levou a seguir as referências de Good, migrando de um artigo a outro, retroativamente, até chegarmos à sua primeira publicação, de 1950.

Foi necessário fazer um trabalho de pesquisa mais extenso do que o previsto, a fim de selecionar não só as publicações de Good mais relevantes ao tema inicial "suporte probabilístico indutivo", mas também para definir, desenvolver e delimitar o novo tópico que nos foi apresentado: o *peso da evidência*. Ampliamos nosso escopo, sem programação prévia, partindo de uma pesquisa da relação dos fundamentos da probabilidade com seu o suporte dedutivo ou indutivo, para um problema aplicado, que envolvia o trabalho estatístico de Turing, na Segunda Guerra Mundial.

Ademais, adentramos em um debate filosófico importante, senão central, do século XX: a tentativa de definir uma medida de grau de corroboração para uma teoria científica ser testada. Este é um dos principais objetivos de Popper com a publicação da "Lógica da pesquisa científica", publicada em inglês, pela primeira vez, em 1959, e escrita originalmente em alemão em 1934. Vejamos como Popper expõe, de forma resumida, o problema do grau de corroboração no contexto da elaboração de seu livro (no apêndice ix, intitulado "Corroboração, peso de evidência e testes estatísticos"):

> Antes mesmo de publicado meu livro, senti que o problema do grau de corroboração era um desses problemas que requer estudo mais aprofundado. Por 'problema do grau de corroboração' pretendo significar o problema (i) de mostrar que existe medida (a ser denominada grau de corroboração) da severidade dos testes a que a teoria foi submetida e da maneira pela qual ela ultrapassou esses testes ou falhou diante deles; e (ii) de mostrar que *essa medida não pode ser uma probabilidade* ou, mais precisamente, que ela não satisfaz as leis formais do cálculo de probabilidades.
>
> Um esboço de solução de ambas essas questões – especialmente da segunda – estava contida em meu livro. Creio, porém, que algo deva ser acrescentado. Não bastava mostrar a insuficiência das teorias de probabilidade existentes – a de Keynes ou de Jeffreys, por exemplo (...). Fazia-se necessário um tratamento da questão em termos gerais. Busquei, por isso mesmo, elaborar um cálculo formal de probabilidades que admitisse uma interpretação em vários sentidos (...)
>
> O objetivo último consistia, naturalmente, em mostrar que o *grau de corroboração não era uma probabilidade*, equivalendo isso a dizer que *não era uma das possíveis interpretações do cálculo de probabilidades*. Sem embargo, dei-me conta de que a tarefa de elaborar um cálculo formal não era só necessária para esse propósito, como era, por si mesma, interessante. (POPPER, 1972, p. 443-444, grifos do autor).

Ou seja, em linhas gerais, Popper dedica-se a elaborar um cálculo formal de probabilidades para mostrar que o grau de corroboração não pode ser medido por uma probabilidade, qualquer que fosse o tipo de interpretação dessa probabilidade. Good concorda com essa afirmação de Popper, mas discorda em relação ao grau de corroboração. Essa é uma das questões que veremos ao longo deste capítulo.

Em muitos dos artigos de Good que selecionamos para esta Dissertação, podemos afirmar que há uma tentativa de dialogar e debater o tema central (grau de corroboração) da maneira que o problema foi apresentado por Popper. Good cita Popper com frequência em seus artigos, faz comentários de artigos de Popper e seus colaboradores (como Miller), e também publica vários artigos como resposta, troca correspondência, enfim, dialoga de muitas formas, constantemente.

Popper define com precisão no que consiste o problema do grau de corroboração:

> O objetivo desta nota é o de propor e discutir uma definição – em termos de probabilidade – do *grau em que um enunciado x é confirmado por um enunciado y*. (Claro que se trata de algo

> equivalente ao *grau em que um y confirma um enunciado x*). Denotarei esse grau pelo símbolo '$C(x,y)$', a ser lido '*grau de confirmação de x por y*'. Em casos particulares, *x* poderá ser uma hipótese, *h*; e *y* poderá ser alguma evidência empírica, *e*, favorável a *h*, ou desfavorável a *h*, ou neutra em relação a *h*. Contudo, '$C(x,y)$' será também aplicável a casos menos típicos. (POPPER, 1972, p. 452, grifos do autor).

Notamos que Popper faz uso do termo "grau de confirmação" no contexto do artigo *Degree of confirmation* (1954). O trecho que acabamos de citar corresponde ao primeiro parágrafo do artigo de 1954. Porém, alguns anos mais tarde, já na introdução do apêndice ix, ele utiliza o termo "grau de corroboração", com o mesmo sentido:

> (...) para definir $C(x,y)$ – o grau de corroboração da teoria *x* pela evidência *y* – tinha que operar com alguma inversa $p(x,y)$, que Fisher denominou '*verossimilhança* de *x*' (à luz da evidência *y* ou dado *y*). (Note-se que tanto minha 'corroboração' como a 'verossimilhança' de Fisher pretendem medir a aceitabilidade da hipótese *x*; *x* é, pois, o que importa, enquanto *y* simplesmente representa a evidência empírica variável ou, como prefiro dizer, os relatórios dos *resultados dos testes*). (POPPER, 1972, p. 444, grifos do autor).

No presente capítulo, a discussão está centrada no problema do grau de corroboração – ou seja, o grau de corroboração de *x* por *y*, ou o grau de corroboração de uma hipótese *h* pela evidência *e*.

Antes de começar, há mais uma observação importante a ser feita. Existe uma polarização na filosofia da ciência que divide os filósofos da ciência em duas linhas: os que seguem uma linha "indutivista", que tem Carnap como um de seus maiores representantes, e os que seguem a linha "dedutivista", representada principalmente por Popper. Essa polarização não constitui obrigatoriamente um dualismo. É plenamente possível concordar parcialmente com Popper e Carnap em determinadas questões, como faz Good. Porém, a polarização é um fator relevante de divisão entre as diferentes linhas: a própria discordância entre Popper e Miller e Good em relação à natureza do suporte probabilístico é um exemplo.

Um dos mais importantes artigos de Good que utilizamos neste capítulo (publicado em 1975)[44], é fruto de uma conferência que teve como tema as metodologias bayesianas e

[44] GOOD, I. J. Explicativity, corroboration, and the relative odds of hypotheses. *Synthese*, Vol. 30, No. 1/2, Methodologies: Bayesian and Popperian (Feb. - Mar., 1975), p. 39-73.

popperianas. A revista *Synthese*, que publica artigos nas áreas de epistemologia, metodologia e filosofia da ciência, lançou uma publicação exclusiva, em 1975, com artigos oriundos das discussões ocorridas em 1973, nesta conferência:

> Os artigos de Joseph Agassi, I. J. Good, Richard Jeffrey e David Miller, (...) constituem os eventos da conferência 'Metodologias: bayesianas e popperianas', realizados na Universidade da Carolina do Sul de 8-10 de novembro.
>
> Ao menos, o material coletado aqui vai um pouco em direção à redefinição de alguns e ao melhoramento de outros assuntos que dividiram bayesianos e popperianos quando Carnap e Popper se enfrentaram em *Bedford College* em 1965. (GOODE et al., 1975, p. 1, tradução nossa).

Seguimos, agora, com a descrição da expressão *peso da evidência*, a partir dos escritos de Good. Não fizemos um estudo da história do termo em si, mas sim um estudo de como, de acordo com a compreensão de Good, o *peso da evidência* foi desenvolvido, de como ele foi debatido historicamente com filósofos e estatísticos, dentro do contexto de polarização entre visões de suporte probabilístico indutivo e dedutivo. Finalmente apresentamos também a visão de Popper e Miller (1983), Fitelson (2001) e Gilles (1990) e uma breve cronologia de outros artigos que se seguiram ao debate, caso o leitor tenha interesse em conhecer outras linhas que não as discutidas aqui em maiores detalhes.

4.2 Peso da evidência: suporte probabilístico indutivo de acordo com Good

Em 1983, Popper e Miller publicaram uma carta na revista *Nature* com uma prova do suporte probabilístico dedutivo de uma evidência *e* em relação a uma hipótese *h*. Os autores negaram a possibilidade de um suporte probabilístico indutivo; o próprio título da carta já revela a intenção dos autores: *A proof of the impossibility of inductive probability*. A carta é seguida de três comentários, na mesma edição: Jeffrey (1984), Levi (1984), e Good (1984b). Levi, que segue a linha de interpretação propensista da probabilidade, assim como Popper, concorda com a prova dos autores. Já Good nunca concordou com a prova de Popper e Miller. Good escreve uma série de artigos com várias objeções à prova da impossibilidade da probabilidade indutiva. Para ele, a probabilidade é sim indutiva. O principal argumento de Good contra a prova de Popper e Miller é, na verdade, uma proposta alternativa de como seria a relação entre evidência e hipótese. O que o autor define como *peso da evidência* é justamente essa relação.

Num breve resumo que reúne as contribuições de Good para a probabilidade, estatística e filosofia da ciência, Lindley (1990, p. 216) observa o quanto esse conceito foi estudado por Good (que àquela época já somava cerca de 1600 publicações): "O peso da evidência, e conceitos relacionados, foram muito estudados nos seus escritos". Em seguida, o autor cita a primeira frase que Good escreve em um dos seus artigos, frase que não deixa dúvida alguma a respeito da importância do *peso da evidência* para o autor (aqui reproduzimos o trecho um pouco mais amplo): "Meu propósito é pesquisar um pouco do trabalho em peso da evidência porque eu penso que o tópico é quase tão importante quanto o da probabilidade ela mesma. A pesquisa ficará longe de ser completa." (GOOD, 1985a, p. 249).

Antes de Good, o conceito de *peso da evidência* já havia sido definido por Peirce (1878), Gibbs (1902), Keynes (1921), Jeffreys (1936), mas geralmente o conceito tinha algumas diferenças em relação ao apresentado por Good. Apesar do termo ser parecido entre os autores citados, fazemos a ressalva de que há bastante diferença entre eles.[45] Good relata, por exemplo, que Levi afirmou que "Peirce era anti-Bayesiano e até um certo ponto antecipou Neyman e Pearson" (GOOD, 1985a, p. 253). Ou seja, apesar da semelhança entre os conceitos, Pierce distanciava-se consideravelmente do bayesianismo.

Seguimos então apresentando a versão de Good (1985a) do *peso da evidência*, que é denotado por:

$$W(H : E)$$

Lemos esta definição de *peso da evidência* como o "peso da evidência a favor de H [hipótese] fornecido por E [evidência]" (GOOD, 1985a, p. 250). Suponha que tenhamos duas hipóteses simples e concorrentes, H e $\overline{H}$. $W(H : E)$ será uma função de $P(E|H)$, a probabilidade da evidência dada a hipótese, e de $P(E|\overline{H})$, a probabilidade da evidência dada a hipótese concorrente, ou seja:

$$W(H : E) = f\big(P(E|H), P(E|\overline{H})\big)$$

[45] Em virtude da limitação do tema do presente trabalho, apresentaremos apenas a versão do conceito de *peso da evidência* desenvolvido por Good. Pode ser encontrada uma história do conceito em Good (1985a), o artigo mais abrangente entre os pesquisados. Para a versão de Peirce, pode-se consultar, além do próprio autor, Kasser (2016); para a versão de Keynes, consultar o próprio autor, Keynes (1921, p. 71), o capítulo "*Weight of Evidence*" em Levi (1967, p. 139-152) e Brady (1993).

A probabilidade inicial ou *a priori* da hipótese H, $P(H)$, é independente da função *f*. Já a probabilidade final ou *a posteriori* da hipótese H relativamente a E, $P(H|E)$, depende de dois fatores, *peso da evidência* e a probabilidade inicial:

$$P(H|E) = g(W(H : E), P(H))$$

Temos então a seguinte igualdade:

$$P(H|E) = g(f(P(E|H), P(E|\overline{H})], P(H))$$

De acordo com o Teorema de Bayes, as probabilidades $P(E|H)$ e $P(E|\overline{H})$ são equivalentes às seguintes igualdades:

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

$$P\left(E\middle|\overline{H}\right) = \frac{P\left(\overline{H}\middle|E\right).P(E)}{P(\overline{H})}$$

Agora, considerando as igualdades $P(H) = x$, $P(E) = y$ e $P(H|E) = z$, e que $P(\overline{H}) = 1 - P(H)$, $P(\overline{H}|E) = 1 - P(H|E)$, temos a igualdade:

$$z = g\left(f\left(\frac{zy}{x}, \frac{y(1-z)}{1-x}\right), x\right)$$

De acordo com Good (1985a, p. 251), podemos deduzir a partir desta equação que a função *f* é uma função monotônica crescente de $P(E|H)/P(E|\overline{H})$.[46] Isso ocorre porque a função f não depende diretamente do valor $y = P(E)$, já que $W(H : E) = f\left(P(E|H), P(E|\overline{H})\right)$. E a função *g* depende mais de *f* do que de $x = P(H)$. Portanto, é razoável supor que a função *g* depende da razão $P(E|H)/P(E|\overline{H})$. A função *f* deve ser monotônica, pois, caso contrário, tomaria o mesmo valor para dois valores diferentes de z, e a equação acima seria impossível — z teria valores distintos e ao mesmo tempo iguais, o que seria absurdo. Desta forma, *f* é uma função monotônica e $W(H : E)$ é uma razão bayesiana de verossimilhança.

[46] Os detalhes desta dedução encontram-se em Good (1968, p. 141) e Good (1984c, p. 298).

Podemos ter na definição de *peso da evidência* a propriedade aditiva. Ao invés de $W(H:E)$, podemos calcular $W(H:E\cap E')$:

$$W(H:E\cap E')=W(H:E)+W(H:E'|E),$$

Ou seja: o *peso da evidência* a favor da hipótese *H* fornecido pelas evidências $E\cap E'$ é igual a soma do *peso da evidência* fornecido por *E* com o *peso da evidência* fornecido por E' dado *E*. Good (1984c, p. 297) destaca que esta equação pode incluir qualquer outra informação condicional *G* dada anteriormente:[47]

$$W(H:E\cap E'|G)=W(H:E|G)+W(H:E'|E\cap G)$$

Good (1985a, p. 251) afirma que a propriedade de adição pode ser satisfeita se definirmos:

$$W(H:E)=\log\frac{P(E|H)}{P(E|\bar{H})}$$

Com aplicações da regra dos produtos das probabilidades, $P(A\cap B)=P(B).P(A|B)$, chegamos à seguinte igualdade:[48]

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

Sendo que $O(H|E)$ significa chance (*odds*) de *H* dado *E*, e $O(H)$, chance de *H*. O primeiro registro da igualdade entre as probabilidades e chance é encontrado em Wrinch e Jeffreys (1921). O seguinte quadro retirado de Good (1985a, p. 251) ilustra a correspondência entre probabilidades e chance:

[47] Ressaltamos que estamos nos referindo a *H* como hipótese e *E* como evidência, porém tanto *H*, *E*, *E*′, e *G* são proposições que "não necessariamente referem-se a eventos e hipóteses" (GOOD, 1968, p. 126).
[48] Este cálculo de probabilidades encontra-se no Apêndice, item B, "Transformação algébrica: Razão bayesiana de verossimilhança em Fator Bayes-Jeffreys-Turing".

Probabilidade (P)	Chance (*Odds*) ($P/(1-P)$)
0	0
$^1/_{10}$	$^1/_9$
$^1/_3$	$^1/_2$
$^1/_2$	1
$^2/_3$	2
$^9/_{10}$	9
1	∞

Good (1985a, p. 252) faz a seguinte afirmação sobre como o lado direito da fórmula, $O(H|E)/O(H)$, pode ser descrito, mencionando inclusive a posição de Turing a respeito:

> O lado direito da equação (3) pode ser descrito em palavras como a razão das chances [*odds*] finais de *H* sobre as suas chances iniciais, ou a razão das chances posteriores sobre as chances a priori, ou o fator pelo qual as chances iniciais de *H* são multiplicadas para resultar nas chances finais. É, portanto, natural chamá-la de *fator a favor de H fornecido por E* e esse foi o nome dado ao fator por A.M. Turing em uma aplicação vital de análise criptográfica na Segunda Guerra Mundial em 1941. Ele não mencionou o teorema de Bayes, com o qual [o fator] está obviamente relacionado de forma muito próxima, porque ele sempre gostava de resolver todas as questões por ele mesmo. Quando eu disse a ele que o conceito era essencialmente uma aplicação do teorema de Bayes ele disse "Suponho que sim". (GOOD, 1985a, p. 252, grifos do autor, tradução nossa).

Good continua a descrição da fórmula $O(H|E)/O(H)$ em termos bayesianos:

> Na literatura bayesiana corrente [o conceito] é geralmente chamado de *fator de Bayes a favor de H fornecido por E*. Dessa forma o peso da evidência [é formado por conceitos que] são Bayesianos porque as probabilidades $P(H), P(H|E), P(E|H)$ e $P(E|\overline{H})$ são todas em

> geral consideradas sem significado por anti-Bayesianos. (GOOD, 1985b, p. 252, grifos do autor, tradução nossa).[49]

Aqui temos, portanto, um outro conceito que está inserido como fundamento no conceito de *peso da evidência*: o *fator de Bayes*, $O(H|E)/O(H)$. Good sugere uma nova denominação para o que é correntemente identificado como fator de Bayes: ele afirma que esse fator poderia incluir os nomes de Jeffreys (1939) e Turing em virtude da colaboração de ambos para o desenvolvimento do conceito, sendo denominado finalmente como o "fator Bayes-Jeffreys-Turing" (GOOD, 1975, p. 52).

Finalmente, temos a seguinte fórmula final do *peso da evidência*:

$$W(H : E) = \log \frac{O(H|E)}{O(H)}$$

Outra forma de representar a equação do *peso da evidência* pode facilitar a compreensão do conceito. O *peso da evidência* é igual às chances finais (*a posteriori*) menos as chances iniciais (*a priori*):

$$W(H : E) = \log O(H|E) - \log O(H)$$

Ou podemos apresentá-la como as chances iniciais, que, somadas ao *peso da evidência*, são iguais às chances finais:

$$\log O(H) + W(H : E) = \log O(H|E)$$

O *peso da evidência* pode ser positivo ou negativo, então as chances finais podem ser maiores ou menores que as chances iniciais. Essa é a ideia fundamental do *peso da evidência*: quanto a ocorrência de uma evidência "pesa" a favor ou contra uma hipótese, ou quanto a chance de uma hipótese aumenta ou diminuiu após a ocorrência da evidência.

Do ponto de vista cronológico, Turing não estava ciente naquela época de que o logaritmo do fator de Bayes, ou seja, $\log O(H|E)/O(H)$, já tinha sido definido por Peirce em

[49] Aqui adaptamos a tradução, pois nos pareceu haver algum erro no texto de Good. Reproduzimos portanto o trecho original que adaptamos: "Thus weight of evidence are Bayesian concepts because the probabilities $P(H), P(H|E), P(E|H)$ and $P(E|\overline{H})$ are all in general regarded as meaningless by anti-Bayesians." (GOOD, 1985b, p. 252).

1878 (embora não de forma simbólica) e também denominado como *peso da evidência* (GOOD, 1979, p. 394). O nome de Peirce provavelmente não está incluído nessa nova denominação sugerida por Good porque ele afirma que o Peirce "falhou em chegar em uma definição correta do peso da evidência devido a um erro" (GOOD, 1985a, p. 252).[50]

Good e Toulmin (1968) justificam da seguinte maneira a escolha de Turing:

> Turing introduziu os log-fatores (pesos da evidência) assim como os fatores parcialmente devido à aditividade e parcialmente porque como uma consequência da aditividade, os pesos da evidência têm distribuições assintoticamente normais. Portanto é usualmente mais apropriado considerar as esperanças dos pesos da evidência no *design* de experimentos do que as esperanças dos fatores de Bayes. As distribuições log-normais são extremamente assimétricas, então a esperança de um fator de Bayes não é um valor típico. (GOOD; TOULMIN, 1968, p. 95, tradução nossa).[51]

Ou seja, os log-fatores são importantes principalmente pelo fato da sua distribuição estatística ser assintoticamente normal, e as aproximações da curva normal são fundamentais para a inferência estatística.[52] Turing também definiu uma unidade de medida para o *peso da evidência*, e chamou-a de *deciban*. Um *deciban* seria um décimo de uma unidade do logaritmo de base 10. Essa seria uma unidade mínima, no sentido de que uma mudança de um *deciban* (ou até metade dele) seria o mínimo possível perceptível para a intuição humana, algo análogo à unidade do decibel, que é menor unidade de diferença de volume perceptível ao ouvido humano. Vejamos em detalhes os motivos de Turing para essa escolha:

> Turing foi o primeiro a reconhecer o valor de nomear as unidades em termos das quais o peso da evidência é medido. Quando a base do logaritmo era *e* ele denominava a unidade de ban natural, e

[50] Para detalhes do erro, consultar Good (1981).
[51] Para uma definição de distribuições normais e log-normais, consultar DeGroot (2002, p. 302-314).
[52] Para mais esclarecimentos a esse respeito, consultar a seção sobre o teorema do limite central em Wackerly et al. (2008. p. 370-373), além da seção sobre a lei dos grandes números em Rice (1995, p. 163-164). Turing, ao final de sua graduação em matemática na Universidade de Cambridge, fez um trabalho sobre o teorema do limite central, estimulado pelas aulas de metodologia da ciência do astrofísico Sir Arthur Eddington, ministradas em 1933. Turing terminou seu trabalho em fevereiro de 1934, e após uma revisão terminada em novembro de 1934, submeteu seu trabalho como uma *Fellowship Dissertation*, conseguindo assim ser eleito como *Fellow* de Cambridge em março de 1935, aos 22 anos. No mesmo ano, ganhou com o seu trabalho o *Smith Prize*, um prêmio da Universidade de Cambridge dado para trabalhos desenvolvidos em qualquer área da matemática ou de suas aplicações, para o qual podem concorrer recém-formados da universidade tanto da graduação como da pós-graduação. Good ganhou o mesmo prêmio em Cambridge. Para mais detalhes sobre o trabalho de Turing em estatística, consultar Zabell (1995).

> simplesmente de ban quando a base era 10. Foi muito mais tarde que uma unidade de informação para a base 2 foi denominada bit e as mesmas unidades podem ser usadas tanto para informação quanto para peso da evidência. Turing introduziu o nome deciban no sentido auto-explicativo de um décimo de um ban, por analogia com o decibel. A razão para o nome ban foi que dezenas de milhares de folhas eram impressas na cidade de Banbury, nas quais os pesos da evidência eram inseridos em decibans para executar um importante processo sigiloso chamado Banburismus.
>
> Um deciban ou metade de um deciban é em torno da menor mudança em peso da evidência que é diretamente perceptível à intuição humana. Eu sinto que isso é uma importante colaboração para o raciocínio humano e irá melhorar eventualmente o julgamento de médicos, advogados e outros cidadãos.
>
> A principal aplicação do deciban foi para a análise sequencial, não para controle de qualidade, mas para discriminar entre hipóteses, assim como em testes clínicos ou em diagnósticos médicos. (GOOD, 1979, p. 394, tradução nossa).

A primeira publicação com a definição de *peso da evidência* como o logaritmo do fator de Bayes é feita por Good (1950). De forma independente, outros autores chegaram a resultados bastante semelhantes. Minsky e Selfridge (1961) utilizam o mesmo termo, Tribus (1969) usou o termo "evidência" para o *peso da evidência*. Kemeny e Oppenheim (1952) usaram a expressão "suporte factual para uma hipótese" (fornecido por evidências), com uma formulação diferente. Há muitas divergências sobre qual seria a melhor medida para o grau de corroboração da evidência em relação à hipótese, e há um extenso debate sobre o assunto, o que torna a tarefa de cobrir todo o debate contraproducente. No entanto, ao final deste capítulo, apresentamos a versão de Popper e Miller (1983) para o problema, além de algumas discussões a respeito da versão de *peso da evidência* de Turing e Good: Fitelson (2001), Gilles (1990) e uma indicação de outros autores que debateram o assunto.

4.2.1 Exemplo de cálculo de peso da evidência

Vejamos então um exemplo de uma aplicação de *peso da evidência*, no qual podemos ver a discriminação entre hipóteses e o uso da unidade *deciban*. Esse exemplo encontra-se descrito em Good (1985a, p. 254).

Suponha que desejamos, após lançar um dado, inferir se ele é um dado enviesado ou não, de acordo com os resultados de múltiplos lançamentos. O dado enviesado foi feito de forma a apresentar o número 6 com maior frequência. E temos a informação a priori da

probabilidade de o dado apresentar o número 6 e de apresentar os outros números. O dado não enviesado é equilibrado e a probabilidade de apresentar qualquer número é a mesma, equiprovável. Temos, então, duas hipóteses a testar: se o dado é enviesado ou se o dado é normal. Com essas informações podemos calcular o *peso da evidência*

Temos os seguintes eventos (que neste caso podemos definir como hipóteses e evidências) e fórmula:

$H =$ o dado é enviesado

$\overline{H} =$ o dado é normal

$E_1 =$ apresentar número 6

$E_2 =$ apresentar outro número

Peso da evidência:

$$W(H : E) = \log \frac{P(E|H)}{P(E|\overline{H})}$$

Good informa qual é a probabilidade de o dado enviesado apresentar o número 6, e sabemos a probabilidade de um dado normal:

$$P(E_1|H) = 1/3$$
$$P\left(E_1\middle|\overline{H}\right) = 1/6$$

Agora, calculamos o *peso da evidência* quando o dado apresenta o número 6:

$$= \log \frac{P(E_1|H)}{P\left(E_1\middle|\overline{H}\right)}$$
$$= \log \frac{1/3}{1/6}$$
$$= \log 2$$
$$\simeq 0.30$$

Para calcular o *peso da evidência* quando o dado apresenta outro número, utilizamos as seguintes probabilidades:

$$P(E_2|H) = 2/3$$
$$P\left(E_2\middle|\overline{H}\right) = 5/6$$

Temos, por fim, o valor do *peso da evidência* quando o dado apresenta outro número:

$$= \log\frac{P(E_2|H)}{P\left(E_2\middle|\overline{H}\right)}$$
$$= \log\frac{2/3}{5/6}$$
$$= \log 4/5$$
$$\simeq -0.09$$

Os valores dos pesos da evidência aos quais chegamos estão medidos na unidade *ban*, que é o resultado com o logaritmo na base 10. Lembramos que a medida utilizada na prática era o *deciban*, que é o valor decimal do *ban* e é abreviado por *db*. Portanto, temos os seguintes pesos da evidência medidos em *decibans*:

$$W(H : E_1) = 3$$
$$W(H : E_2) = -1$$

Isso significa que, quando lançamos o dado e ele apresenta o número 6, temos 3 unidades *db a favor* da hipótese de que o dado é enviesado, e quando lançamos o dado e ele apresenta outro número, temos 1 unidade *contra* a hipótese de que o dado é enviesado, ou -1 unidade *a favor* da hipótese do dado ser enviesado.

Seguindo com o exemplo de Good, se lançarmos o dado 20 vezes, e se em 10 vezes ele apresentar o número 6, e nas outras 10 vezes apresentar outros números, teremos 20 *db* a favor da hipótese do dado ser enviesado.

4.2.2 Variações da fórmula do peso da evidência

Vimos, até agora, a fórmula geral e mais simples do *peso da evidência*:

$$W(H : E) = \log \frac{O(H|E)}{O(H)}$$

Existem outras formas de representar o *peso da evidência* de acordo com Good. Nesta última fórmula, por exemplo, também é possível adicionar uma probabilidade condicional qualquer G da qual se tenha conhecimento prévio:

$$W(H : E|G) = \log \frac{O(H|E \cdot G)}{O(H|G)}$$

Neste caso, portanto, temos o *peso da evidência* a favor da hipótese H, fornecido pela evidência E, dado o conhecimento prévio[53] de G.

As duas fórmulas do *peso da evidência* que acabamos de ver, $W(H : E)$ e $W(H : E|G)$, são válidas no caso das hipóteses H e $\overline{H}$ serem mutuamente exclusivas. De acordo com Good (1975, p. 52), há uma generalização do *peso da evidência* que inclui hipóteses que não são obrigatoriamente mutuamente exclusivas, H_1 e H_2:

$$W(H_1/H_2 : E\,|G) = \log \frac{(H_1/H_2\,|E \cdot G)}{(H_1/H_2\,|G)}$$

A fórmula acima, significa, portanto, o *peso da evidência* a favor da hipótese H_1 comparada a uma outra hipótese H_2, fornecido pela evidência E, dado o conhecimento prévio G.

4.2.3 Peso da evidência e quantidade de informação

Good apresenta outra forma de representar o conceito de *peso da evidência*: ele o relaciona ao conceito de *quantidade de informação*. Em um artigo publicado em 1960 (*Weight of evidence, corroboration, explanatory power, information and the utility of experiments*), Good afirma que este debate iniciou-se com uma iniciativa de Popper, que, "por motivos

[53] O termo em inglês utilizado nos textos para "conhecimento prévio" é *background knowledge*.

filosóficos (...) discutiu o que poderia se entender por grau de corroboração de uma hipótese fornecido por uma observação ou resultado experimental" (GOOD, 1960, p. 320).

Good afirma que o primeiro artigo publicado por Popper sobre corroboração foi publicado em 1954, porém não informa a referência. No entanto, conseguimos identificar o artigo: *Degree of confirmation*. Popper publica mais dois artigos em sequência a este, em 1957 e 1958, como duas notas complementares que esclarecem os pontos discutidos no primeiro deles. Os três artigos compõem o apêndice ix da principal publicação de Popper, "A lógica da pesquisa científica". Outros autores também debateram o assunto: "A história diretamente relevante [sobre corroboração] pode ser traçada por Good (1950, 1955a, 1955b), Lindley (1956), Popper (1959), Mallows (1959), e a história de diferentes medidas de informação por Kullback (1959, p. 1-2)."[54].

Mas por que estamos citando o histórico da discussão sobre a corroboração de uma hipótese quando nos propomos a apresentar a relação entre *peso da evidência* e quantidade de informação? Porque Good apresenta estes dois conceitos, de *peso da evidência* e quantidade de informação, em muitos momentos referindo-se aos *desiderata* propostos por Popper e aos *explicata* possíveis em relação ao conceito de corroboração. Os *desiderata* são regras que uma teoria científica teria que respeitar para ser corroborada, e, de acordo com Good, encontram-se todas no apêndice ix de Popper (1972). Já os *explicata* são as diferentes explicações possíveis para o conceito de corroboração.

No artigo de 1960, Good afirma que a quantidade de informação a que ele se refere é diferente da teoria da comunicação de Shannon (1948), porque esta última leva em consideração a *expectativa* da quantidade de informação. Porém, em Good e Toulmin (1968), encontramos referência à relação direta entre *peso da evidência* e a teoria de Shannon. O artigo trata justamente desta questão:

> Provas de alguns teoremas de codificação, de alguma forma mais gerais que o Teorema Fundamental de Shannon, são dadas em esboço em termos de 'pesos da evidência" (logaritmos de fatores de Bayes). Embora a abordagem não tenha a intenção de ser completamente rigorosa, ela mostra como o conceito de quantidade de informação surge como uma aproximação bem próxima de peso da evidência quando a 'codificação aleatória' é utilizada (...). (GOOD, TOULMIN, 1968, p. 94, tradução nossa).

[54] Ibid., p. 320.

Seguimos com a definição, de acordo com Good, de informação a respeito da hipótese H fornecida pela evidência E dado o conhecimento prévio G:

$$I(H : E|G) = \log\frac{P(E|H.G)}{P(E|G)}$$

Agora, podemos transformar o *peso da evidência* em termos de quantidade de informação:

$$\begin{aligned} W(H : E|G) &= \log\frac{P(E|H.G)}{P\left(E\middle|\overline{H}.G\right)} \\ &= \log\frac{O(H|E.G)}{O(H|G)} \\ &= I(H : E|G) - I\left(\overline{H} : E\middle|G\right) \end{aligned}$$

Também temos a relação entre a generalização do *peso da evidência* para duas hipóteses que não são mutuamente exclusivas e a quantidade de informação:

$$W(H_1/H_2 : E\,|G) = I(H_1 : E|G) - I(H_2 : E|G)$$

Por fim, temos a quantidade de informação em termos da propriedade aditiva:

$$I(H : E.F|G) = I(H : E|G) + I(H : F|E.G)$$

4.3 A relação entre indução, corroboração e probabilidade de acordo com Good

Em muitos dos artigos publicados por Good nos quais ele afirma, em contraposição ao artigo de Popper e Miller (1983), que o suporte probabilístico é indutivo, a sua argumentação concentra-se majoritariamente em negar a prova de Popper e Miller. No entanto, Good também apresenta a sua versão de suporte probabilístico, que é o *peso da evidência*, que acabamos de discutir. Nos artigos que constituem uma espécie de resposta, ou seja, os direcionados à discussão iniciada por Popper e Miller, Good não deixa totalmente claro o que constitui o

aspecto indutivo do suporte probabilístico. Conseguimos traçar, consultando outros artigos, a linha de raciocínio de Good de uma forma mais clara.

A definição de Good sobre o que ele considera indução é o primeiro passo para essa compreensão: "Por indução científica refiro-me a mudar a probabilidade das hipóteses à luz de evidência ou observações e, desta forma, também mudar as probabilidades de futuras observações. O problema é parcialmente resolvido por meio do teorema de Bayes" (GOOD, 1988, p. 388). O *peso da evidência*, como vimos na seção anterior, é justamente a mudança da probabilidade de uma hipótese em relação à evidência.

Lindley (1990) também ajuda a esclarecer o aspecto indutivo do suporte probabilístico de acordo com a visão de Good:

> Vários cientistas apoiaram a visão de Popper da filosofia da ciência e, em particular, sua afirmação de que uma hipótese só pode ser falsificada [*disproved*]. Isso está em desacordo com a visão Bayesiana de simetria entre uma hipótese e a sua negação, entre evidência a favor e contra. Recentemente Popper e Miller (1987) argumentaram que a inferência não pode ser baseada na probabilidade. Good defendeu a visão Bayesiana admiravelmente e é um dos poucos estatísticos que se preocuparam com a abordagem de Popper e conduziu uma defesa admirável. O seguinte teorema, que Good chama de Primeiro Teorema da Indução (...), é central a qualquer resposta. Se $P(H) > 0$ e se E_1, E_2, ... são implicados por H e não por $\overline{H}$, então
>
> $$P(E_{n+1}, E_{n+2}, \dots, E_{n+m} | E_1, E_2, \dots, E_n) \to 1$$
>
> onde m e n tendem ao infinito de qualquer modo. Note-se que o resultado não se refere a probabilidade de H exceto que inicialmente ela não seja zero.
>
> Muitos estatísticos acham que a filosofia não é relevante. Good aparenta adotar a visão oposta e acha que a abordagem filosófica de alguém é importante no trabalho prático. (LINDLEY, 1990, p. 218, tradução nossa).

Ou seja, a cada nova evidência dada a evidência anterior – por exemplo $(E_2|E_1)$, $(E_3|E_2)$ e assim sucessivamente –, a probabilidade de uma nova evidência dada a anterior tende a um. Para isso, a hipótese H deve ser maior que zero e implicar a evidência e quando isso

ocorre, $\overline{H}$ não implica a evidência. Esse seria o aspecto indutivo do suporte probabilístico de acordo com Good.[55]

Outro conceito que precisamos mencionar é o conceito de corroboração. Good deixa claro em seus escritos que a probabilidade *não* significa corroboração, por mais que ela possa ter um componente indutivo como acabamos de expor. Corroboração para Good tem o mesmo sentido de *confirmação*. Good tece críticas a Carnap pelo uso da palavra confirmação para definir a probabilidade lógica, explicando que "o sentido comum em inglês de confirmação é muito mais próximo de *peso da evidência* do que de probabilidade. Prevejo que o mal-uso do termo 'confirmação' irá continuar até o ano de 2002." (GOOD, 1988, p. 393).

Discordando de Carnap nesse aspecto, por outro lado, Good concorda com Popper no sentido de que a probabilidade não significa corroboração, e afirma que é o *peso da evidência* que significa corroboração, algo que foi descoberto em Bletchley Park de forma prática antes da afirmação de Popper:

> O filósofo Karl Popper (1954) propôs um [conjunto de] *desiderata* para corroboração e disse (1959, p. 394) 'Eu considero a doutrina de que o *grau de corroboração ou aceitabilidade não pode ser uma probabilidade* como uma das mais importantes descobertas da filosofia do conhecimento'. Essa contribuição fundamental para a filosofia foi tomada como garantida por uma dúzia de criptoanalistas dezoito anos antes de Popper publicar o seu comentário, sendo usado o nome 'pontuação' ou '*decibannage*'. Além disso nós utilizávamos o melhor *explicatum* [peso da evidência], o qual não é mencionado por Popper apesar de satisfazer os seus *desiderata*. (GOOD, 1985a, p. 253, grifos do autor, tradução nossa).

A origem do termo *peso da evidência* ajuda a compreender a oposição entre a filosofia da ciência de Popper, que considera que uma hipótese pode apenas ser falsificada (ou não confirmada), e a filosofia da ciência defendida por Good, na qual uma hipótese pode ter evidência a favor ou contra ela:

> O mais antigo uso da expressão 'peso da evidência' citado pelo *Oxford English Dictionary* (1971) é uma menção de T. H. Huxley em 1878. Naquele mesmo ano Peirce (1878) publicou uma definição

[55] Além do Primeiro Teorema da Indução, Good (1988) elabora o Segundo Teorema da Indução. Ambos são formalizações do autor do conceito de "indução pura", introduzido por Keynes (1921) e tratado matematicamente também por Wrinch e Jeffreys (1921), de maneira independente, e por Huzurbazar (1955). Keynes (1921, p. 233 et seq.) faz críticas ao posicionamento de Hume em relação à indução, investigando ao longo capítulo XX "as condições e a maneira pelas quais a mera repetição de instâncias pode adicionar à força do argumento".

> formal, então ou a expressão estava em uso regular ou talvez é simplesmente autoexplicativa num sentido qualitativo. Claramente o conceito era familiar aos gregos antigos porque Themis, a deusa da justiça, é representada segurando um par de bandejas nos quais ela pesa argumentos opostos. (GOOD, 1988, p. 389, tradução nossa).

4.4 Visão antagônica de Popper e Miller: o suporte probabilístico dedutivo

Popper e Miller (1983) consideram que o suporte probabilístico dado por uma evidência *e* para uma hipótese *h*, considerando um conhecimento prévio *b*[56], é um suporte probabilístico dedutivo. Vamos seguir a argumentação dos autores a esse respeito, utilizando a notação original.

Os autores apresentam uma sequência de 8 equações, que mostram as relações entre evidência, hipótese e conhecimento prévio. Destas equações, derivam dois teoremas e um corolário.

A primeira equação formaliza a ideia de que, quando a evidência e é deduzida de h considerando o conhecimento prévio b, a probabilidade da evidência dada a hipótese e o conhecimento prévio é igual a 1:

1. $p(e, hb) = 1$

Se a evidência e está logicamente contida na hipótese h, então a probabilidade da hipótese e da evidência dado o conhecimento prévio é igual à probabilidade da hipótese dado o conhecimento prévio.

2. $p(he, b) = p(h, b)$

Agora, partir da seguinte regra do produto:

3. $p(he, b) = p(h, eb)p(e, b)$

E considerando que a probabilidade $p(e, b)$ assuma os seguintes valores:

[56] "*b*" de *background knowledge*, e significa para os autores mais precisamente "as condições iniciais necessárias para derivar as previsões *e* de *h*" (POPPER; MILLER, 1983, p. 687).

4. $0 < p(e, b) < 1$

Pode-se obter a partir das duas últimas equações (desde que $p(h, b) > 0$):

5. $p(h, eb) > p(h, b)$

De acordo com os autores, é uma ilusão acreditar que esta última equação justifique a crença na indução probabilística. Para provar que este não é o caso, eles admitem que uma hipótese possa ser fatorizada da seguinte forma:

6. $h = (h \leftarrow e)(h \vee e)$

Sendo que $(h \leftarrow e)$ significa "h se e" ou "se e então h". O segundo fator desta equação segue dedutivamente de e $(h \vee e)$, e o primeiro fator, $(h \leftarrow e)$, é dedutivamente equivalente a h. Ou seja, se e está contido em h, podemos adicionar e ao primeiro fator, e essa probabilidade será igual a $p(h, e)$, não faz diferença:

7. $p(h \leftarrow e, e) = p\big((h \leftarrow e)e, e\big) = p(he, e) = p(h, e)$

Além disso, quando a evidência e é adicionada aos fatores de h, mantém-se a igualdade com a probabilidade condicional $p(h, e)$, os fatores de h são independentes:

8. $$p\big((h \leftarrow e)(h \vee e), e\big) = p(h, e) = p(h \leftarrow e, e)$$
$$= p(h \leftarrow e, e)p(h \vee e, e)$$

A conclusão dos autores em relação ao caráter dedutivo da probabilidade é a seguinte: "nós separamos h em relação a e em dois fatores. Um deles segue dedutivamente de e, e pode ser, na presença de e (e de b), ignorado: o outro é, na presença de e (ou de e e b) equivalente a h." (POPPER, MILLER, 1983, p. 687)

O primeiro teorema diz, então, que e "contra-suporta" $h \leftarrow e$, ou seja, que dado o primeiro fator, se adicionamos a evidência, a sua probabilidade será menor que a sua probabilidade inicial:

Teorema 1:

Se $p(h,e) \neq 1 \neq p(e)$,
então $p(h \leftarrow e, e) < p(h \leftarrow e)$

O grau de contra-suporte é igual ao excesso (h,e), ou seja, a diferença entre a probabilidade do primeiro fator menos a probabilidade do fator dado a evidência é maior que zero:

Teorema 2:

$$p(h \leftarrow e) - p(h \leftarrow e, e) = p(h \leftarrow e) - p(h,e)$$
$$= p(-h,e)p(-e) = Exc(h,e) > 0$$

Em seguida, temos o corolário que adiciona o conhecimento prévio *b* ao teorema 2:

Corolário:

Se $p(h,eb) \neq 1 \neq p(e,b)$,
então $p(h \leftarrow e,b) - p(h \leftarrow e,eb) = p(h \leftarrow e,b) - p(h,eb)$
$= p(-h,eb)p(-e,b) = Exc(h,e,b) > 0$

Por fim, ou autores apresentam mais uma equação que mede o "contra-suporte" da evidência *e* dado o conhecimento prévio *b*:

9. $ct(e,b) = 1 - p(e,b)$

Essa equação significa que

> o fator que contém tudo de *h* que não segue dedutivamente de *e* é fortemente contra-suportado por *e*. Ele é mais contra-suportado quanto maior o conteúdo de *e* (...) o contra-suporte aumenta com o

> conteúdo de e, se e suporta h ou não. (POPPER, MILLER, 1983, p.688).

4.5 Outras visões sobre medidas de corroboração

Fitelson (2001) faz um estudo sobre as características de quatro medidas de corroboração, entre elas o *peso da evidência* e uma outra medida, que é ao mesmo tempo uma medida de corroboração para alguns bayesianos e uma medida de suporte probabilístico para Popper. Um dos objetivos da proposta de Fitelson é mostrar quais são as medidas de corroboração que satisfazem os *desiderata* de Peirce para o suporte probabilístico de evidências independentes, entre outras condições.

O autor denota todas as medidas como bayesianas, incluindo a medida de Popper. No entanto, observamos que o autor não menciona Popper e atribui a medida a Gilles (1986), um defensor da linha filosófica de Popper e do bayesiano Jeffrey (1992), entre outros. Não consideramos o uso do termo "bayesiano" para esta medida de corroboração como o mais adequado, justamente por incluir anti-bayesianos como Popper e Gilles. Na verdade, a medida que Fitelson denomina como sendo uma medida de "confirmação", os bayesianos consideram como corroboração, já Popper e Gilles a consideram como um suporte probabilístico:

> Popper e Miller (...) argumentaram que a indução probabilística é impossível (...) Gilles segue Popper e Miller no uso, como uma medida de suporte de uma hipótese h, fornecida pelas observações e, a expressão $s(h : e) = p(h|e) - p(h)$. (GOOD, 1987, p. 470, tradução nossa).

Este fato não afeta a proposta de Fitelson, porém, é importante mencioná-lo, já que uma de nossas propostas é justamente apontar as diferenças filosóficas e matemáticas entre a medida do grau de corroboração de Good e de Popper.[57] Essas controvérsias são comuns. Veremos mais adiante, por exemplo, que Gilles (1990) não considera o *peso da evidência* como uma medida bayesiana.

Após esta ressalva, voltamos à proposta de Fitelson. O autor estuda "o grau de confirmação incremental fornecido por proposições evidenciais E para as hipóteses sob teste

[57] Fitelson também utiliza o termo "confirmação" para o que consideramos como corroboração. Apresentamos o tema como um problema de corroboração, porém em seguida optamos pela utilização do termo "confirmação" de acordo com o uso Fitelson, para que corresponda mais fielmente à proposta original do autor.

H, dadas as evidências prévias K, de acordo com as medidas de relevância do grau de confirmação c." (FITELSON, 2001, p. S123). A medida c é uma medida de relevância do grau de confirmação se, e somente se, satisfaz as restrições (no caso de E confirmar, desconfirmar ou ser irrelevante para H dadas as evidências prévias K):

A medida c confirma quando $c(H, E|K) > 0$, se $P(H|E \cap K) > P(H|K)$,
A medida c desconfirma quando $c(H, E|K) < 0$, se $P(H|E \cap K) < P(H|K)$,
A medida c é irrelevante quando $c(H, E|K) = 0$, se $P(H|E \cap K) = P(H|K)$.

Dessa forma, Fitelson reduz a questão do grau de confirmação de E para a hipótese H (sob evidências prévias K) à relação entre probabilidade condicional de H dado $E \cap K$ e à probabilidade condicional de H dado K, ou seja, à relação entre $P(H|E \cap K)$ e $P(H|K)$. O autor apresenta então quatro formas de representar o grau de confirmação, quatro medidas diferentes:

Medida da diferença:

$$d(H, E|K) = P(H|E \cap K) - P(H|K)$$

Medida da log-razão:

$$r(H, E|K) = \log \frac{P(H|E \cap K)}{P(H|K)}$$

Medida da log-razão de verossimilhança:

$$l(H, E|K) = \log \frac{P(E|H \cap K)}{P\left(E\middle|\overline{H} \cap K\right)}$$

Medida da diferença normalizada:

$$s(H, E|K) = P(H|E \cap K) - P\left(H\middle|\overline{E} \cap K\right)$$
$$= \frac{1}{P\left(\overline{E}\middle|K\right)} . d(H, E|K)$$

A medida $d(H, E|K)$ é a medida de Popper e a medida $l(H, E|K)$ é o *peso da evidência*. Ambas satisfazem o que Fitelson chama de *desiderata* fundamental de Peirce. Vemos que essa estrutura aditiva da confirmação está presente muito antes dos trabalhos de Good, que segue uma linha independente da de Peirce. De acordo com Fitelson, Peirce (1878) apresenta duas intuições importantes sobre o "suporte indutivo independente". Seguimos com a formalização da primeira intuição de Peirce, no formato de uma definição, ao afirmar que quando duas evidências E_1 e E_2 são independentes, o suporte indutivo de uma ou de outra para uma hipótese H não depende se alguma delas é parte da evidência prévia:

Definição: E_1 e E_2 são (mutuamente) confirmativamente independentes a respeito de H de acordo com c se, e somente se, $c(H, E_1|E_2) = c(H, E_1)$ e $c(H, E_2|E_1) = c(H, E_2)$.

A segunda intuição é sobre a aditividade. Se as evidências são independentes, a confirmação da hipótese a partir da conjunção das evidências é igual à soma da confirmação das duas evidências:

Se E_1 e E_2 são confirmativamente independentes a respeito de H de acordo com c, então $c(H, E_1 \cap E_2) = c(H, E_1) + c(H, E_2)$.

Fitelson segue o artigo testando as quatro medidas em mais três outras condições. A medida $l(H, E|K)$ é a que satisfaz o maior número de condições. Vamos nos aprofundar um pouco mais nas diferenças entre a medida de Popper e o *peso da evidência*. Gilles (1990) aponta uma relação importante entre a medida de Popper e o conceito do *princípio da testagem severa*.

A medida de Popper é conhecida como $Q = P(E|H \cap K) - P(E|K)$. Gilles inicia sua argumentação sobre o significado da medida citando um trecho da *Lógica da Pesquisa Científica*. O trecho encontra-se no capítulo X, que trata exatamente de corroboração. Na seção 82, "Teoria positiva da corroboração: como uma hipótese pode 'assegurar sua qualidade'", temos enfim a definição de testagem severa: "(...) não é tanto o número de casos corroboradores que determina o grau de corroboração, mas sim a *severidade dos vários testes* a que a hipótese em pauta pode ser e foi submetida" (POPPER, 1972, p. 293). De acordo com Gilles, é a medida $Q = P(E|H \cap K) - P(E|K)$ que é capaz de medir a severidade do teste τ:

> Q é grande se $P(e|h \cap k)$ é grande, i.e., se e é altamente provável dado $h \cap k$, e se $P(e|k)$ é pequena, i.e. se e é altamente improvável dado k, mas não h. Se e satisfaz essas condições, então o teste correspondente τ é severo, e Q pode ser tomada como uma medida da severidade de um teste. (GILLES, 1990, p. 145, tradução nossa).

O objetivo de Gilles ao explicar a medida Q é mostrar que a medida de *peso da evidência* [58] está relacionada de forma muito próxima de Q. Supondo que a hipótese H seja uma hipótese universal, do tipo "Todos os corvos são pretos", a probabilidade da hipótese é igual a zero, ou seja, $P(H) = 0$. Esse fato é aceito tanto para Popper, Carnap e os subjetivistas, de acordo com Gilles. Além disso, quando omitimos K apenas por conveniência, obtemos a seguinte probabilidade de $P(E)$:

$$P(E) = P(E|H).P(H) + P\left(E\middle|\overline{H}\right).P\left(\overline{H}\right)$$
$$= P\left(E\middle|\overline{H}\right)$$

Neste caso, o *peso da evidência* seria igual a

$$W(H : E) = logP(E|H) - logP(E)$$

E a medida Q de Popper

$$Q = P(E|H) - P(E)$$

Gilles chega à conclusão de que, no caso de $P(H) = 0$, as duas funções são essencialmente equivalentes. Para ele, o uso do logaritmo é apenas uma "conveniência matemática" e não difere do ponto de vista de interpretação filosófica. Discordamos destes dois pontos, além de uma outra afirmação que Gilles faz ao longo do artigo dizendo que o *peso da evidência* não é uma medida bayesiana.

Em primeiro lugar, no caso do *peso da evidência*, o logaritmo é aplicado à razão bayesiana de verossimilhança. A razão $P(E|H)/P(E|\overline{H})$ é essencialmente bayesiana. As probabilidades desta razão bayesiana ou das probabilidades que compõem as referidas

[58] Gilles chama o *peso da evidência* de "função Turing-Good de peso da evidência".

probabilidades condicionais podem assumir qualquer valor igual ou maior que 0 e menor ou igual a 1, e mesmo que tenhamos valores os extremos como $P(H) = 0$ e $P(\overline{H}) = 1$, a razão não deixa de ser bayesiana por este motivo, por mais que fique mais semelhante à medida Q de Popper: a interpretação filosófica mantém-se. Além disso, Good afirma que o *peso da evidência* não depende de uma avaliação prévia de $P(H)$:

> Outra vantagem de W [peso da evidência] em relação a s [medida de Popper/Gilles] e σ [medida de Redhead] é que ela não depende de um julgamento da probabilidade inicial de h, a qual é sujeita a uma grande variação de um avaliador para outro. (GOOD, 1987, p. 471-472, tradução nossa).

O logaritmo não é uma "conveniência", mas uma necessidade. Sem o logaritmo, a expectativa do *peso da evidência* não assume uma distribuição assintoticamente normal. E sabemos que "Para decidir antecipadamente se o *Banburismus* seria provavelmente bem-sucedido, Turing estimou a expectativa do peso da evidência" (GOOD, 1985a, p. 257). No próximo capítulo, os motivos empíricos que justificam a utilização do *peso da evidência* serão esclarecidos em maiores detalhes.

5. A QUEBRA DA CRIPTOGRAFIA DO ENIGMA

5.1 Introdução

O *peso da evidência* e o processo de quebra da criptografia do Enigma estão intimamente relacionados. Ambos foram elaborados por Turing e aplicados por ele e por sua equipe de trabalho em Bletchley Park (BP), embora o conceito de *peso da evidência* tenha sido elaborado em maiores detalhes e divulgado por Good (1950). Não é simples relatar tanto a história quanto a descrição da elaboração e aplicação dos métodos estatísticos utilizados por Turing e sua equipe, neste período, por vários motivos.

O primeiro deles é a necessidade, por um período de cerca de 30 anos, de absoluto sigilo. A quebra da criptografia ocorreu no início da década de 1940 e as primeiras informações sigilosas começaram a ser reveladas em 1974, com o lançamento do livro *The Ultra Secret*, escrito por F. W. Winterbotham, um oficial da *Royal Air Force* que havia participado das operações em BP com a distribuição das mensagens decriptografadas. Hinsley e Stripp (1993) afirmam em data relativamente recente (1992), ou seja, à época em que escreviam o prefácio de seu livro *Codebreakers: the inside history of Bletchley Park*, que algumas informações ainda demandavam sigilo:

> O sigilo absoluto preservado por todos que trabalharam lá [em BP] foi e ainda é um feito extraordinário. Ainda que a equipe do período de guerra de Bletchley Park esteja agora em contraste absolvida de reticência em relação ao seu trabalho, o governo ainda requer que eles não divulguem detalhes de alguns métodos pelos quais a inteligência foi obtida. (HINSLEY; STRIPP, 1992, p. vii, tradução nossa).

Kahn (1996, p. 2.173) afirma que esse sigilo pode ser classificado como sendo "provavelmente o melhor exemplo de sigilo geral na história". As consequências dessa necessidade de sigilo tiveram diferentes vertentes. Muitos acadêmicos e profissionais que saíram de Bletchley Park faleceram sem que as pessoas mais próximas a eles, como familiares, amigos e colegas de profissão tivessem qualquer conhecimento das suas contribuições. Alguns sofreram no ambiente acadêmico, enfrentando muitas dificuldades, por exemplo: tiveram dificuldade de formar e manter uma boa equipe de trabalho, sofreram com descrença de pares

em sua pesquisa, entre outros problemas. Tanto Turing como Good passaram por adversidades, mas de maneiras diferentes. De acordo com McGrayne (2011), o sigilo teve

> um efeito catastrófico em Turing. (...) [ele] juntou-se ao *National Physical Laboratory* em Londres. Devido ao *Official Secrets Act* ele chegou lá como um ninguém. Se ele tivesse recebido o título de *Sir* ou sido honrado de alguma outra forma ele certamente teria mais facilidade em ter mais do que dois engenheiros como equipe de suporte. Ignorante das realizações de Turing, o diretor do laboratório, Charles Galton Darwin, o neto de Charles Darwin, reprimia Turing repetidamente por seus atrasos matinais após trabalhar até tarde na noite anterior. (...)
>
> No laboratório, Turing projetou o primeiro computador digital eletrônico *stored-program*[59] relativamente completo para quebrar códigos em 1945. No entanto, Darwin considerou o projeto muito ambicioso, e após muitos anos Turing deixou [o laboratório] com sentimento de repugnância. Quando o laboratório construiu finalmente o seu projeto em 1950, ele era o computador mais rápido do mundo, e, espantosamente, tinha a memória de um primeiro Macintosh construído três décadas depois. (MACGRAYNE, 2011, p. 84-85, tradução nossa).

A falta de reconhecimento, que frequentemente era uma consequência da necessidade de sigilo, também afetou muitos outros colaboradores de BP. Good (1993, p. 163), por exemplo, cita o engenheiro A.W.M. Coombs, que trabalhou com a decriptografia de outro sistema alemão de criptografia, que eles chamavam de *Fish*; a falta de crédito da sua contribuição à comunicação transatlântica pela administração de Dollis Hill fez com que Coombs pensasse em suicídio.

De acordo com MacGrayne (2001, p. 100), o próprio Good não pôde ser identificado, à época do lançamento do livro *The Codebreakers* de Kahn em 1967, como um dos três principais criptoanalistas do Reino Unido. Uma passagem original do livro que fazia essa referência à Good foi censurada pela *National Security Agency*. A falta de reconhecimento também atingiu Good. O relato de Lindley a respeito de Good mostra que ele não usufruía de uma posição de destaque mesmo entre seus pares:

> Se eu olhar retrospectivamente [década de 1960, nas reuniões da *Royal Statistical Society*] para o que *eu* disse, é frequentemente rude. Eu fiquei envergonhado quando vi uma das discussões a respeito

[59] Ou seja, o mesmo que uma máquina de Turing universal. Para informações a respeito de máquina de Turing, consultar Davis (1958) e Carnielli e Epstein (2005).

> dos artigos de Jack Good. Good é alguém que eu, e talvez Savage, não compreendemos e menosprezamos. Ele não passou as ideias dele para nós. Nós deveríamos ter tido muito mais respeito para o que ele estava dizendo porque ele estava muito avançado em relação a nós em muitos aspectos. Esse é um dos meus arrependimentos, mas eu tive a oportunidade de reparar isto escrevendo um resumo da sua obra em 1990 (...). Olhando para trás, acho que muitos de nós tínhamos pouco apreço por Good. Não sei se as suas explicações não eram claras ou se nós não nos dávamos ao trabalho, mas a mensagem não chegou a nós (...). (SMITH, 1995, p. 312, tradução nossa).

Embora não possamos ter certeza quanto a uma relação direta de causa e consequência entre a censura de Good como um eminente criptoanalista e a falta de reconhecimento vinda de outros estatísticos bayesianos, é possível que se o trabalho criptográfico de Good não tivesse sido censurado e ele tivesse um destaque maior, Lindley e outros estatísticos teriam prestado mais atenção a ele.[60]

A outra vertente da consequência que advém da necessidade sigilo é a dificuldade de descrição e a memória dos métodos e procedimentos realizados em BP:

> A compilação desse livro [*Codebreakers: the inside history of Bletchley Park*] baseou-se quase completamente em memórias pessoais; e isso não é usual em um relato que pretende ter qualquer tipo de acurácia. (...) Para todos nós a recordação clara e acurada de fatos secretos altamente especializados por cinquenta anos tem sido uma tarefa de alta demanda, requerendo bastante verificação cruzada. (...) Dentro de cada bloco, cabana ou cômodo tínhamos pouco tempo e menos encorajamento para discutir assuntos de interesse em comum a não ser que o trabalho demandasse isto especificamente. A maioria de nós não tinha ideia 'do que acontecia na porta ao lado', especialmente à medida que aumentava a variedade de empregos e mais funcionários chegavam. (HINSLEY; STRIPP, 1992, p. vi-vii, tradução nossa).

Há falta de acurácia na descrição dos métodos e procedimentos não só pelo tempo decorrido e pela necessidade de sigilo, mas também pela própria complexidade da tarefa. BP chegou a ter cerca de 10.000 pessoas trabalhando simultaneamente; o código gerado pelo Enigma naval era um entre outros códigos que estavam tentando quebrar em BP. Além do código do Enigma, existia o *Fish*, que era mais complexo ainda. As máquinas que produziam

[60] Para mais informações a respeito de Good, pode-se consultar uma entrevista feita por Banks (1996). A entrevista é bastante detalhada, ela acompanha desde a infância de Good até os seus últimos trabalhos, e inclui todo o período que ele passou em BP.

código *Fish* eram utilizadas pelo exército alemão; elas não eram portáteis como a máquina Enigma e possuíam duas formas diferentes: uma chamada pelos britânicos de *Sturgeon* e a outra de *Tunny*. Good trabalhou também com esta última.

Os códigos japoneses também foram interceptados e havia um grupo trabalhando com a sua decriptografia. Edward Simpson (1922-2019) trabalhou em BP entre 1942 e 1945, inicialmente na quebra de códigos italianos, e em seguida, com o código naval japonês, conhecido como JN 25 (*Japanese Naval*). O trabalho de decriptografia do JN 25 era bem menos complexo que o trabalho do Enigma. O teorema de Bayes era utilizado para decriptografar todos os códigos: Enigma, *Fish* e JN 25. Nas palavras de Simpson:

> Eu consegui reconstruir o que a contribuição de Bayes para o ataque ao Enigma pode ter sido. Bayes foi utilizado também em *Tunny*, mas eu não tenho detalhes a esse respeito. Porém aqui, porque eu sei em primeira mão, descrevo o uso feito de Bayes no ataque criptoanalítico à principal cifra naval japonesa JN 25 em 1943-1945, pelo grupo do bloco B, o qual eu liderei. (SIMPSON, 2010, p. 77, tradução nossa).

Simpson reconhece o uso do teorema de Bayes em todos os códigos citados. Porém ressaltamos, mais uma vez, a dificuldade de descrição do processo de decriptografia do Enigma: mesmo Simpson, com tanta experiência, esforçou-se para "reconstruir" o uso de Bayes neste caso. Em mais um trecho do relato de Simpson, podemos confirmar a falta de troca de conhecimento à época a respeito dos trabalhos realizados pelas diferentes equipes de decriptografia:

> A criptoanálise do Enigma naval alemão era significativamente mais complexa que aquela do JN 25. (...) Novamente menciono apenas três das muitas pessoas envolvidas: Alan Turing, que dispensa qualquer introdução; Hugh Alexander, que além de ser campeão britânico de xadrez foi tanto um mestre em criptoanálise e um administrador mestre de criptoanalistas; e Jack Good, matemático e posteriormente professor de estatística na Virginia Tech. Em virtude da compartimentação de todas as várias equipes em Bletchley Park eu não sabia de nada do trabalho deles naquela época. (SIMPSON, 2010, p. 79, tradução nossa).

É interessante notar, apesar da falta de troca de conhecimento entre as equipes, que Simpson (2010, p. 78) cita inclusive o *peso da evidência* como um método utilizado para decifrar o JN 25:

> Os matemáticos no grupo do JN 25 eram Ian Cassels (posteriormente *Fellow* de *Trinity* e professor *Sadleirian* em Cambridge), Jimmy Whitworth e eu. Nossa solução para o processo necessário de julgamento iniciou-se com uma aplicação do teorema de Bayes. (...) O peso da evidência fornecido por Q [evidência] a favor de A [hipótese] sendo verdadeira era o fator de Bayes p_t [probabilidade de Q ocorrer se A fosse verdadeiro] dividido por p_f [probabilidade de Q ocorrer se A fosse falso]. (SIMPSON, 2010, p. 78, tradução nossa).

A complexidade da tarefa de criptografia era tamanha que até mesmo pessoas que trabalhavam muito próximas, na mesma equipe, compreendiam apenas superficialmente determinados aspectos da criptografia do Enigma. Milner-Barry (1993), um campeão de xadrez que trabalhava diretamente com o Enigma, fez o seguinte relato:

> Welchman tinha, eu acho, sido recrutado para o trabalho de inteligência antes da guerra eclodir, e ele já havia feito um bom trabalho de pesquisa nos mistérios da máquina Enigma. Esses [mistérios] vinham com facilidade para Alexander, mas eu sou (apesar da minha ancestralidade) quase um enumerado. Por esse motivo achei as explicações pacientes de Gordon Welchman muito difíceis de acompanhar, e até hoje eu não posso afirmar que eu entendi completamente como a máquina funcionava, muito menos o que estava envolvido nos problemas de quebra e leitura da cifra do Enigma. Felizmente isso não importava muito, dado que eu era capaz de ser útil de outras maneiras – principalmente porque eu tinha conhecimento da língua alemã. (MILNER-BARRY, 1993, p. 89-90, tradução nossa).

Nossa intenção em enfatizar este tipo de relato não é nos eximir quanto a explicar da melhor forma possível a quebra da criptografia do Enigma, mas de expor o grau de dificuldade e de complexidade da tarefa. Portanto, nosso relato será concentrado mais nos aspetos estatísticos da quebra da criptografia, e não no processo todo: vamos apresentar um segmento da quebra da criptografia.

5.2 Algumas características básicas da máquina Enigma

A primeira máquina Enigma foi patenteada em 1919. Ela passou a ser adotada em 1926 pela marinha, em 1928 pelo exército e em 1935 pela força aérea alemã, além de ser

utilizada para outros serviços do governo. A máquina sofreu diversas modificações nesse período, até 1945. Sua função era codificar e decodificar mensagens. Ela possuía um teclado de 26 letras, sem incluir pontuação ou números. A cada letra digitada, outra letra aparecia num painel com o mesmo padrão das 26 letras; era um painel no qual a letra correspondente já codificada aparecia iluminada. Quando a mensagem original repetia uma letra, a letra codificada era sempre diferente da codificação anterior, e a sequência só voltava a se repetir depois de 16.900 tecladas. Somando todos os tipos diferentes de configurações iniciais da máquina que poderiam ser selecionados para fazer uma codificação, existia um total de 159 quintilhões de combinações possíveis.[61]

A German Enigma coding machine. Courtesy of Bletchley Park Museum

Figura 5 – A máquina alemã Enigma. FONTE: Simpson (2010).

Para quebrar o código, era necessário saber quais eram as configurações iniciais da máquina. A equipe de Turing tentava descobrir qual era a configuração de uma das três rodas

[61] Cf. Stripp (1993, p. 83-88).

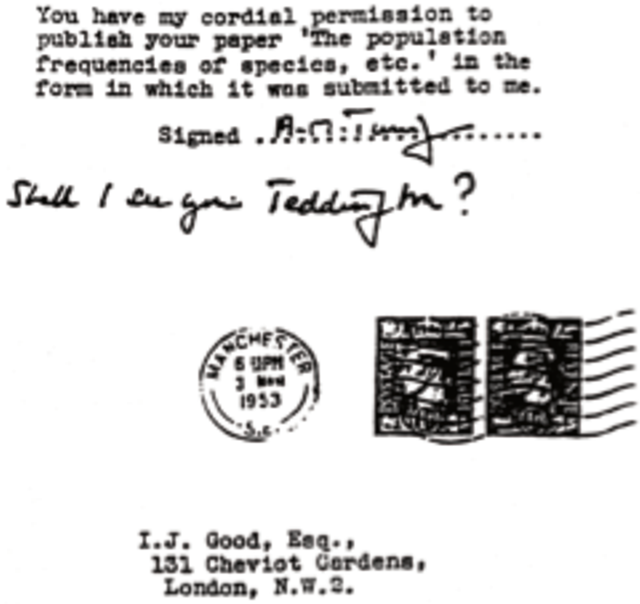
(*wheel*) iniciais. Se eles conseguissem descobrir uma delas, a que ficava no lado direito em relação às outras, os próximos cálculos eram feitos pela máquina *bombe*:

> Como disse acima, o Banburismus identificaria a posição do entalhe da roda direita (que causava o movimento da roda do meio). (...) A identificação da roda direita reduzia o número de ordens possíveis das rodas de $8 \times 7 \times 6 = 336$ para 42 ou 126. Essa redução de trabalho era especialmente valiosa quando o tempo da *bombe* era escasso. (GOOD, 1993, p. 158, tradução nossa).

Em seguida, vamos ver que tipo de cálculo e estimativas eram feitos para descobrir qual era a posição da roda direita.

5.3 Descrição dos métodos desenvolvidos por Turing de acordo com Good

Em 1953, Good pediu a permissão de Turing para publicar um artigo no qual descrevia o método estatístico que fora desenvolvido e utilizado por Turing para quebrar a criptografia do Enigma naval. Turing nunca publicou o seu trabalho estatístico desenvolvido em BP. Good só começou a dedicar-se a escrever o artigo após certificar-se que os interesses de Turing estavam concentrados em outras áreas, e também, após "a estatística não ser mais considerada um assunto sigiloso" (GOOD, 2000, p. 106).


You have my cordial permission to publish your paper 'The population frequencies of species, etc.' in the form in which it was submitted to me.

Signed .A.M.Turing.....

Shall I see you Teddington?

I.J. Good, Esq.,
131 Cheviot Gardens,
London, N.W.2.

Figura 6 – Autorização de Alan Turing, datada de 3 de março de 1953, para que Good pudesse publicar o artigo "The population frequencies of species and the estimation of population parameters" (1953). FONTE: Good (2000).

O primeiro artigo que citamos, de 1953, é um dos mais completos de todos sobre o assunto, contendo mais de 100 fórmulas. Devido ao sigilo, a descrição dos métodos estatísticos foi aplicada ao caso de uma amostra de uma população de animais e o que poderia ser inferido a respeito da população, como, por exemplo, a frequência de uma determinada espécie de animal. Apesar deste artigo ser extenso e detalhado, não existe um único artigo no qual Good faça a descrição passo a passo de todo o processo de decriptografia. Mesmo porque a decriptografia envolvia outros processos além da estatística e a colaboração de Turing também incluía outras áreas. O sigilo também impediu que as informações fossem transmitidas com clareza e de forma completa, portanto, consultamos vários artigos e selecionamos as informações mais importantes que estão diretamente relacionadas ao conceito de *peso da evidência*.

Antes de iniciarmos a descrição do *Banburismus* relacionado ao *peso da evidência*, é importante ressaltar que Turing e o trabalho realizado em BP não foi iniciado do zero, ou seja, sem nenhuma informação prévia. Os poloneses enviaram aos franceses e britânicos todas as informações que tinham, incluindo uma réplica da máquina Enigma, uma máquina chamada *bomba*, precursora da máquina *bombe* que já mencionamos e um método iniciado de quebra de criptografia. Vejamos um relato de Welchman publicado pela primeira vez em 1986 a esse respeito:

> Os poloneses mantiveram o segredo com eles até julho de 1939, quando, com a eminente invasão alemã de seu país, eles deram todo o seu conhecimento, assim como réplicas em funcionamento da máquina Enigma, aos franceses e britânicos. Na Inglaterra, em Bletchley Park, nós fomos rápidos em explorar a oportunidade de ouro que os poloneses nos deram.
>
> (...) se os poloneses não tivessem nos dado os detalhes do Enigma em Pyry, o GC&CS britânico provavelmente continuaria achando que o problema do Enigma não tinha esperança sem uma captura. Mesmo se Turing tivesse pensado na sua bombe, teria tido pouca ou nenhuma justificativa para o desenvolvimento de sua engenharia. (WELCHMAN, 2019, p. 195 e 223 respectivamente, tradução nossa).

Até mesmo o *Banburismus* era um aperfeiçoamento de um método já desenvolvido pelos poloneses: "O Banburismus era uma elaboração de um método chamado 'método do relógio' por Rejewski, que atribui o método a Jerzy Różycki." (GOOD, 1993, p. 155). Vamos nos concentrar, portanto, somente nas inovações atribuídas a Turing e Good em relação ao método.

Uma das descobertas de Turing quando trabalhava com o *peso da evidência* no processo do *Banburismus* foi um teorema interessante: Turing descobriu que o fator esperado a favor de uma hipótese incorreta é igual a um. Suponha que os resultados possíveis de um experimento sejam E_1, E_2, E_3, ... e que a hipótese H seja verdadeira. Temos a seguinte definição de fator contra H, se E_i é um resultado observado:

$$F\left(\overline{H}: E_i\right) = \frac{P\left(E_i \middle| \overline{H}\right)}{P(E_i|H)}$$

Agora, a expectativa deste fator, dada a hipótese verdadeira H é igual a um:

$$E\left(F(\overline{H}: E_i)\middle|H\right) = \sum_i \frac{P\left(E_i \middle| \overline{H}\right)}{P(E_i|H)} . P(E_i|H)$$

$$= \sum_i P\left(E_i \middle| \overline{H}\right) = 1$$

Retomando o exemplo de lançamento de um dado (seção 4.2.1), podemos verificar este teorema. Vamos supor que o dado seja normal, e que, portanto, a hipótese correta seja essa e que a hipótese incorreta é a de que o dado é enviesado:

H = o dado é normal

$\overline{H}$ = o dado é enviesado

E_1 = apresentar número 6

E_2 = apresentar outro número

$P(E_1|H) = 1/6$

$P(E_1|\overline{H}) = 1/3$

$P(E_2|H) = 5/6$

$P(E_2|\overline{H}) = 2/3$

Então, com estes dados calculamos a esperança do fator:

$$E(F(\overline{H}:E_i)|H) = \frac{P(E_1|\overline{H})}{P(E_1|H)}.P(E_1|H) + \frac{P(E_2|\overline{H})}{P(E_2|H)}.P(E_2|H)$$
$$= \frac{1/3}{1/6}.1/6 + \frac{2/3}{5/6}.5/6 = 1$$

Este teorema pode ser utilizado para conferir se o cálculo do fator de Bayes está correto. Wald (1944) também chegou ao mesmo teorema.

Outra conclusão de Turing que foi apresentada a Good em 1941, é que a expectativa do *peso da evidência* a favor da hipótese verdadeira é não-negativa (só desaparece quando o *peso da evidência* é igual a zero para todo E de probabilidade positiva):

$$E(W(H:E)|H) \geq 0$$

De acordo com Good,

> Esta inequação foi apontada a mim por Turing em 1941, com uma prova diferente. Considerada como uma álgebra é apenas uma inequação elementar. O que a torna interessante é a sua interpretação

> em termos de inferência humana, e, portanto, estatística. É por este motivo que considero sensato atribuí-la a Turing apesar de ter sido aplicada também à estatística mecânica por Gibbs (1902, p. 136). (GOOD, 1985a, p. 256, tradução nossa).

Turing descobriu outro teorema, a partir da necessidade de estimar a expectativa do *peso da evidência*. Considere que o *peso da evidência* a favor da hipótese H, quando a hipótese H é verdadeira, tem uma distribuição normal com média μ e variância σ^2, com a unidade de ban natural (logaritmo com base e). Então a relação entre a variância e a expectativa será a seguinte:

$$var(W(H:E)|H) = 2E(W(H:E)|H)$$

Ou seja, $\sigma^2 = 2\mu$. Outro resultado refere-se a uma propriedade da expectativa do *peso da evidência* a favor da hipótese falsa, quando a hipótese é verdadeira:

$$E\left(W(\overline{H}:E)\middle|H\right) = -E(W(H:E)|H) = -\mu$$

Lemos este resultado como "a expectativa do *peso da evidência* a favor de uma hipótese falsa (incorreta) quando a hipótese é verdadeira é igual à média negativa". Em um estudo sobre radares, os mesmos resultados foram encontrados por Peterson, Birdsall e Fox (1954). A relação entre média e variância muda se considerarmos como medida de *peso da evidência* os decibans (logaritmo base 10) ao invés do ban natural (logaritmo base e):

$$\sigma = \sqrt{\mu 20 log_{10} e} = \sqrt{8{,}686\mu} \cong 3\sqrt{\mu}$$

Neste caso, o desvio padrão σ é aproximadamente 3 vezes a raiz da média, isso significa que o desvio padrão é grande.

Good indica alguns dos pontos relevantes em relação à expectativa e à variância do *peso da evidência*:

> Quando Turing julgou o valor do Banburismus pela estimação da expectativa do peso da evidência ele estava na verdade tratando o peso da evidência como se fosse uma utilidade. Ele [peso da evidência] pode ser considerado como uma *quase-utilidade*, ou seja,

> uma substituição aditiva para a utilidade expressa em termos de probabilidades. (...) No design de um experimento toda a distribuição do peso da evidência, em particular a sua variância, é de interesse, não apenas a sua expectativa. (GOOD, 1985a p. 258, grifos do autor, tradução nossa).

Este último teorema de Turing sobre expectativa e variância de *peso da evidência* pode ser generalizado para um caso mais realista, no qual a distribuição normal do *peso da evidência* só ocorre na região próxima da média μ.[62] Há também um estudo de Kullback (1959) que faz uso não bayesiano da expectativa do *peso da evidência*.

[62] Para mais detalhes, consultar Good (1961).

6. CONCLUSÃO

Good afirma que "todo teste bayesiano de uma hipótese pode ser considerado como uma aplicação do conceito de peso da evidência" (GOOD, 1985a, p. 267). De fato, podemos traçar um paralelo entre o conceito de *peso da evidência*, que consiste essencialmente em um cálculo para encontrar a hipótese mais provável dado as evidências disponíveis, e um teste bayesiano como descrito por O'Hagan e Forster (2004), no qual inferimos algo sobre o parâmetro θ (hipótese) de acordo com os dados x (evidências) disponíveis.

Outra forma de ver essa semelhança entre o *peso da evidência* e os testes bayesianos é a relevância da própria evidência, que altera a chance ou probabilidade inicial para uma chance final ou probabilidade posterior. No caso do *peso da evidência*, temos:

$$\text{chances finais} = \text{peso da evidência} + \text{chances iniciais}$$

E, no teste bayesiano:

$$\text{probabilidade posterior} = \text{verossimilhança} \times \text{probabilidade inicial}$$

A evidência, em ambos os métodos estatísticos, é formalizada e quantificada. Tradicionalmente, os filósofos não costumam se ocupar com a discussão filosófica dos métodos estatísticos. Na maioria das vezes, o debate é restrito aos fundamentos da probabilidade, aos paradoxos e discussões de casos especiais e hipotéticos. No entanto, existe não apenas espaço para debater questões filosóficas no campo da estatística, mas consideramos este debate necessário, conforme exemplo de resultados estatísticos em teste de uma vacina: com um método estatístico chega-se à conclusão de que vacina tem eficácia, e no outro método, conclui-se que é bem pouco provável que ela tenha eficácia.

A medida Q de Popper, $Q = P(E|H \cap K) - P(E|K)$, afirma que o seu valor é maior quanto maior a "severidade" dos testes aos quais uma hipótese foi submetida, porém, Popper não avança na avaliação da metodologia dos testes. Um teste pode ser considerado severo para um grupo de pessoas que defende determinado paradigma da estatística, e, no entanto, apresentar falhas ou até mesmo ser inválido no caso de outro paradigma. Os métodos bayesianos diferem da concepção da filosofia da ciência de Popper por vários motivos e nesta Dissertação nos concentramos nas seguintes diferenças:

1) Uma hipótese pode ser confirmada pela evidência, e não apenas falsificada;

2) O suporte probabilístico é indutivo e não dedutivo, pois a evidência tem caráter indutivo. A indução, portanto, faz parte dos métodos científicos bayesianos.

Da mesma forma que os criptoanalistas em Bletchley Park sabiam que a probabilidade não poderia ser uma confirmação ou corroboração muito antes da afirmação feita por Popper, existem reflexões e resultados importantes do ponto de vista filosófico sendo produzidos pelos estatísticos, fora do âmbito filosófico. Por este motivo, é importante que os filósofos leiam os textos estatísticos e possam dialogar com eles tanto a respeito dos diferentes métodos estatísticos como a respeito de conceitos como corroboração de hipóteses, evidência científica e métodos indutivos. Devido à crescente complexidade e variedade dos métodos estatísticos existentes, grupos de pesquisa interdisciplinares podem ser um primeiro passo nesta direção.

O *peso da evidência* teve uma aplicação prática para um problema de difícil resolução à época, e o seu cálculo foi eficiente para corroborar as hipóteses que estavam sendo testadas. Além disso, o *peso da evidência* não se restringe apenas à criptografia, pode ser aplicado a outros casos. A sua característica de aditividade consegue capturar a intuição de cada evidência apresentar um peso a favor ou contra uma hipótese e a sua somatória consegue definir a corroboração ou não de uma determinada hipótese.

Talvez o *peso da evidência* não seja a melhor medida existente para o grau de corroboração e a sua comparação com outros métodos é uma pesquisa que pode ser desenvolvida futuramente. Stern e Pereira (2013), por exemplo, desenvolvem o método bayesiano FBST (*Full Bayesian Significant Test*)[63], um teste que apresenta uma medida chamada *e*-valor ou valor epistêmico, $ev(H)$. Uma característica fundamental desse tipo de teste é que ele

> faz uma distinção clara entre o espaço da hipótese e o espaço do parâmetro, adotando medidas distintas para cada um deles, nomeadamente, a medida natural de probabilidade posterior bayesiana no espaço do parâmetro, e a medida possibilística do o *e*-valor no espaço da hipótese. (STERN, 2011, p. 644, tradução nossa).

[63] A primeira proposta deste tipo de teste consta em Pereira e Stern (1999).

Além desta pesquisa, também há a possibilidade de trabalhar, como mencionamos na seção 2.4, com o estabelecimento de uma relação entre o conceito de *peso da evidência* e a lógica da inconsistência formal, no sentido de atribuir graus de evidência a proposições que envolvam contradições.

REFERÊNCIAS

BANKS, D. L. A conversation with I. J. Good. *Statistical Science*, Vol. 11, No. 1 (Feb., 1996), p. 1-19.

BRADY, M. E. J. M. Keynes's theoretical approach to decision-making under conditions of risk and uncertainty. *The British Journal for the Philosophy of Science*, Vol. 44, No. 2 (Jun., 1993), p. 357-376.

BUENO-SOLER, J.; CARNIELLI, W. Paraconsistent probabilities: Consistency, contradictions and Bayes' theorem. *In*: STERN, J. (Ed), *Special Issue Statistical Significance and the Logic of Hypothesis Testing. Entropy* 18(9), 2016.

CARD, W. I.; GOOD, I. J. A logical analysis of medicine. *In*: PASSMORE, R.; ROBSON, J. S. (Eds.). *A Companion to Medical Studies.* 3, Chapter 60. Oxford: Blackwell, 1974.

CARNAP, R. *Logical foundations of probability.* Chicago: University of Chicago Press, 1950.

_____. *The continuum of inductive methods.* Chicago: The University of Chicago Press, 1952.

_____. A basic system of inductive logic, part II. *In*: JEFFREY, R. C. (Ed). *Studies in Inductive Logic and Probability,* volume II. Berkeley-Los Angeles-London: University of California Press, 1980. p. 7-156.

CARNIELLI, W.; BUENO-SOLER, J. Paraconsistent probabilities, their significance and their uses. *In*: GOUVEIA, P.; CALEIRO, C.; DONISIO, F. (Eds), *Logic and Computation: Essays in Honour of Amilcar Sernadas*, p. 197–230. College Publications, 2017.

CARNIELLI, W.; EPSTEIN, R. L. *Computabilidade, funções computáveis, lógica e os fundamentos da matemática.* 2.ed. revista. São Paulo: Editora UNESP, 2005.

CARNIELLI, W.; RODRIGUES, A. An epistemic approach to paraconsistency: a logic of evidence and truth. *Synthese*, 196:3789–3813, 2017.

_____. On epistemic and ontological interpretations of intuitionistic and paraconsistent paradigms. *Logic Journal of the IGPL*, 2019.

COURSERA. *Bayesian statistics:* from concept to data analysis. 2019. Disponível em: <https://www.coursera.org/learn/bayesian-statistics>. Acesso em: 13 set. 2019.

DAVIS, M. *Computability and undecidability.* New York: McGraw-Hill, 1958.

DEGROOT, M. H.; SCHERVISH, M. J. *Probability and statistics.* 4th ed. Boston: Addison-Wesley, 2012.

DEVLIN, K. *Introduction to mathematical thinking.* Palo Alto: Keith Devlin, 2012.

EDWARDS, A. W. (1992). *Likelihood.* Expanded ed. Baltimore: Johns Hopkins University Press, 1992.

FITELSON, B. A Bayesian Account of Independent Evidence with Applications. *Philosophy of Science* 68, 2001, p. S123-140.

FISHBURN, P. C. The axioms of subjective probability. *Statistical Science*, 1986, Vol. 1, No. 3, 3:35-358.

GALAVOTTI, M. C. *Philosophical introduction to probability.* Stanford: CSLI Publications, 2005.

_____. Confirmation, probability, and logical empiricism. *In*: RICHARDSON, A.; UEBEL, T. (Eds). *The Cambridge Companion to Logical Empiricism.* New York: Cambridge University Press, 2007. p. 117-135.

_____. Probability and pragmatism. *In*: DIEKS, D.; GONZALEZ, W. J.; HARTMANN, S.; UEBEL, T.; WEBER, M. (Eds). *Explanation, Prediction, and Confirmation*. Dordrecht: Springer, 2011. p. 499-510.

_____. Probabilistic epistemology: a european tradition. *In*: GALAVOTTI, M. C.; NEMETH, E.; STADLER, F. (Eds). *European Philosophy of Science* – Philosophy of Science in Europe and the Viennese Heritage. Cham: Springer, 2014. p. 77-88.

GIBBS, J. W. *Elementary principles in statistical mechanics*. London: Constable, 1902. Dover reprint 1960.

GILLIES, D. In defense of the Popper-Miller argument. *Philosophy of Science* 53, 1986, p. 110-113.

_____. The Turing-Good weight of evidence function and Popper's measure of the severity of a test. *The British Journal for the Philosophy of Science*, Vol. 41, No. 1 (Mar., 1990), p. 143-146.

GLANVILL, J. *The vanity of dogmatizing, or, Confidence in opinions manifested in a discourse of the shortness and uncertainty of our knowledge, and its causes*: with some reflexions on peripateticism, and an apology for philosophy. Hove: The Harvester Press, 1970.

GOOD, I. J. *Probability and the weighing of evidence*. London: Charles Griffin, 1950.

_____. The Population Frequencies of Species and the Estimation of Population Parameters. *Biometrika*, 40, 1953, p. 237-264.

_____. Review of a paper by K. R. Popper. *Math. Rev.*, 16, 1955a, 376.

_____. Some notation and terminology of information theory. *Proc. Inst. Elec. Engrs. C*, 103, 1955b, 200-205, or Monograph No. 155R.

_____. Weight of Evidence, Corroboration, Explanatory Power, Information and the Utility of Experiments. *Journal of the Royal Statistical Society*. Series B (Methodological), Vol. 22, No. 2 (1960). p. 319-331.

_____. Weight of evidence, causality, and false-alarm probabilities. *Information Theory, Fourth London Symposium*. London: Butterworth, 1961.

_____. Corroboration, explanation, evolving probability, simplicity and a sharpened razor. *The British Journal for the Philosophy of Science*, Vol. 19, No. 2 (Aug., 1968), p. 123-143.

_____. Explicativity, corroboration, and the relative odds of hypotheses. *Synthese*, Vol. 30, No. 1/2, Methodologies: Bayesian and Popperian (Feb. - Mar., 1975), p. 39-73.

_____. Studies in the history of probability and statistics. XXXVII A. M. Turing's statistical work in World War II. *Biometrika*, Vol. 66, No. 2 (Aug., 1979), p. 393-396.

_____. An error by Peirce concerning "weight of evidence". *Journal of Statistical Computation and Simulation*, 13:2, 1981, p. 155-157.

_____. *Good thinking:* the foundations of probability and its applications. Minneapolis: University of Minnesota Press, 1983.

_____. A bayesian approach in the philosophy of inference. *The British Journal for the Philosophy of Science*, Volume 35, Issue 2, June 1984, p. 161–166. 1984a.

_____. Letter concerning Popper and Miller (1983). *Nature* 310, August 2 1984b, p. 433.

_____. The best explicatum for weight of evidence. *C197 in Journal of Statistical Computation and Simulation*, 1984c. 19: 294-299; 20: 89.

_____. Weight of evidence: a brief survey. *In*: BERNARDO, J. M.; DEGROOT, M. H.; LINDLEY, D. V.; SMITH, A. F. M. (eds.), *Bayesian Statistics 2:* Proceedings of the Second Valencia International Meeting, September 1983. New York: North-Holland, 1985a. p. 249-269.

_____. Probabilistic induction is inevitable. *C216 in Journal of Statistical Computation and Simulation*, 1985b. 20: p. 323-324.

_____. The whole truth. *IMS Bull*. 15, 1986, p. 366-373.

_____. A reinstatement, in response to Gillies, of Redhead's argument in support of induction. *Philosophy of Science*, Vol. 54, No. 3 (Sep., 1987), p. 470-472.

_____. The interface between statistics and philosophy of science. *Statistical Science*, Vol. 3, nº 4 (Nov., 1988), p. 386-397.

_____. A suspicious feature of the Popper/Miller argument. *Philosophy of Science*, Vol. 57, No. 3 (Sep., 1990), p. 535-536.

_____. Enigma and Fish. *In*: HINSLY, F. H.; STRIPP, A. (Eds). *Codebreakers*: the inside story of Bletchley Park. Oxford University Press, 1993. p. 149-166.

_____. Turing's anticipation of empirical bayes in connection with the cryptanalysis of the naval enigma. *Journal of Statistical Computation and Simulation*, 2000, Vol.66:2, p. 101-111.

GOOD, I. J.; CARD, W. I. The diagnostic process with special reference to errors. *Methods Information Med.* 10, 1971, p. 176-188.

GOOD, I. J.; TOULMIN, G. H. Coding theorems and weight of evidence. *J. Ins. Maths Applics*. 1968, 4, p. 94-105.

GOODE, T. M.; LOEWER, B. M., ROSENKRANTZ, R. D.; WETTERSTEN, J. R. Editorial introduction. *Synthese*, Vol. 30, n. 1/2, Methodologies: Bayesian and Popperian (Feb. – Mar., 1975), p. 1.

HACKING, I. *The emergence of probability:* a philosophical study of early ideas about probability, induction and statistical inference. 2nd ed. New York: Cambridge University Press, 2006.

HINSLY, F. H.; STRIPP, A. (Eds). *Codebreakers*: the inside story of Bletchley Park. Oxford University Press, 1993.

HORWICH, P. *Probability and evidence*. New York: Cambridge University Press, 1982.

HOWSON, C.; URBACH, P. *Scientific reasoning: the bayesian approach*. La Salle: Open Court, 1989.

HUME, D. *Philosophical Essays Concerning Human Understanding*. 2nd ed. London: M. Buckle, 1751.

_____. *Investigações sobre o entendimento humano e sobre os princípios da moral*. Tradução de José Oscar Almeida Marques. São Paulo: UNESP, 2004.

HUZURBAZAR, V. S. On the certainty of an inductive inference. *Proc. Cambridge Philos. Soc*. 51, 1955, p. 761-762.

JEFFREY, R. C. Letter concerning Popper and Miller (1983)**.** *Nature* 310 August 2 1984, p. 433.

JEFFREYS, H. *Theory of probability*. Oxford: Clarendon, 1939.

KASSER, J. Confidence, evidential weight, and the theory-practice divide in Peirce. *Transactions of the Charles S. Peirce Society*, Vol. 52, No. 2 (Spring 2016), p. 285-308.

KAHN, D. *The codebreakers*: the comprehensive history of secret communication from ancient times to the internet. New York: Scribner, 1996.

KEMENY, J.C. & OPPENHEIM, P. Degrees of factual support. *Philosophy of Science*, 19, 1952. p. 307-324.

KEYNES, J. M. *A treatise on probability*. London: Macmillan and Co., Limited, 1921.

KOLMOGOROV, A. N. *Foundations of the theory of probability*. New York: Chelsea Publishing Company, 1950.

KULLBACK, S. *Information theory and statistics*. New York: Wiley, 1959.

LEIBNIZ, G. W. F. Die philosophischen Schriften von G. W. Leibniz. *In*: GEHARDT, C. I. (ed.). Berlim, 1875-90, 7 vols apud HACKING, I. *The Emergence of Probability:* A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference. 2nd ed. New York: Cambridge University Press, 2006.

LEVI, I. *Gambling with truth* – an essay on induction and the aims of science. Cambridge: The MIT Press, 1967.

_____. Letter concerning Popper and Miller (1983)**.** *Nature,* 310, August 2 1984, p. 433.

LINDLEY, D. V. Statistical inference (with discussion). *JRSS*, Series B (15), 1953, p. 30-76.

_____. On a measure of the information provided by an experiment. *Ann. Math. Statist.*, 27, 1956, p. 986-1005.

_____. Good's work in probability, statistics and the philosophy of science. *J. Statist. Plann. Inference* 25, 1990, p. 211-223.

_____. The philosophy of statistics. *Journal of the Royal Statistical Society*. Series D (The Statistician), Vol. 49, nº 3, 2000. p. 293-337.

MALLOWS, C. L. The information in an experiment. *J. R. Statist. Soc.,* B, 21, 1959, p. 67-72.

MARTIN, E. A. Review: Pure inductive logic. *MathSciNet*. MR3379655, 03-02 03B48 60A05, 2016.

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. New Haven & London: Yale University Press, 2011.

MILNER-BARRY, S. Hut 6: early days. *In*: HINSLY, F. H.; STRIPP, A. (Eds). *Codebreakers:* the inside story of Bletchley Park. Oxford University Press, 1993. p. 89-99.

MINSKY, M.; SELFRIDGE. 0. G. Learning in random nets. *In*: *Information Theory*: Fourth London Symposium (Colin Cherry, ed.). London: Butterworths, 1961. p. 335-347.

O'HAGAN, A.; FORSTER, J. *Kendall's advanced theory of statistics*, Vol. 2B: Bayesian inference. 2nd ed. London: Arnold, 2004.

PARIS, J. B.; VENCOVSKÁ, A. *Pure inductive logic*. Cambridge: Cambridge University Press, 2015.

PEREIRA, C. A. B.; STERN, J. M. Evidence and credibility: full Bayesian significance test for precise hypotheses. *Entropy Journal*, 1, 1999, p. 69-80.

PETERSON, W. W.; BIRDSALL, T. G.; FOX, W. C. The theory of signal detectability. *Trans. Inst. Radio Engrs*. PGIT-4, 1954, p. 171-212.

PEIRCE, C. S. (1878). The probability of induction. *In*: NEWMAN, J. R. (Ed). *The World of Mathematics 2*. New York: Simon and Shuster, 1956. p. 1341-1354.

_____. *The Essential Peirce v.1*. Ed. by N. Houser and C. Kloesel. Bloomington, Ind.: Indiana University Press, 1992.

POPPER, K. Two autonomous axiom systems for the calculus of probabilities. *The British Journal for the Philosophy of Science*, 6, 1955. p. 51-57.

_____. *The logic of scientific discovery*. London: Hutchinson & Co, 1959.

_____. The non-existence of probabilistic inductive support. *In*: Dorn G., Weingartner P. (eds) *Foundations of Logic and Linguistics*. Boston: Springer, 1985. p. 303-318.

POPPER, K.; MILLER, D. A proof of the impossibility of inductive probability, *Nature*, 1983. 302: p. 687-688.

_____. Reply to Levi (1984), Jeffrey (1984) and Good (1984). *Nature*, Lond., 1984. 310, p. 434.

_____. Why probabilistic support is not inductive. *Philosophical Transactions of the Royal Society*, London, 1987. A 321: p. 569-591.

PRICE, R. *Observations on reversionary payments on schemes for providing annuities for widows, and for persons in old age; on the method of calculating the values of assurances on lives; and on the national debt*. 1771. Disponível em: <https://openlibrary.org/books/OL7198407M/Observations_on_reversionary_payments>. Acesso em: 23 jan. 2021.

REDHEAD, M. L. G. On the impossibility of inductive probability. *British Journal for the Philosophy of Science* 36, 1985, p. 185-191.

RÉNYI, A. On a new axiomatic theory of probability. *Acta Mathematica Academiae Scientiarum Hungaricae*, 6, 1955, p. 286-335.

RICE, J. A. *Mathematical statistics and data analysis*. 2nd ed. Belmont: Wadsworth Publishing Company, 1995.

RODRIGUES, A.; CARNIELLI, W. A.; BUENO-SOLER, J. Measuring evidence: a probabilistic approach to an extension of Belnap-Dunn logic. *Synthese*, 2020, no prelo.

SAVAGE, L. J. *The foundations of statistics*. New York: Dover Publications, 1954.

SHANNON, C. E. A mathematical theory of communication. *Bell System Technical Journal* (27). p. 379-423, 623-56, 1948.

SIMPSON, E. Edward Simpson: Bayes at Bletchley Park. *Significance*, Vol. 7, Issue 2, June 2010. p. 76-80.

SKRYMS, B. *Choice and chance* – an introduction to inductive logic. Belmont: Dickenson Publishing Company, 1966.

SMITH, A. A conversation with Dennis Lindley. *Statistical Science*, Vol. 10, nº 3, Aug., 1995, p. 305-319.

SPIEGELHALTER, D. J.; KNILL-JONES, R. P. Statistical and knowledge-based approaches to clinical decision-support systems, with an application to gastroenterology (with discussion). *J. Roy. Statist. Soc*. Ser. A., 147, 1984, p. 1-34.

STERN, J. M. Constructive verification, empirical induction, and falibilist deduction: a threefold contrast. *Information* 2011, 2, p. 635-650.

STERN, J. M.; PEREIRA, C. A. B. Bayesian epistemic values: focus on surprise, measure probability! *Logic Journal of IGPL* 21(2), August 2013.

STIGLER, S. M. The Epic Story of Maximum Likelihood. *Statistical Science* 2007, Vol. 22, No. 4, p. 598–620.

STRIPP, A. The Enigma machine: its mechanism and use. *In*: HINSLY, F. H.; STRIPP, A. (Eds). *Codebreakers: the inside story of Bletchley Park*. Oxford University Press, 1993. p. 81-88.

THOMAS, D. O.; PEACH, B. *The Correspondence of Richard Price*. Volume 1, July 1748–March 1778. Duke Univ. Press, 1983. p. 45–47.

TRIBUS, M. *Rational Descriptions, Decisions and Designs*. New York: Pergamon Press, 1969.

WACKERLY, D. D.; MENDENHALL III, W.; SCHEAFFER, R. L. *Mathematical statistics with applications*. 7th ed. Belmont: Thomson, 2008.

WALD, A**.** On cumulative sums of random variables. *Ann. Math. Statist*. 15, 1944, p. 283-96.

_____. *Sequential analysis***.** New York: Wiley, 1947.

WELCHMAN, G. From Polish bomba do British bombe: the birth of ultra (1986). In: WELCHMAN, G. *The hut 6 story*: breaking the Enigma codes. Witney: M&M Baldwin, 2019.

WRINCH, D.; JEFFREYS, H. On certain fundamental principles of scientific inquiry. *Philosophical Magazine*, Series 6, 42, 1921, p. 369-390.

ZABELL, S. L. Alan Turing and the Central Limit Theorem. *The American Mathematical Monthly*, Vol. 102, No. 6 (Jun. - Jul., 1995), p. 483- 494.

APÊNDICE

A. Axiomas da probabilidade: probabilidade absoluta e probabilidade relativa

Vamos definir dois sistemas de axiomas: os axiomas da probabilidade absoluta e os da probabilidade relativa. Os axiomas da probabilidade são considerados os fundamentos da estatística, a partir dos quais podemos fazer todos os cálculos de probabilidade – como derivar teoremas. Os axiomas da probabilidade relativa podem ser derivados a partir dos axiomas da probabilidade absoluta. Historicamente, o sistema de axiomas da probabilidade absoluta é definido mais cedo, sendo o mais conhecido deles o sistema de Kolmogorov (publicado originalmente em 1933, já a tradução para a língua inglesa ocorre em 1950).[64] Os axiomas da probabilidade relativa são definidos mais tarde, por Carnap (1952), Rényi (1955) e Popper (1955).

A escolha que fizemos dos axiomas deve-se a dois motivos principais: a simplicidade e a clareza. A simplicidade dos axiomas é apontada por Lindley (2000, p. 298) para a sua escolha dos axiomas da probabilidade relativa e seguimos, de igual modo, este princípio também para os axiomas da probabilidade absoluta. Além disso, nossa escolha dos axiomas nos pareceu conferir uma melhor compreensão dos conceitos apresentados na Dissertação. Muitos sistemas de axiomas acabam sendo equivalentes e para um estudo mais aprofundado do assunto temos várias recomendações.[65]

A.1 Probabilidade absoluta

Vamos seguir então com os axiomas, definições e alguns teoremas da probabilidade absoluta de acordo com DeGroot e Schervish (2012, p. 16-18). O que nos pareceu um diferencial neste sistema em relação ao de Kolmogorov é que ele é mais sucinto. Os autores definem apenas três axiomas – não-negatividade, normalização e adição infinita, enquanto no sistema de Komogorov são necessários quatro axiomas – não-negatividade, normalização, adição finita e adição infinita (esse último é denominado "Axioma da Continuidade" por

[64] Segundo Galavotti (2014, p. 79), antes da publicação de Kolmogorov, há registro de uma publicação de axiomatização da probabilidade em 1932 feita pelo matemático Stefan Mazurkiewicz: Zur Axiomatik der Wahrscheinlichkeitsrechnung, *In*: *Comptes rendues des séances de la Société des Sciences et des Lettres de Varsovie*, Classe III, 25, 1932, p. 1-4.

[65] Para os axiomas da probabilidade subjetiva, por exemplo, pode-se consultar Fishburn (1986).

Kolmogorov). O axioma de Kolmogorov da adição finita é, no sistema de DeGroot e Schervish, um teorema derivado da adição infinita.

Em primeiro lugar, vamos considerar um *experimento*, ou seja, um "processo, real ou hipotético, no qual os possíveis resultados podem ser identificados de maneira prévia" (DEGROOT.; SCHERVISH, 2012, p. 5). O experimento pode ser, por exemplo, um lançamento de um dado ou de uma moeda (ou uma série de lançamentos), um sorteio de um conjunto de números numa loteria, uma estimativa do valor de compra do dólar no dia seguinte, ou uma estimativa da quantidade de peças defeituosas numa linha de produção industrial.

O conjunto de todos os eventos possíveis que podem ocorrer como resultado de um experimento qualquer é chamado de *espaço amostral* (Ω). Um evento é um subconjunto do espaço amostral e será denotado por letras maiúsculas. No caso do lançamento do dado, o espaço amostral é o conjunto das possibilidades do lançamento, se formos considerar um dado de 6 lados, o espaço amostral terá 6 elementos, ou 6 eventos possíveis:

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Vamos utilizar letras maiúsculas para nos referir a um evento possível qualquer $(A, B, C, \dots)$. Para cada evento é atribuída uma probabilidade, $P(A)$. A atribuição dessa probabilidade deve respeitar os três axiomas seguintes:

Axioma 1 (não-negatividade):

$$P(A) \geq 0.$$

Axioma 2 (normalização):

$$P(\Omega) = 1.$$

Axioma 3 (adição infinita):

para toda sequência infinita de eventos disjuntos A_1, A_2, ...,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Definição 1: Uma probabilidade $P(A)$ é qualquer número correspondente a um evento no espaço amostral Ω que satisfaça os axiomas 1, 2 e 3.

Teorema 1: a probabilidade do conjunto vazio é igual a zero.

$$P(\emptyset) = 0.$$

Isso significa que, se o evento for impossível, sua probabilidade será igual a zero. Notamos, no entanto, que o inverso não é verdadeiro: uma probabilidade pode ser igual a zero e o evento não ser impossível de ocorrer.

Seguimos com a prova do teorema. Seja um conjunto infinito de eventos $A_1, A_2, ...$ tal que $A_i = \emptyset$ para todo $i = 1, 2, ...$ e considerando o axioma 3:

$$P(\emptyset) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} P(\emptyset).$$

Teorema 2 (adição finita):

para toda sequência finita de eventos disjuntos A_1, A_2, ...,

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

A partir do axioma 3 e do Teorema 1, podemos dividir o conjunto infinito em duas somatórias, sendo uma delas a somatória da sequência A_1, A_2, ..., para $i = 1, \dots, n$ e a outra a somatória de $A_i = \emptyset$ para $i > n$:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

$$= \sum_{i=1}^{n} P(A_i) + \sum_{i=1}^{\infty} P(A_i)$$

$$= \sum_{i=1}^{n} P(A_i) + 0$$

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

Vamos apresentar mais alguns teoremas fundamentais da probabilidade. Dado que este Apêndice foi elaborado para esclarecer e complementar o texto principal que tem como objetivo discutir o conceito de *peso da evidência* de acordo com Good, vamos adaptar a notação tanto de Good quanto dos textos de estatística e probabilidade para ter uma notação única que não confunda o leitor.

Desta maneira, o evento que na teoria de conjuntos conhecemos como A^c, que lemos "A complementar", tem equivalência com $\sim A$ ("não A") e com $\overline{A}$, que podemos ler como "A barra". Essa última notação é utilizada por Good, no sentido de ser um evento disjunto de A e, portanto, "concorrente": o *peso da evidência* é uma medida que pondera o quanto a evidência E favorece uma hipótese H em relação a sua hipótese concorrente, $\overline{H}$. Os teoremas a seguir seguirão essa adaptação de notação.

Teorema 3:

para qualquer evento A, $P(\overline{A}) = 1 - P(A)$.

Teorema 4:

se $A \subset B$, então $P(A) \leq P(B)$.

Teorema 5:

para qualquer evento A, $0 \leq P(A) \leq 1$.

Teorema 6:

para todos os eventos A e B,

$$P(A \cap \overline{B}) = P(A) - P(A \cup B).$$

Teorema 7:

para todos os eventos A e B,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

A.2. Probabilidade relativa

Optamos por reproduzir o sistema de axiomas da probabilidade relativa de acordo com Lindley (2000). Ressaltamos que Lindley não denomina o seu sistema como um sistema de axiomas, mas um sistema com "regras", que de acordo com ele, "são deduções, exceto pela regra 4, de outras, mais básicas, presunções." (Lindley, 2000, p. 299). Porém, aqui faremos algumas adaptações, entre elas denominar as regras de Lindley como axiomas.

Lindley, assim como outros estatísticos bayesianos, parte dos axiomas da probabilidade relativa e não da probabilidade absoluta. Isso porque a "probabilidade é sempre condicional (...). Ela depende de dois argumentos: o elemento cuja incerteza está sendo descrita e o conhecimento no qual essa incerteza é baseada"[66]. Ou, mais sucintamente, a "probabilidade depende de dois elementos: a incerteza do evento e as condições sob as quais você o está considerando"[67].

Axioma 1 (convexidade):

para todo evento A e B,
$0 \leq P(A|B) \leq 1$,
$P(A|A) = 1$,
e $P(A|B) = 1$ se A é uma consequência lógica de B.

Axioma 2 (adição):

se A e B são exclusivos, dado C,

$$P(A \cup B|C) = P(A|C) + P(B|C).$$

[66] Ibid., p. 301-302.
[67] Ibid., p. 298.

Axioma 3 (multiplicação):

para todo A, B e C,

$$P(AB|C) = P(A|BC).P(B|C).$$

Axioma 4 (conglomerabilidade):

se $\{B_n\}$ é uma partição, possivelmente infinita, de C e $P(A|B_nC) = k$, o mesmo valor para todos n, então $P(A|C) = k$.

Também podemos derivar alguns teoremas de maneira semelhante às derivações da probabilidade absoluta:

Teorema 1:

para todo evento A e B,

$$P\left(\overline{A}\middle|B\right) = 1 - P(A|B).$$

B. Transformação algébrica: Razão bayesiana de verossimilhança em Fator Bayes-Jeffreys-Turing

Seguimos aqui com a transformação algébrica da seguinte igualdade:

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

A primeira equação é a razão bayesiana de verossimilhança, e a segunda equação o fator Bayes-Jeffreys-Turing.

Vamos começar transformando o fator Bayes-Jeffreys-Turing, que é representado por chance (*odds*) em probabilidade.

Considerando a relação entre chance e probabilidade, $O(A) = (P(A)/(1 - P(A)))$, temos:

$$\frac{O(H|E)}{O(H)} = \frac{P(H|E)}{1 - P(H|E)} \div \frac{P(H)}{1 - P(H)}$$

Aplicando o teorema 1 da probabilidade relativa e o teorema 3 da probabilidade absoluta, temos:

$$\frac{P(H|E)}{P(\overline{H}|E)} \div \frac{P(H)}{P(\overline{H})}$$

Agora, aplicando o teorema de Bayes no nominador da divisão acima:

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

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

$$\frac{P(E|H).P(H)}{P(E)} \div \frac{P(E|\overline{H}).P(\overline{H})}{P(E)}$$

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

E, finalmente, finalizamos a transformação chegando na fórmula da razão bayesiana de verossimilhança:

$$\frac{P(E|H).P(H)}{P(E|\overline{H}).P(\overline{H})} \div \frac{P(H)}{P(\overline{H})}$$

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