



Universidade do Estado do Rio de Janeiro

Centro de Tecnologia e Ciências

Instituto de Matemática e Estatística

Monica Almeida Gama

Probabilidade: uma abordagem mais próxima do cotidiano

Rio de Janeiro

2018

Monica Almeida Gama

Probabilidade: uma abordagem mais próxima do cotidiano



Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Mestrado Profissional em Matemática em Rede Nacional – PROFMAT, da Universidade do Estado do Rio de Janeiro.

Orientadora: Prof.^a Dra. Patrícia Nunes da Silva
Coorientador: Prof. Dr. André Luiz Cordeiro dos Santos

Rio de Janeiro

2018

CATALOGAÇÃO NA FONTE
UERJ / REDE SIRIUS / BIBLIOTECA CTC-A

G184 Gama, Mônica Almeida.
Probabilidade: uma abordagem mais próxima do cotidiano / Mônica Almeida Gama – 2018.
99f. : il.

Orientadora: Patrícia Nunes da Silva.
Coorientador: André Luis Cordeiro dos Santos.
Dissertação (Mestrado Profissional em Matemática em Rede Nacional - PROFMAT) - Universidade do Estado do Rio de Janeiro, Instituto de Matemática e Estatística.

1. Probabilidades - Teses. 2. Análise combinatória - Teses. I. Silva, Patrícia Nunes da. II. Santos, André Luis Cordeiro dos. III. Universidade do Estado do Rio de Janeiro. Instituto de Matemática e Estatística. IV. Título.

CDU 519.2

Patrícia Bello Meijinhos CRB-7/5217 – Bibliotecária responsável pela elaboração da ficha catalográfica

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação, desde que citada a fonte.

Assinatura

Data

Monica Almeida Gama

Probabilidade: uma abordagem mais próxima do cotidiano

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Mestrado Profissional em Matemática em Rede Nacional – PROFMAT, da Universidade do Estado do Rio de Janeiro.

Aprovada em 24 de agosto de 2018.

Banca Examinadora:

Prof.^a Dra. Patrícia Nunes da Silva (Orientadora)
Instituto de Matemática e Estatística – UERJ

Prof. Dr. André Luiz Cordeiro dos Santos (Coorientador)
Centro Federal de Educação Tecnológica Celso Suckow da Fonseca

Prof. Dr. Helvécio Rubens Crippa
Instituto de Matemática e Estatística – UERJ

Prof. Dr. Juscelino Bezerra dos Santos
Escola Nacional de Ciências Estatísticas

Rio de Janeiro

2018

DEDICATÓRIA

Dedico essa dissertação aos meus pais, Isabel Cabral e Wilson Pinho, pelo carinho e incentivo. Por sempre me impulsionarem para longe da inércia.

Dedico também ao meu amado marido, Arlan Almeida Sousa, companheiro leal com o qual a vida me presenteou.

AGRADECIMENTOS

Agradeço aos meus amigos, Tatiana e Glauco, por me aconselharem a seguir a minha carreira acadêmica. Essa jornada não teria começado sem o incentivo de vocês. Nossa conversa foi um divisor de águas para mim.

Agradeço a ajuda da minha turma do curso de Mestrado. Sem ela não seria possível a conclusão deste sonho. Agradeço, em especial, ao meu amigo, Roberto Nascimento, por todas as vezes que teve paciência em explicar o conteúdo para mim. Por sua simplicidade e altruísmo.

Agradeço a minha orientadora Patrícia Nunes da Silva por sua energia constante, por não desistir de mim, por sonhar comigo novos desafios. Entrei neste curso buscando o título e saí com muito mais. Hoje tenho importantes perspectivas e uma amiga.

Agradeço ao meu coorientador André Luiz Cordeiro dos Santos por ter melhorado consideravelmente este trabalho e o meu desempenho nesta jornada.

Agradeço a todos os parentes e amigos que estiveram na torcida sincera pelo meu sucesso.

RESUMO

GAMA, Monica Almeida. *Probabilidade: uma abordagem mais próxima do cotidiano*. 2018. 99 f. Dissertação (Mestrado Profissional em Matemática em Rede Nacional – PROFMAT) – Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2018.

Esta dissertação analisou alguns problemas apresentados no livro “O Andar do Bêbado” de Leonard Mlodinow. Preencheu lacunas, pois se propôs a explicar o cálculo realizado nos exemplos propostos pelo autor. Em alguns casos, houve no livro a omissão de passagens do enunciado, dificultando o entendimento do que se pretendia mostrar. O tema principal é probabilidade. A forma como se apresenta no cotidiano, contrariando o imaginário popular de que é possível se distanciar da Matemática no dia-a-dia. Este trabalho se dispôs a quebrar o paradigma de que é preciso realizar cálculos para se entender probabilidade. Em diversos momentos é possível analisar situações nas quais não se percebe a Matemática. Foi possível verificar como a intuição pode ser uma aliada para o entendimento da probabilidade, mas pode prejudicá-lo em outras ocasiões. Mostrou como as pessoas se posicionam diante do acaso. Assim como há indícios de que não entendem o conceito de aleatório, uma vez que buscam padrões onde não há. É o caso do conceito de “matching”. Um aparente ganho no desenvolvimento cognitivo da espécie humana que é mal empregado quando estamos diante do acaso.

Palavras-chave: Probabilidade. Combinatória. “Matching”.

ABSTRACT

GAMA, Monica Almeida. *Probability: a closer approach to everyday life*. 2018. 99 f. Dissertação (Mestrado Profissional em Matemática em Rede Nacional – PROFMAT) – Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2018.

This dissertation analyzed some problems presented in Leonard Mlodinow's book "The Drunk Walk". It filled in gaps because it set out to explain the calculation made in the examples proposed by the author. In some cases, the book omitted passages of the utterance, making it difficult to understand what it was intended to show. The main theme is probability. The way it presents itself daily, contrary to the popular imagination that it is possible to distance yourself from mathematics in everyday life. This paper set out to break the paradigm that one must perform calculations to understand probability. At various times it is possible to analyze situations in which mathematics is not perceived. It was possible to see how intuition can be an ally for understanding probability, but it can harm it on other occasions. It showed how people position themselves in the face of chance. As well as there is evidence that they do not understand the concept of random as they seek patterns where there are none. This is the case with the concept of "matching". An apparent gain in the cognitive development of the human species that is misused when faced with chance.

Keywords: Probability. Combinatorics. Matching.

LISTA DE FIGURAS

Figura 1	- Roleta	26
Figura 2	- Alvo	27
Figura 3	- Astrágalos	28
Figura 4	- Relação entre conjuntos	44
Figura 5	- Abordagem geométrica	46
Figura 6	- Os pontos alcançáveis estão em preto.	65
Figura 7	- Alguns caminhos até $(5, 1)$	66
Figura 8	- Exemplo do Princípio de Reflexão	68
Figura 9	- Caminho típico	71
Figura 10	- Caminho s	72
Figura 11	- Caminho s' partindo da nova origem $(1, 1)$	73
Figura 12	- O caminho s não pertence a \mathcal{N}_{2m}	74
Figura 13	- Construção do caminho s'	74
Figura 14	- O caminho s'	75
Figura 15	- Última época em que $s_t = \bar{k}$	76
Figura 16	- Deslize o trecho refletido até a origem	76
Figura 17	- O caminho $s' = F^{-1}(s)$	77
Figura 18	- Vitória de Peter	82
Figura 19	- Derrota de Peter	82
Figura 20	- “Permanecer na liderança por metade do tempo”	88
Figura 21	- “Permanecer na liderança por metade do tempo”	88
Figura 22	- Princípio da reflexão, $y = -2$	90
Figura 23	- Princípio da reflexão, $y = 2$	91
Figura 24	- Sistema de coordenadas anterior: origem em $(1, 1)$	92
Figura 25	- Indução em c	93

LISTA DE TABELAS

Tabela 1	- Caso dos Collins	42
Tabela 2	- Frequência dos primeiros dígitos significativos	57
Tabela 3	- Frequência dos primeiros dígitos significativos nos balancetes	58
Tabela 4	- Frequência dos primeiros dígitos significativos no balancete de abril de 2017	59

SUMÁRIO

	INTRODUÇÃO	10
1	O QUE É PROBABILIDADE?	12
2	ESPAÇO AMOSTRAL	21
2.1	Tudo a seu tempo	21
2.1.1	<u>Espaços Amostrais</u>	21
2.2	Espaços não-equiprováveis	23
2.2.1	<u>Jogada de Vênus</u>	27
3	COMBINATÓRIA	29
3.1	Executivos e Moedas	29
3.2	Resultados de um teste	30
3.3	Outra analogia a moedas	34
4	PROBABILIDADE CONJUNTA E CONDICIONAL	38
4.1	Policia! Divorciado e Morto	40
4.2	Caso dos Collins	41
4.3	Duas Meninas	42
4.4	Ler Ment	44
4.5	Menos é mais	44
5	DIANTE DO ACASO	48
5.1	Estratégias diante do acaso	50
5.2	Exemplo das lâmpadas	53
6	LEI DE BENFORD	56
6.1	Apresentação da Lei	56
6.2	Aplicações da Lei de Benford	58
7	O ANDAR DO BÊBADO	60
7.1	O que é um passeio aleatório?	61
7.2	Caminhos	62
7.2.1	<u>Pontos “alcançáveis”</u>	63
7.3	Princípio da Reflexão	67
7.4	Apuração parcial da votação	69
7.4.1	<u>Contando caminhos</u>	69
7.4.2	<u>Lema principal</u>	75
7.4.3	<u>Empates</u>	79
7.5	Na liderança	81
7.6	Troca de liderança	87
	CONCLUSÃO	95
	REFERÊNCIAS	98

INTRODUÇÃO

A ciência é uma grande montanha de açúcar; dessa montanha só conseguimos retirar insignificantes pedacinhos.

Malba Tahan

É indiscutível que a probabilidade é um tema que seduz e prega peças nas mentes mais brilhantes. Grandes matemáticos como D’Alembert já foram traídos por um pensamento equivocado. Um problema de simples enunciado pode pedir soluções elaboradas. Assim como, enunciados complicados podem ser associados a situações extremamente fáceis. Esta dissertação se baseia em problemas apresentados no livro “O Andar do Bêbado” de Leonard Mlodinow. A obra tenta deixar a Matemática, em especial, a probabilidade, mais acessível para o público em geral. Sua linguagem é de fácil entendimento além da narrativa tender à comicidade diversas vezes. Apresenta uma grande variedade de exemplos e curiosidades. Este trabalho visa preencher algumas lacunas deixadas pelo autor. A ideia é facilitar o entendimento dos exemplos, visto que o livro omite muitos cálculos. Apresenta soluções sem se aprofundar. Esta dissertação se propõe a ser uma ferramenta que dialogue com matemáticos. Não apresenta um método de ensino. O objetivo é aprofundar temas e suscitar ideias.

O capítulo “O que é a probabilidade?” apresenta o início do conhecimento matemático sobre esse tema. Aborda o mito de que os cálculos são necessários para uma abordagem inicial. Mostrará que a definição de probabilidade não é um consenso.

“Espaço Amostral” é um capítulo que menciona o porquê a genialidade dos gregos antigos não foi suficiente para estruturar a probabilidade. Além disso, define e exemplifica espaço amostral, equiprováveis e não equiprováveis.

“Combinatória” mostra como três exemplos tão distintos podem ser associados ao lançamento de moedas. Utiliza um princípio caro à Matemática. O fato de associar situações aparentemente complexas a exemplos simples. Não aplicável às Ciências ditas Humanas.

“Probabilidade Conjunta e Condicional” foi um capítulo que buscou a solução de agregar situações diferentes apresentadas por Mlodinow. Também se propõe a mostrar erros comuns das pessoas nesse tema. Como pode ser difícil definir qual de duas situações é a mais provável.

“Diante do Acaso” mostra a junção entre a Matemática e a Psicologia. O ser humano não tem domínio sobre o conceito de aleatório. Não o entende completamente. Há uma busca frequente por padrões mesmo em contexto onde não existem. Neste capítulo é possível conhecer o conceito de “matching” e como está enraizado no cotidiano.

Qual o primeiro passo para definir se uma empresa fraudou sua contabilidade? O

capítulo “Lei de Benford” mostra como esse teorema pode ser eficiente na análise inicial das auditorias fiscais.

Por fim, o capítulo “O Andar do Bêbado”. Ele recebe o mesmo nome do livro. A denominação se refere ao andar aleatório. Uma imagem mental muito clara do que é estudado pela probabilidade. Permite elucidar o que é randômico. A falta de controle da capacidade motora leva a pessoa embriagada a se mover de forma não previsível. A expressão também pode ser associada à falta de controle simplesmente. Assim como o bêbado se move sem garantia de seguir o caminho desejado, mentes brilhantes podem ser levadas a cometer enganos.

1 O QUE É PROBABILIDADE?

“No fundo, a teoria das probabilidades é apenas o senso comum expresso em números.”

Pierre Simon Laplace

Este capítulo visa buscar a definição de probabilidade. Em geral, cálculos são utilizados para a explicação deste ramo da Matemática. Para resolver uma situação-problema, o método utilizado é a comparação com outros exercícios semelhantes já resolvidos. Como se fosse possível criar um catálogo. No entanto, isso acarreta em erros, pois é comum haver falha na interpretação do que se quer calcular. A proposta é que se defina a probabilidade para evitar tais equívocos. Além disso, assim como a Geometria e qualquer conhecimento matemático, procura-se definir cada ente de forma criteriosa para que este conceito não possa ser refutado posteriormente. Apesar deste foco, deseja-se associar, sempre que possível, os exemplos a situações do cotidiano para garantir o entendimento. Essa prática é utilizada em toda a dissertação.

Segundo Carvalho et al. (2016), os experimentos podem ser de dois tipos: determinísticos ou aleatórios. “Um experimento é determinístico quando repetido em condições semelhantes conduz a resultados essencialmente idênticos.” O ato de soltar objetos a uma determinada altura do solo faz com que o mesmo siga em sentido descendente. A natureza do objeto pode retardar o momento de chegada, mas não há dúvida que encontrará o chão. A ação da gravidade no planeta Terra garante que esse experimento é determinístico. Já os aleatórios são aqueles que ao serem repetidos nas mesmas condições podem gerar resultados diferentes, como o lançamento de um dado. Não existe a garantia de que irá sair um determinado número em um dado honesto¹. O mesmo se pode dizer sobre os números na loteria e outros tantos experimentos sobre os quais não se tem certeza a respeito dos seus resultados. No entanto, em muitos casos, o conjunto dos possíveis resultados do experimento é conhecido previamente.

É de conhecimento geral que a origem da probabilidade foram os jogos de azar. Começou no século XVI com Gerolamo Cardano tentando maximizar seus ganhos em apostas. Segundo Boyer e Merzbach (2012), Pierre Simon Laplace foi o primeiro a estruturar a teoria da probabilidade. Isso ocorreu no século XVIII. Para muitos, até hoje, a probabilidade se restringe a cálculos numéricos, utilizados para realizar escolhas sobre o lançamento de um determinado produto ou definir previsões meteorológicas. Para além

¹ Um dado cujas faces têm a mesma chance de sair é chamado de dado honesto ou não-viciado. Se o dado apresentar uma face maior do que as demais, ou mais pesada, essa tenderá a cair para baixo mais vezes do que as outras. Assim, o número contido na face oposta irá aparecer mais vezes. Esse dado é chamado de viciado.

dessa visão pragmática, é preciso entender a probabilidade como uma área da matemática que não tem como objetivo apenas cálculos. A probabilidade não pode se resumir a contas. Não deve apenas buscar valores específicos para cada problema apresentado, mas sim modelos que podem ser aplicados em situações gerais.

Segundo Feller (1957), a probabilidade é semelhante à geometria. Deve ser regida por axiomas e partir de experiências mais simples, como o lançamento de dados para posteriormente ser generalizada em modelos abstratos. Os axiomas são verdades não questionáveis. A geometria é construída a partir de conceitos primitivos e axiomas. O ponto é um conceito primitivo para a geometria. Não se define ou demonstra o que é um ponto. Dizer que o ponto é feito quando se apoia a caneta no papel é apenas um recurso estilístico para passar a ideia do conceito. Não é uma definição. Em sua apresentação axiomática da probabilidade, Kolmogorov (1956) considera evento aleatório e sua probabilidade como conceitos primitivos.

Feller (1957) entende haver três aspectos que devem ser considerados no estudo da probabilidade: (a) o conteúdo lógico formal, (b) o conhecimento intuitivo (experiências anteriores) e (c) as aplicações.

O conteúdo lógico formal é a maneira como a Matemática se estrutura. A geometria se baseia em axiomas e a partir deles demonstra teoremas. Fossa (2009) diz que “axioma é uma proposição aceita sem demonstração.” São alicerces. A partir deles são construídas as teorias matemáticas.

Existem conceitos na Matemática que são muito abstratos. No entanto, a probabilidade e a geometria são temas que são mais intuitivos. Muitas vezes, é possível relacionar conceitos dessas áreas a situações cotidianas. Considere que um jogador vence se ao lançar um dado, sair o número 5. É comum ouvir que a chance dele ganhar é “uma em seis.” Expressões como esta são vagas, mas são importantes para o conhecimento inicial de probabilidade. As pessoas entendem que a oferta de faces que garantem vitória é de apenas uma em um total de seis. Por ser um objeto familiar, percebem também que isso não significa uma garantia. Pode-se jogar o dado dez vezes e sair sempre o número dois, mesmo para um dado honesto. O conhecimento intuitivo é importante para o estudo inicial da probabilidade, mas há outro fator central. Em suas aulas sobre o tema, o professor Augusto César Morgado dizia ser essencial que a pessoa que fosse resolver um problema de combinatória ou probabilidade se colocasse na posição de quem vai efetivamente lidar com aquela situação na prática. Exercícios sobre como acomodar pessoas em uma fila ou lançamento de moedas, por exemplo, devem ser tratados como se a pessoa fosse de fato organizar uma fila ou apostar em moedas.

Para o entendimento do que é probabilidade, inicialmente é preciso utilizar o conhecimento intuitivo de situações simples. Valorizar o conhecimento intuitivo garante um ganho posterior sobre as generalizações. Por exemplo, para entender o teorema do produto, pode-se utilizar uma situação hipotética. Considere uma adolescente que goste de

tirar fotos. O que seria mais provável para a vida adulta dela? Ser mãe ou ser fotógrafa e mãe? As pessoas têm uma tendência a acreditar que a segunda opção é a mais provável. No entanto, ela é mais específica que a primeira. Ser mãe é muito mais provável do que ser mãe e ainda ser fotógrafa. Outra vantagem desse exemplo é que não envolve números. Ele está associado a um importante teorema, mas não utiliza cálculos. Em geral, a aceitação para esse formato é melhor. É interessante observar que muitos concursos públicos no Brasil substituíram a Matemática por Raciocínio Lógico. O público que estuda para tais provas ficou aliviado com a troca, mas continuam sendo problemas matemáticos com outra roupagem. Muitas vezes sem números ou cálculos. O raciocínio intuitivo é importante, mas não pode ser um fim em si mesmo. É preciso haver um jeito formal de explicar estes exemplos, ou seja, suas generalizações. Somente assim, é possível aplicar essas estruturas em uma conjuntura mais complexa.

Segundo Singh (1946), não há uma definição formal universalmente aceita para a probabilidade. Há matemáticos que são mais categóricos. Azevedo (2004) afirma que não se define probabilidade. Ela é um axioma. Singh (1946) discute quatro tentativas de grandes matemáticos para formalizar a probabilidade: as de Bernoulli-Laplace, Von Mises, Kolmogorov e a de Jeffreys. Elas serão agora brevemente apresentadas.

- **Bernoulli-Laplace**

O tratamento do tema proposto por Bernoulli e Laplace é conhecido como a definição clássica de probabilidade. Na definição de Bernoulli e Laplace considera-se conhecido o conjunto de todos os n resultados possíveis e equiprováveis de um determinado experimento. Se m deles são favoráveis a um evento A , a probabilidade de A acontecer é $\frac{m}{n}$.

Uma das críticas feitas a esta definição é o uso da ideia de equiprovável antes mesmo de se definir probabilidade. Ao se falar em equiprovável, já se utiliza o conceito de probabilidade. Singh (1946) afirma que, para resolver isso, alguns autores retiram a palavra equiprovável. Porém isso gera outro problema. Causa confusão sobre os conjuntos de resultados possíveis nos quais os elementos aparecem mais de uma vez. No caso dos números sorteados na Mega Sena, cada um tem a mesma chance de ser sorteado. Diz-se que são resultados equiprováveis. Por outro lado, em uma caixa na qual há 4 bolas verdes e 1 branca, a probabilidade de sortear verde é maior do que sair branca. Ao se retirar a palavra equiprovável, não é possível diferenciar esses conjuntos de possíveis resultados.

A definição não contempla os conjuntos de resultados possíveis nos quais um caso tem probabilidade maior de ocorrer do que outro. Como exemplo, pode-se citar uma apólice de seguro de vida. Imagina-se que a seguradora liste n possíveis causas de morte. Um dos motivos de óbito poderia ser a pessoa estar viajando de avião e o

mesmo cair. Neste caso, a probabilidade não é de $\frac{1}{n}$. Mortes por causas naturais ou acidentes de carro são mais comuns. Cada item da lista apresenta diferentes chances de ocorrer.

Azevedo (2004) lembra outra crítica à definição clássica. Ela vale somente para conjuntos de resultados finitos. A teoria de Laplace apresenta ainda outro fator limitador. Para aplicá-la é preciso conhecer todos os possíveis resultados de um experimento. Ela foi criada em um contexto de jogos. Experiências sobre as quais se tem conhecimento sobre o que irá sair. No entanto, a teoria clássica não consegue determinar a probabilidade de um senhor de 50 anos viver até os 61.

- **Von Mises**

Para Von Mises, a probabilidade de um evento está associada à frequência relativa em que ele ocorre. Seja s a quantidade de vezes que o evento A ocorre e n o número de repetições do experimento. A probabilidade de A é o limite de $\frac{s}{n}$ quando n tende ao infinito.

Ele é o precursor da noção frequentista de probabilidade. Von Mises avança esta discussão ao exigir que o limite proposto seja invariante com relação a “permutações” no conjunto de repetições considerado. O limite citado não é o usual da Matemática. A ideia é repetir o experimento um número de vezes suficientemente grande e utilizar a proporção de um determinado resultado como sua probabilidade. Pode-se observar a quantidade de meninas que nasceram no Brasil durante o ano de 2017. Daí se define que a proporção de bebês desse sexo é a probabilidade de nascer uma menina no país. Segundo Azevedo (2004)

Dizer que $P(A) = \lim_{n \rightarrow \infty} f_n(A)$ equivale a dizer que para todo $\epsilon > 0$,

$$P(|P(A) - f_n(A)| > \epsilon) \rightarrow 0$$

A probabilidade do módulo acima ser diferente de zero é praticamente nula. Dessa forma, a probabilidade do evento A e a frequência relativa de A diante de um número suficientemente grande de repetições de um experimento são praticamente o mesmo valor.

- **Kolmogorov**

Segundo Corry (1997), no Congresso de Matemática de 1900, Hilbert propôs 23

problemas matemáticos. O sexto deles se referia à axiomatização da probabilidade²:

As investigações sobre os fundamentos da geometria sugerem o problema: tratar da mesma maneira, por meio de axiomas, aquelas ciências físicas nas quais a matemática desempenha um papel importante; no topo da lista estão as teorias das probabilidades e da mecânica (CORRY, 1997, p. 84)

Kolmogorov definiu probabilidade a partir da teoria da medida. Concordando com Hilbert, Kolmogorov (1956) defende que a probabilidade é uma disciplina da Matemática tal como a Geometria ou a Álgebra e precisa ser regida por axiomas. Eles têm que servir como alicerces para teoremas posteriores.

A teoria de Laplace é aplicada para conjuntos finitos. Kolmogorov vai além. Sua definição de probabilidade também se aplica a conjuntos infinitos enumeráveis e não enumeráveis. Além disso, a definição de Laplace está contemplada pela de Kolmogorov.

Kolmogorov utiliza em sua teoria álgebras e σ -álgebras de subconjuntos. Se X é um conjunto não vazio, uma álgebra de conjuntos em X é uma coleção não vazia \mathcal{A} de subconjuntos de X que é fechada sob uniões finitas e complementares. Quando uma álgebra é fechada também sob uniões enumeráveis, ela é chamada σ -álgebra. Os eventos serão elementos de uma σ -álgebra aos quais será atribuído um número que representa a probabilidade associada a eles.

Narens (2007) afirma que a teoria de Kolmogorov se baseia em seis axiomas, são eles:

1. A incerteza é expressa através de uma única função P ;
2. O domínio da função P é uma σ -álgebra de conjuntos;
3. O contradomínio de P é o intervalo $[0, 1]$;
4. $P(\emptyset) = 0$ (evento impossível);
5. $P(X) = 1$ (evento certo de ocorrer);
6. Se $A_i \in \mathcal{A}$ para cada $i \in \mathbb{N}$ é uma sequência de conjuntos dois a dois disjuntos, então:

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

O axioma 6 nos diz que a função de probabilidade é aditiva.

² The investigations on the foundations of geometry suggest the problem: To treat in the same manner, by means of axioms, those physical sciences in which mathematics plays an important part; in the first rank are the theory of probabilities and mechanics.

Para Kolmogorov, X é o espaço amostral e a classe de eventos para os quais definimos a probabilidade é uma σ -álgebra \mathcal{A} em X . A probabilidade é uma medida P que satisfaz aos seis axiomas citados. Neste caso, para cada evento A pertencente à σ -álgebra \mathcal{A} , sua medida $P(A)$ corresponde à probabilidade desse evento.

A teoria de Laplace satisfaz os axiomas de Kolmogorov. Seja X um conjunto finito e não vazio de todos os possíveis resultados de uma única repetição do experimento. O espaço amostral será a álgebra \mathcal{A} formada por todos os subconjuntos de X . Isto é, \mathcal{A} é o conjunto das partes da X . Para cada subconjunto A de X , Laplace definiu $P(A) = \frac{\#A}{\#X}$ ($\#A$ denota a quantidade de elementos do conjunto A). A função P se aplica a subconjuntos de X , aos eventos. Isto é, o domínio de P é \mathcal{A} . Resta verificar que a função P satisfaz os demais axiomas. Se o evento for \emptyset , aplicando a teoria clássica, tem-se $P(\emptyset) = \frac{\#\emptyset}{\#X} = 0$. Da mesma forma, $P(X) = \frac{\#X}{\#X} = 1$. O contradomínio de P é o intervalo $[0, 1]$. De fato, para um subconjunto A qualquer, tem-se $\#\emptyset \leq \#A \leq \#X$, logo $0 \leq P(A) \leq 1$. A aditividade de P é facilmente verificada.

- **Jeffreys**

Howie (2002) diz que para entender a probabilidade é preciso observar que há dois grupos principais de classificação. Para alguns matemáticos, a probabilidade é uma fração baseada em uma sequência de repetições de um experimento. Ou seja, uma vez que a probabilidade de lançar uma moeda e aparecer cara é de 50%, após uma sequência de jogadas tendendo ao infinito, a quantidade de caras será, aproximadamente, a metade delas.

No entanto, para outro grupo há uma perspectiva diferente. Para um meteorologista afirmar que há probabilidade de 50% de chover amanhã, ele não analisa o clima de uma sequência de dias. O profissional se baseia em uma série de dados climáticos que o fazem atribuir um número numa escala entre 0 e 1. Essa abordagem atribui um grau de conhecimento ou “crença” ou interpretação Bayesiana.

Para Howie (2002), o cálculo da probabilidade de jogos de azar é facilmente realizado devido ao conhecimento de todos os resultados possíveis. Essa abordagem não pode ser utilizada para o exemplo de chover em um determinado dia. Neste caso, há uma inversão do cálculo da probabilidade, devido à incerteza sobre os resultados possíveis. Chover deriva de uma conjuntura de fatores. Howie (2002) afirma que a primeira menção à “probabilidade inversa” foi feita em um artigo de 1764 de Bayes. Seu método poderia ser aplicado ao sorteio de bolas de uma urna. Após se retirar cada bola, a mesma é repostada na urna. Há bolas pretas e brancas e não se conhece a proporção de cada cor. A partir da observação dos primeiros sorteios, verifica-se

que saíram p bolas pretas e q brancas. A probabilidade da próxima ser preta³ é de $p+1/p+q+2$.

Neste contexto é possível entender a perspectiva de Jeffreys. Para Jeffreys (1950 apud HOWIE, 2002), o teorema de probabilidade inversa de Bayes “está para a probabilidade assim como o de Pitágoras está para a geometria”⁴.

Howie (2002) conta que no início da década de 1930, a teoria de Bayes estava associada à inferência científica e não era considerada útil para a análise de dados. Jeffreys mudou isso. Seu livro *Theory of Probability* foi o primeiro a aplicar a teoria de Bayes. Somente por volta de 1950 que as teorias de Bayes voltaram a ter seu valor e o status de estatístico foi atribuído novamente à Jeffreys. No entanto, Singh (1946) afirma que a maioria dos matemáticos não aceita a abordagem do autor.

A hipótese principal da teoria de Jeffreys é que a toda proposição p feita sobre dados q podemos atribuir uma única probabilidade e que probabilidades de diferentes proposições sobre diferentes dados são comparáveis. Mas ele não deu qualquer indicação de como esse número é calculado. (SINGH, 1946, p. 258)⁵

Dentre essas abordagens sobre probabilidade, a mais utilizada é a de Kolmogorov, sendo a de Laplace considerada um caso particular para conjuntos finitos de resultados possíveis. A noção de probabilidade baseada na teoria de conjuntos está presente em textos didáticos. Para ilustrar este fato, vamos considerar a abordagem de Carvalho et al. (2016) para probabilidade. Inicialmente, eles citam Cardano em linguagem moderna: o Espaço Amostral é o conjunto de todos os resultados possíveis de um determinado experimento. Já a probabilidade é o quociente da quantidade de casos favoráveis pelo total de resultados possíveis de um experimento aleatório. Para entender o que significam casos favoráveis, pode-se pensar em uma aposta. São os casos nos quais o jogador vence. A situação procurada, é chamada de “evento”. Neste exemplo, é a vitória do participante. O evento é um subconjunto do espaço amostral. É importante lembrar que essa definição de probabilidade se refere aos espaços amostrais finitos e equiprováveis⁶.

Mais formalmente, Carvalho et al. (2016) consideram A um subconjunto de um espaço amostral S e definem:

A probabilidade é uma função que associa a cada evento A um número $P(A)$ (a probabilidade do evento A ocorrer) de forma que:

³ Este resultado também foi obtido independentemente por Laplace em 1774 e é conhecido como Regra de Sucessão de Laplace. Para mais detalhes consulte Gorroochurn (2016).

⁴ o Theorem of Inverse Probability de Bayes is “to the theory of probability what Pythagoras’s theorem is to geometry

⁵ Jeffreys’ fundamental hypothesis is that every proposition p on data q has a unique numerical probability and that probabilities of different propositions on different data are comparable. But he gives no indication how this number is to be assessed in any given case.

⁶ Espaços Equiprováveis são conjuntos nos quais cada elemento tem a mesma probabilidade de ocorrer.

1. Para todo evento A , $0 \leq P(A) \leq 1$
2. $P(S) = 1$
3. Se A e B são eventos mutuamente excludentes, isto é, eventos que não podem ocorrer simultaneamente (isto é, $A \cap B = \emptyset$) então: $P(A \cup B) = P(A) + P(B)$.

Admitem que S é finito e equiprovável, e definem a probabilidade $P(A)$ de um dos resultados contidos no subconjunto A ocorrer por

$$P(A) = \frac{\text{número de casos favoráveis}}{\text{total de resultados possíveis}}$$

Essencialmente, eles apresentam a definição clássica revestida do formalismo de Kolmogorov. Vamos tratar um exemplo concreto sob esta perspectiva. Considere o lançamento de dois dados de seis faces. Neste caso, o espaço amostral possui 36 resultados distintos. Considerando-se o par ordenado (resultado do “dado I”, resultado do “dado II”), tem-se $S = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), \dots, (6, 6)\}$. O evento considerado será saírem números iguais para ambos os dados. Este evento corresponde ao subconjunto A do espaço amostral S que possui seis elementos, a saber

$$A = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}.$$

Sua probabilidade é dada por $P(A) = \frac{6}{36} = \frac{1}{6}$. Esta probabilidade é um número racional contido no intervalo $[0, 1]$. O limite inferior da probabilidade é zero. Para entender isso, pode-se pensar no mesmo experimento, o lançamento dos dados, mas agora o evento A é sair um número maior do que 7 para o primeiro dado. Não há nenhuma face com esse número, logo, o conjunto que representa este evento é vazio. Assim, $P(A) = \frac{0}{36} = 0$.

O limite superior da probabilidade é 1. Neste caso, o evento A é igual a S . $P(S)$ é a probabilidade de sair algum dos resultados do espaço amostral. Uma vez que S é o conjunto composto por todos os resultados possíveis do experimento, então, todos são casos favoráveis. Daí, tem-se que $P(S) = \frac{36}{36} = 1$. Neste caso, a probabilidade é equivalente a 100 % de chance de acerto.

Sejam A e B dois eventos mutuamente excludentes. Isso significa que os conjuntos A e B não possuem elementos em comum. Pode-se pensar no evento A ainda como os números dos dois dados serem iguais e o B como números diferentes. Ao se lançarem os dados, não é possível ocorrer o evento A e B ao mesmo tempo em uma tentativa. Todos os resultados estarão no conjunto A ou no B , logo $A \cup B = S$, assim:

$$1 = P(S) = P(A \cup B) = \frac{36}{36} = \frac{6}{36} + \frac{30}{36} = P(A) + P(B).$$

O mesmo argumento pode ser utilizado para o caso de eventos dois a dois disjuntos.

2 ESPAÇO AMOSTRAL

Na vida, não existe nada a temer, mas a entender.

Marie Curie

Assim como a probabilidade foi definida no capítulo anterior, o conceito de espaço amostral o é neste trecho da dissertação. Muitas vezes, ele é tratado como um conjunto finito. Há a preocupação de se buscar definições e não focar apenas no caso particular de conjuntos finitos. Decidiu-se abordar espaços amostrais não-equiprováveis por serem pouco explorados, além de formalizar um método para lidar com eles.

2.1 Tudo a seu tempo

Os gregos contribuíram brilhantemente para o desenvolvimento da Matemática, mas não inventaram a probabilidade. Segundo Mlodinow (2009), um dos motivos é que eles ainda não conheciam a aritmética necessária. Na Grécia Antiga, nem o zero era representado. Os conceitos de espaço amostral e probabilidade foram criados por Gerolamo Cardano no século XVI. Mlodinow (2009) afirma que a Matemática daquele século apresentava recursos adequados para a descoberta de Cardano. Embora ainda fossem rudimentares, na época do matemático já era possível contar com a ideia de frações, essencial para a probabilidade. No entanto, o sinal de igual só foi inventado depois.

Além disso, Cardano não era diferente das demais pessoas do mesmo período histórico. Era supersticioso. Para ele, o uivo de um cão significava que uma pessoa amada iria morrer. Aparentemente, a racionalidade necessária para a criação de um conceito científico inovador associada à fé em algum tipo de crença pode parecer um paradoxo. Pode soar contraditório, mas todos sabem que os seres humanos são complexos. Muitos tentam classificar fé e razão como conceitos antagônicos. No entanto, irracional é não admitir que a maioria das pessoas carregam ambos. Estudiosos renomados que muito contribuíram para as ciências foram adeptos a credos. Cardano é só um exemplo desse grupo. Apesar das adversidades, ele foi o inventor da probabilidade.

2.1.1 Espaços Amostrais

A obra de Cardano se referia a espaços amostrais finitos. São conjuntos nos quais conseguimos contar todos os seus elementos. Mais precisamente, existem um natural $n \in$

\mathbb{N} e uma função bijetora f definida em $[n] = \{1, 2, \dots, n\}$ com valores no espaço amostral. Há outros tipos de espaços amostrais. Meyer (2006) classifica os espaços amostrais em finitos, infinitos enumeráveis e infinitos não-enumeráveis. Nos conjuntos infinitos não-enumeráveis não é possível “contar” seus elementos, mas nos infinitos enumeráveis, sim. Associa-se cada elemento a um único número inteiro positivo.

Para ilustrar espaços amostrais infinitos enumeráveis, Meyer (2006) cita um experimento de fabricação contínua de peças até que sejam produzidas 10 peças perfeitas. Neste caso, não se sabe quantas unidades serão feitas até se obter a meta. Em teoria, poderia ser uma quantidade infinita. Como conjunto não-enumerável, Meyer (2006) cita o exemplo do lançamento de um míssil. Os registros da altura em relação ao solo a cada instante é um conjunto não-enumerável. Como outro exemplo, pode-se recorrer ao ramo da Análise Matemática. Considere um conjunto A . Denotamos por $\mathcal{P}(A)$ o conjunto de todos os subconjuntos formados pelos elementos de A . Lima (2016) afirma que o conjunto $\mathcal{P}(\mathbb{N})$, ou seja, o conjunto de todos os subconjuntos formados com os números naturais é não-enumerável. Para provar isso, é preciso mostrar que não existe uma correspondência biunívoca entre os elementos de $\mathcal{P}(\mathbb{N})$ com os naturais, ou seja, deseja-se mostrar que não existe uma função bijetiva $f : \mathbb{N} \rightarrow \mathcal{P}(\mathbb{N})$. Para a função ser bijetiva, é preciso ser sobrejetiva e injetiva. Será provado que ela não é sobrejetiva, mas antes é preciso considerar uma proposição.

Proposição 1 (Lima (2016, p. 37)). *Dado um conjunto não-vazio A , não existe uma função sobrejetiva $f : A \rightarrow \mathcal{P}(A)$.*

Prova. Seja $f : A \rightarrow \mathcal{P}(A)$ uma função qualquer, mostraremos que f não é sobrejetiva, ou seja, mostraremos que existe algum subconjunto de A , denotado por Z_f , tal que $Z_f \neq f(x)$, para todo $x \in A$. Afirmamos que:

$$Z_f = \{x \in A : x \notin f(x)\}$$

tem as propriedades desejadas. De fato, dado $x_0 \in A$, temos uma das seguintes possibilidades:

- (i) $x_0 \in f(x_0) \Rightarrow x_0 \notin Z_f$, logo, $f(x_0) \neq Z_f$
- (ii) $x_0 \notin f(x_0) \Rightarrow x_0 \in Z_f$, logo, $f(x_0) \neq Z_f$.

□

A demonstração anterior diz que para qualquer função f definida em A e com valores em $\mathcal{P}(A)$, sempre existe um conjunto Z_f , subconjunto de A que não é imagem, através de f , de nenhum elemento de A . O conjunto Z_f é o subconjunto formados por elementos de A que não pertencem à sua própria imagem. Convém notar que se a função f for tal $x \in f(x)$, para todo $x \in A$, Z_f seria o conjunto vazio.

Só há duas possibilidades para um elemento x_0 qualquer, ou pertence a $f(x_0)$ ou não pertence. No item (i), como $x_0 \in f(x_0)$, por definição, não pertence a Z_f . No item (ii), $x_0 \notin f(x_0)$, por definição, está em Z_f .

Deste modo, Lima (2016) demonstra que o conjunto $\mathcal{P}(\mathbb{N})$ é não-enumerável, “Se $\mathcal{P}(\mathbb{N})$ fosse enumerável, então existiria uma função bijetiva $f : \mathbb{N} \rightarrow \mathcal{P}(\mathbb{N})$, sendo bijetiva, ela é sobrejetiva, o que contraria a proposição.”

2.2 Espaços não-equiprováveis

Para se aplicar a definição Laplaciana de probabilidade é preciso que o espaço amostral utilizado seja equiprovável. No entanto, mesmo matemáticos renomados se enganam. Mlodinow (2009) diz que D’Alembert definiu o espaço amostral S do lançamento de duas moedas pela quantidade de caras, $S = \{0, 1, 2\}$. Afirmou que a probabilidade de sair uma cara nesse experimento era de $\frac{1}{3}$. De fato é um possível espaço amostral, o erro foi aplicar a definição clássica em um conjunto não-equiprovável.

Neste caso, é preciso utilizar um outro espaço S_1 que seja equiprovável. Uma abordagem inicial a problemas desse tipo, seria listar os casos possíveis. Assim, admitindo que os elementos de S_1 sejam os pares ordenados (resultado da primeira moeda, resultado da segunda moeda), tem-se:

$$S_1 = \{(cara, cara), (cara, coroa), (coroa, cara), (coroa, coroa)\}$$

Em S_1 , há dois casos com uma cara, aplicando a definição Laplaciana, a probabilidade do evento uma cara é $\frac{2}{4} = \frac{1}{2}$.

Segundo Meyer (2006), é muito comum a referência ao espaço amostral como o único. Diz-se que é “o” espaço amostral. O correto é definir como “um”. No exemplo acima tem-se dois espaços amostrais distintos para o mesmo experimento.

Mesmo que a pessoa não seja supersticiosa como Cardano, pode errar. Ao longo da história, muitos renomados estudiosos cometeram equívocos. D’Alembert claramente dominava os conceitos matemáticos, mas cometeu enganos como qualquer um.

Voltemos ao lançamento de dois dados. Dessa vez será considerada a soma dos resultados. Logo, o espaço amostral é

$$S_1 = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.$$

Se o evento A é a soma ser igual a 6, qual a probabilidade de ele ocorrer? É importante observar que as somas apresentam probabilidades distintas. Para sair 2 é preciso sair 1 em cada dado, mas a soma 6 pode ser obtida de cinco formas diferentes a saber $(1, 5), (5, 1), (2, 4), (4, 2)$ ou $(3, 3)$, onde as coordenadas de cada par ordenado representam os resultados obtidos em cada um dos dados. Se consideramos S_1 como espaço amostral, o evento $E_1 = \{6\}$ é cinco vezes mais provável do que o evento $E_2 = \{2\}$. Como já foi

visto, há 36 resultados possíveis para o experimento “lançamento de dois dados”. Cada um deles gera uma soma. Daí, $P(E_1) = \frac{5}{36}$. Já a probabilidade de sair a soma 2 é $\frac{1}{36}$.

Uma estratégia para lidar com espaços não-equiprováveis é recorrer a um que seja equiprovável. O espaço amostral S_1 está bem definido, mas não é possível aplicar a teoria Laplaciana diretamente. Se fosse feito assim, $P(E_1)$ seria $\frac{1}{11}$. Para não cometer este erro, é preciso associar o evento a soma 6 aos resultados que a geram. Na prática, foi utilizado o espaço S no qual todos os resultados têm a mesma probabilidade de ocorrer. Dessa forma, foi possível utilizar a definição clássica de probabilidade usando em S o evento $A = \{(1, 5), (5, 1), (2, 4), (4, 2), (3, 3)\}$ para calcular em S_1 a probabilidade de E_1 .

Outro exemplo de espaço amostral não equiprovável é uma experiência na qual há três bolas azuis e uma branca em uma caixa. Neste caso, só é possível distinguir as bolas por sua cor após o sorteio. Além disso, as bolas serão devolvidas à caixa após cada retirada. Qual a probabilidade de uma pessoa sortear duas bolas brancas em duas tentativas?

A fim de facilitar a escrita, a cor azul será representada por A e a branca por B . Será utilizada a notação (cor da primeira bola sorteada, cor da segunda bola sorteada) para definir os possíveis resultados. Neste exemplo, um espaço amostral é

$$S = \{AA, AB, BA, BB\}$$

Ao visualizar este conjunto, pode-se ter a ideia de que a probabilidade de saírem duas bolas brancas é de $\frac{1}{4}$. No entanto, a probabilidade de cada par é diferente. S é não equiprovável, pois o caso BB tem menos chances de sair do que o AA . Para mostrar essa diferença, as bolas azuis serão diferenciadas por índices, a saber, A_1, A_2 e A_3 .

Para calcular a probabilidade do evento duas bolas brancas, será utilizado o espaço amostral equiprovável S_1 . Assim, tem-se:

$$S_1 = \{(B, B), (A_1, B), (A_2, B), (A_3, B), (B, A_1), (B, A_2), (B, A_3), (A_1, A_1), (A_2, A_2), \\ (A_3, A_3), (A_1, A_2), (A_1, A_3), (A_2, A_1), (A_3, A_1), (A_2, A_3), (A_3, A_2)\}$$

Obviamente que o caso (A_2, B) e (A_3, B) serão visualmente iguais. A ideia de listar os elementos de S_1 é mostrar mais claramente as probabilidade de cada caso. A abordagem de listar é interessante no primeiro contanto com a combinatória e a probabilidade. Assim, utilizando o espaço amostral S_1 equiprovável e a definição laplaciana de probabilidade, é possível verificar que a probabilidade de saírem duas bolas brancas em duas retiradas é de $\frac{1}{16}$.

No entanto, em vários problemas, é difícil enumerar os elementos do espaço amostral. Pretende-se utilizar posteriormente, técnicas mais eficazes. Construir modelos matemáticos que possam ser usados em situações semelhantes. O problema acima pode ser resolvido pelo teorema do produto. A probabilidade de sair a cor branca em cada tentativa é de $\frac{1}{4}$, então a probabilidade do evento mencionado é de $\frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$.

O teorema do produto assim como outros são apresentados de forma prematura aos alunos. Não é explorado o conceito de equiprovável antes. Assim, a combinatória e a probabilidade se tornam um conjunto de fórmulas sem sentido. Os estudantes tentam aplicá-las de qualquer maneira sem analisar cada experimento.

Ao recorrer a um espaço equiprovável subjacente, temos um entendimento concreto dos espaços não-equiprováveis. A partir do momento que essa ideia for familiar, pode-se recorrer à conjugação da regra do produto com as permutações simples. Para entender essa prática, considere uma caixa com 6 bolas verdes, 4 amarelas e 2 azuis. Todas elas apresentam forma e massa iguais, ou seja, não é possível distingui-las pelo tato, apenas pela cor. Só é possível saber a cor após a bola ser sorteada. Será retirada uma de cada vez. Admitindo que há reposição da bola a cada sorteio, qual a probabilidade de se retirar três cores distintas? O espaço amostral composto por todos os resultados é não-equiprovável. Sortear três bolas azuis é menos provável do que saírem somente bolas verdes.

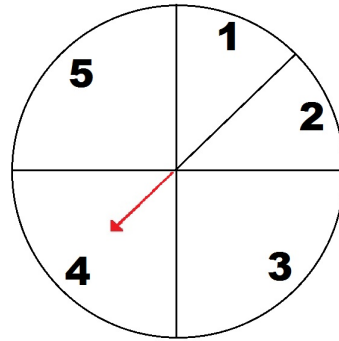
Neste caso, é muito trabalhoso listar os resultados de um espaço equiprovável associado. Além disso, listar as triplas ordenadas (a, b, c) com $a, b, c \in \{\text{verde, amarela, azul}\}$ não captura o fato de as bolas verdes terem mais chance de serem sorteadas. Ainda assim podemos calcular probabilidade de sair uma tripla (verde, amarela, azul) da seguinte maneira:

$$\frac{6}{12} \cdot \frac{4}{12} \cdot \frac{2}{12} = \frac{1}{36}$$

Mas a ordem das cores não importa neste caso. Aplicando a permutação das três cores, há 6 formas distintas desse resultado ocorrer. Daí, $\frac{1}{36} \cdot 6 = \frac{1}{6}$ é a probabilidade deste evento.

Há um ganho no uso dessa estratégia, mas não é recomendado que seja a primeira opção de abordagem. Entende-se que é preciso um avanço gradativo para de fato se conhecer o espaço amostral trabalhado. É preciso aprofundar os conhecimentos sobre espaços amostrais equiprováveis. Depois compará-los aos que não o são. Por fim, para resolver diretamente um problema referente a um espaço não-equiprovável, é interessante associar a ele um conjunto auxiliar equiprovável para aplicar a definição clássica de probabilidade e, posteriormente, usar diretamente resultados de combinatória.

Figura 1 - Roleta



Fonte: Captura de tela de vídeo do youtube⁷.

Um exemplo de espaço amostral não equiprovável foi apresentado pelo professor Conrad Pinheiro. Ele utiliza uma roleta na qual cada região tem probabilidades diferentes de sair. Utilizando a roleta da Figura 1, qual a probabilidade de sair o número 2?

Um espaço amostral para esse experimento é $S = \{1, 2, 3, 4, 5\}$. As probabilidades de cada região são diferentes. Para resolver esse problema, deve-se recorrer ao conceito de probabilidade geométrica. Tunala (1992) a define para um segmento de reta, área ou volume. Considere um segmento L e outro contido nele l . Escolhendo-se um ponto aleatório em L , a probabilidade de ser um ponto também de l é:

$$P(l) = \frac{\text{comprimento de } l}{\text{comprimento de } L}$$

Ao definir a probabilidade geométrica, Tunala (1992) considera um espaço amostral equiprovável. Isso se deve ao fato de ter padronizado as unidades de medida dos comprimentos. Ao usar a mesma medida, o autor utiliza o conjunto de pontos de cada segmento. Cada ponto tem a mesma probabilidade de sair.

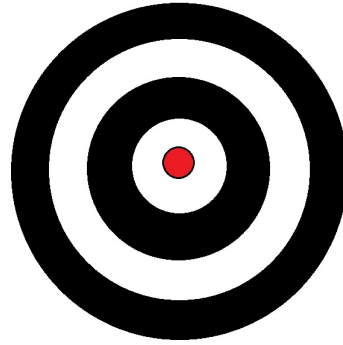
De forma análoga, Tunala (1992) definiu a probabilidade geométrica para regiões. Considere as figuras planas A e a . Sendo que a está contida em A . Escolhendo-se um ponto aleatório de A , a probabilidade de pertencer a a é

$$P(a) = \frac{\text{área de } a}{\text{área de } A}$$

No espaço amostral S , não se pode aplicar diretamente a definição clássica. A proba-

⁷ Disponível em: <<https://www.youtube.com/watch?v=UtYqfbck0TQ>>. Acesso em: 30 out. 2017.

Figura 2 - Alvo



Fonte: Captura de tela de vídeo do youtube⁸

bilidade de sair o número 2 não é $\frac{1}{5}$. Considere a circunferência pela qual a ponta da seta gira. Ela pode ser dividida proporcionalmente para cada região do círculo. O círculo foi repartido em quatro partes, sendo que as regiões 1 e 2 ocupam uma delas e são congruentes. Utiliza-se a estratégia de repartir a figura em oito regiões cujas áreas são congruentes a 2. Proporcionalmente, a circunferência pode ser dividida em oito pedaços. Assim, $P(2) = \frac{1}{8}$.

Ainda utilizando o conceito de probabilidade geométrica. Considere uma pessoa que nunca jogou dardos. Admitindo que ela acertou o alvo em uma tentativa, qual a probabilidade de ter sido no centro? Neste caso, não se tem informações sobre o desempenho do jogador em questão. Não é possível definir a precisão de sua tentativa. Se fosse um profissional desse esporte, a probabilidade de acertar o centro seria alta. Assim, será utilizada a probabilidade geométrica. Para isso, seria preciso calcular a área do centro O e de todo o alvo A . Dessa forma, a probabilidade do evento $O = \{\text{acertar o dardo no centro do alvo}\}$ é $P(O) = \frac{\text{área de } O}{\text{área de } A}$.

2.2.1 Jogada de Vênus

Nosso último exemplo de espaço amostral não equiprovável está associado a dados usados na Grécia Antiga. Os gregos não possuíam dados. Utilizavam ossos chamados astrágalos. Consiste em um objeto tridimensional com seis faces, porém duas delas não apresentam estabilidade para que o osso se apoie sobre elas. As demais faces têm probabilidades não equiprováveis, a saber duas têm 10% e as outras duas, 40%. A jogada de

⁸ Disponível em: <https://commons.wikimedia.org/wiki/File:Kempton_Astragaloi.jpg>. Acesso em: 30 out. 2017

Figura 3 - Astrágalos



Fonte: Imagem do Wikimidia Commons⁸.

Vênus correspondia a lançar quatro desses ossos e encontrar faces diferentes pra cada um. A probabilidade desse evento é $\frac{384}{10.000}$ (MLODINOW, 2009, p. 35).

Esse exemplo é diferente da quantidade de caras no lançamento de duas moedas. Aqui não é conveniente recorrer ao espaço amostral equiprovável, ou seja, listar os casos possíveis. A estratégia que será utilizada é a regra do produto e a permutação simples das quatro faces.

Pode-se calcular a probabilidade de se ter 4 resultados distintos:

$$\frac{10}{100} \cdot \frac{10}{100} \cdot \frac{40}{100} \cdot \frac{40}{100} = \frac{16}{10.000}$$

É preciso lembrar que esses resultados podem aparecer em ordens diferentes nos 4 ossos, usando a permutação simples, tem-se: $4 \cdot 3 \cdot 2 \cdot 1 = 24$. Daí, $\frac{16}{10.000} \cdot 24 = \frac{384}{10.000}$.

3 COMBINATÓRIA

Busque a simplicidade, e desconfie dela.
Alfred North Whitehead

Quando lidamos com espaços amostrais finitos, o cálculo de probabilidade depende essencialmente da caracterização adequada do espaço amostral e da determinação da cardinalidade do espaço amostral e do subconjunto associado ao evento de interesse. Neste capítulo, vamos analisar alguns dos problemas apresentados por Mlodinow (2009) dando especial atenção a estes dois aspectos.

3.1 Executivos e Moedas

É de conhecimento geral que é muito difícil prever qual filme será um campeão de bilheteria, porém a indústria de cinema tenta fazer isso frequentemente. Para Mlodinow (2009), é difícil definir se o sucesso de um produto se deve à sorte ou à habilidade de alguém. No entanto, os estúdios medem a competência de um executivo pelo número de sucessos de bilheteria atribuídos a ele. Mlodinow (2009) entende que é um experimento aleatório um filme ser um sucesso ou não. Isso não determina a capacidade profissional de um executivo. Uma vez que ele entende ser um processo randômico, faz uma analogia ao lançamento de moedas.

Mlodinow (2009) apresenta o exemplo de 10 executivos que jogaram 10 moedas. Considere o evento de, pelo menos um, conseguir, no mínimo, 8 caras ou, no mínimo, 8 coroas. Ele afirma que a probabilidade desse evento é $\frac{2}{3}$. Ao resolver esse problema, é preciso entender que o executivo ganharia se obtivesse 8 ou mais resultados iguais. Assim, 8 a 10 caras garantiriam sua vitória. Da mesma forma se fossem coroas.

Para calcular essas sequências de lançamentos que garantem vitórias, será utilizada a permutação com repetição. As sequências serão associadas a palavras. Quando o resultado for cara, será utilizada a letra K . Se for coroa, C . Assim, um resultado que atende ao evento é $KKKKKKKKCK$, no qual ocorreu cara para quase todos os lançamentos. Apenas a nona moeda teve coroa como resultado. A permutação irá possibilitar saber o número de resultados possíveis para cada caso.

Para calcular a quantidade de sequências de lançamentos de 10 moedas em que ocorrem 8 ou mais caras, consideramos as seguintes situações:

- **1º caso: exatamente 8 caras**

Nesta sequência tem-se o K aparecendo exatamente 8 vezes e o C , 2. Um total de 10 lançamentos.

$$P_{10}^{8,2} = \frac{10!}{8! \cdot 2!} = 45 \text{ resultados possíveis de sucesso.}$$

- **2º caso: exatamente 9 caras**

Aqui são 9 *Ks* e 1 *C*.

$$P_{10}^{9,1} = \frac{10!}{9! \cdot 1!} = 10 \text{ sequências que atendem ao evento.}$$

- **3º caso: 10 caras**

Apenas 1 caso.

Sendo assim, temos 56 casos para 8 caras ou mais e outros 56 para 8 coroas ou mais, totalizando 112 casos.

Ao invés de calcular a probabilidade do evento ocorrer, será utilizada a probabilidade do evento complementar. Será calculado o evento de nenhum deles conseguir esse resultado. Ao se jogar uma moeda, há dois resultados possíveis. Considerando 10 moedas, tem-se $2^{10} = 1024$ resultados distintos. Destes, 112 atendem ao evento, logo, 912 não atendem.

Cada um dos executivos pode ter um dos 1.024 resultados. Ainda que a ordem na qual essas sequências saem para cada executivo não importe no “mundo real”, é preciso levá-la em conta na composição do espaço amostral para capturar a real proporção de ocorrência entre os possíveis resultados. O espaço amostral Ω de interesse será formado por 10-uplas cujas componentes são possíveis sequências obtidas por cada um dos 10 executivos. A cardinalidade de Ω é igual a 1024^{10} . A probabilidade de nenhum deles obter 8 ou mais caras ou coroas é

$$\frac{912^{10}}{1024^{10}} \cong 0,31401$$

Logo, $1 - 0,31401 = 0,68599$ que resulta em⁹, aproximadamente, uma probabilidade de $\frac{2}{3}$. À primeira vista, emplacar 8 ou mais sucessos de bilheteria nos pareceria resultado e sinal da efetiva competência de um executivo. Com esse exemplo, Mlodinow (2009) nos mostra que em um equipe de 10 executivos, pelo menos um deles apresentar esse resultado em um conjunto de 10 filmes lançados não é tão improvável como podemos imaginar.

3.2 Resultados de um teste

É de conhecimento geral que muitas vezes as pessoas não estudam na frequência adequada. É muito comum turmas inteiras participarem de avaliações para as quais não

⁹ Os cálculos foram realizados no *octave*.

se prepararam adequadamente. Eventualmente estas provas são objetivas. Dessa forma, não é necessário contar apenas com o preparo. A sorte pode ajudar neste tipo de situação.

Suponha que apliquemos uma prova com 10 questões de verdadeiro ou falso a uma classe de 25 alunos do sexto ano. Façamos um registro dos resultados que uma aluna em particular poderia obter: ela poderia responder corretamente a todas as perguntas; poderia errar só uma – o que pode acontecer de 10 maneiras, pois ela poderia errar qualquer uma das 10 perguntas –, poderia errar duas perguntas – o que pode acontecer de 45 maneiras, pois existem 45 pares distintos de perguntas – e assim por diante. O resultado é que, em média, em um grupo de estudantes que estão dando chutes ao acaso, para cada aluno que acertar 100% das perguntas encontraremos cerca de 10 acertando 90% e 45 acertando 80%. Naturalmente, a chance de que um aluno obtenha uma nota próxima a 50% é ainda mais alta, mas numa classe de 25 alunos, a probabilidade de que ao menos um deles acerte 80% ou mais, se todos os alunos estiverem apenas chutando as respostas, encontra-se ao redor de 75%. (MLODINOW, 2009, p. 34)

Inicialmente, vamos quantificar todos os possíveis desempenhos de uma aluna específica, a saber:

- **Não errar nenhuma questão**

Apenas um caso. A aluna acertou exatamente o gabarito.

- **Errar apenas uma questão**

A forma mais simples de analisar esse problema é que ela pode errar qualquer uma das 10 questões, então são 10 casos para este item. Outra forma de abordar o problema é associar a uma “palavra” na qual cada acerto será representado por A e cada erro por E . Para o caso no qual ela erra a primeira questão, essa estratégia gera a palavra $EAAAAAAAAA$. Pode-se calcular todos os anagramas dessa palavra utilizando a permutação com repetição:

$$P_{10}^{9,1} = \frac{10!}{9! \cdot 1!} = 10$$

- **Errar exatamente duas questões**

Nesse exemplo, fica um pouco mais difícil calcular os casos possíveis. Pode-se recorrer a ideia de que tem-se 2 possibilidades de erro. Pode-se analisar como se ela escolhesse qual questão errar. Para o primeiro engano, a aluna pode errar qualquer uma das 10 questões. Para o segundo, 9. Teria-se 90 resultados. Nesse raciocínio, há uma relação de ordem. Errar as questões 1 e 2 é diferente de errar as perguntas 2 e 1. No entanto, numa prova não faz diferença a ordem na qual os erros se sucederam. Assim, cada um dos pares de questões foram contados duas vezes. Para corrigir este método, dividi-se os 90 casos por 2. O total de vezes que a aluna pode errar duas perguntas é 45.

Outra forma de pensar é usar anagrama. Será usada a permutação com repetição novamente. Nesta situação, temos como exemplo, a palavra *AEEAAAAAAA*. O número de organizações possíveis são:

$$P_{10}^{8,2} = \frac{10!}{8! \cdot 2!} = 45$$

• **Errar exatamente três questões**

Neste exemplo, pode-se, novamente, utilizar as duas abordagens mencionadas. Para o primeiro erro, tem-se 10 possibilidades. Sobram 9 questões para ocorrer o segundo erro. Já para o terceiro, 8. Assim, são $10 \cdot 9 \cdot 8 = 720$ resultados. Neste caso, pode-se considerar um único trio de erros. Imagine que a estudante errou as questões 1 – 3 – 4. Há 6 formas diferentes de organizar essas questões. Seria como se a aluna estivesse fazendo a prova em outra ordem. A sequência 4 – 3 – 1 é uma das 6. Considerando cada trio, tem-se 6 organizações distintas $3 \cdot 2 \cdot 1 = 6$. A ordem na qual as questões foram erradas não importa. Todos esses grupos correspondem a apenas um grupo de erros. Ao dividir 720 por 6, tem-se 120 casos possíveis.

A outra maneira teria como exemplo a palavra *EEEAAAAAAA*. A quantidade de anagramas é:

$$P_{10}^{7,3} = \frac{10!}{7! \cdot 3!} = 120$$

Para os demais casos, se forem exatamente, respectivamente, 4, 5, 6, 7, 8, 9 e 10 questões erradas serão 210 casos, 252, 210, 120, 45, 10 e 1.

Pode-se fazer uma comparação com o Triângulo de Pascal¹⁰. Consiste em um triângulo formado pelos valores da combinação de n e p , C_n^p , sendo $n, p \geq 0$.

$$\begin{array}{ccc} C_0^0 & & 1 \\ C_1^0 & C_1^1 & 1 \quad 1 \end{array}$$

¹⁰ O triângulo de Pascal era chamado de aritmético. Era um problema já conhecido há 600 anos antes de Pascal. Foi estudado na China por Zhu Shijie (1280 – 1303) e Yang Hui (viveu por volta de 1261 – 1275) (BOYER; MERZBACH, 2012, p. 148). O estudo sobre esse assunto também aparece em uma obra alemã *Rechnung* (1527), quase um século antes do nascimento de Blaise Pascal (BOYER; MERZBACH, 2012, p. 199). A justificativa para o triângulo ter sido renomeado para Triângulo de Pascal é que o matemático descobriu propriedades do mesmo (BOYER; MERZBACH, 2012, p. 254).

$$\begin{array}{cccccc}
 C_2^0 & C_2^1 & C_2^2 & & & & 1 & 2 & 1 \\
 C_3^0 & C_3^1 & C_3^2 & C_3^3 & & & 1 & 3 & 3 & 1 \\
 C_4^0 & C_4^1 & C_4^2 & C_4^3 & C_4^4 & & 1 & 4 & 6 & 4 & 1
 \end{array}$$

Para preencher o triângulo, usa-se o teorema (Relação de Stifel)

$$C_n^p + C_n^{p+1} = C_{n+1}^{p+1}$$

Por exemplo, para encontrar o termo C_4^1 soma-se os termos C_3^0 e C_3^1 .

Para demonstrar a relação de Stifel:

Considere um conjunto A de $n + 1$ elementos, um dos quais é x . O número de subconjuntos de A com $p + 1$ elementos é C_{n+1}^{p+1} . Esse número é igual à soma do número de subconjuntos nos quais x não figura, C_n^{p+1} , com o número de subconjuntos nos quais x figura, C_n^p (CARVALHO et al., 2016, p. 135).

Outra forma de provar a Relação de Stifel é algebricamente.

$$\begin{aligned}
 C_n^p + C_n^{p+1} &= \frac{n!}{(n-p)! \cdot p!} + \frac{n!}{(n-p-1)! \cdot (p+1)!} \\
 &= \frac{n!}{(n-p)! \cdot p!} \cdot (p+1) + \frac{n!}{(n-p-1)! \cdot (p+1)!} \cdot (n-p) \\
 &= \frac{n! \cdot (p+1) + n! \cdot (n-p)!}{(n-p)! \cdot (p+1)!} = \frac{n! \cdot (p+1+n-p)}{(n-p)! \cdot (p+1)!} = \frac{n! \cdot (n+1)}{(n-p)! \cdot (p+1)!} \\
 &= \frac{(n+1)!}{(n-p)! \cdot (p+1)!} = C_{n+1}^{p+1}
 \end{aligned}$$

O Triângulo de Pascal poderia ser utilizado para determinar a quantidade de erros possíveis da aluna no teste mencionado. Bastava observar a 11ª linha do triângulo. Embora o teste tenha 10 questões, a contagem no triângulo é a partir do zero, por isso utiliza-se a 11ª, a saber

$$1 \quad 10 \quad 45 \quad 120 \quad 210 \quad 252 \quad 210 \quad 120 \quad 45 \quad 10 \quad 1$$

Considere que os alunos não estudaram nada. Toda resposta será aleatória. O Triângulo de Pascal também define a distribuição das notas dos alunos. Assim, em uma turma, tem-se algo próximo dessa distribuição. Um aluno não errou nada, cerca de 10 erraram apenas uma, então acertaram 90%. Se um aluno errou duas, garantiu 80% da prova e assim por diante.

Neste contexto, a probabilidade de um aluno acertar 80% ou mais da prova é próxima de 75%. O total de formas de responder essa prova é 2^{10} , uma vez que para cada uma das 10 questões cabem duas respostas. Os 25 alunos podem responder de qualquer uma das 1024 formas.

No caso do desempenho dos 25 alunos, será analisada a probabilidade complementar. Ao invés de calcular a probabilidade de pelo menos um atingir 80% ou mais na prova, o evento analisado será nenhum aluno atingir este índice.

O espaço amostral Ω de interesse será formado por 25-uplas cujas componentes são possíveis respostas apresentadas por cada um dos 25 alunos. A cardinalidade de Ω é igual a 1024^{25} . Pelo Triângulo de Pascal, há 56 casos para os quais os alunos tirariam 80% ou mais. São os primeiros três números da 11^a linha. Sobram, 968 resultados abaixo de 80%. Dessa forma, a probabilidade de nenhum aluno atingir 80% ou mais na prova¹¹ é:

$$\frac{968^{25}}{1024^{25}} \cong 0,24512$$

Em uma prova na qual todos os alunos não tem ideia do conteúdo dela, suas respostas são aleatórias. A probabilidade de não atingirem nem 80% corresponde a, aproximadamente, 25%. Para alcançar a probabilidade de pelo menos um aluno acertar 80% ou mais, tem-se $1 - 0,24512 = 0,75488$. Aproximadamente 75%.

3.3 Outra analogia a moedas

Bill Miller foi considerado o maior administrador de fundos dos anos 90 do século XX nos Estados Unidos, pois conseguiu ter excelentes resultados durante um período de 15 anos. Cogitava-se que seu desempenho não poderia ser obra de mero acaso (MLODINOW, 2009, p. 189). Para analisar esta situação, faremos uma analogia com o lançamento de moedas. Consideraremos como bem sucedido o administrador que obtiver uma sequência de 15 caras seguidas. Se for considerada a quantidade de administradores de fundo daquele país, a probabilidade de um deles tirar 15 caras seguidas é $\frac{1}{32.768}$, mas ao se considerar que, em um grupo de 1.000 administradores de fundos, por exemplo, um deles consiga o resultado de 15 caras seguidas, a probabilidade é muito maior, cerca de 3% (MLODINOW, 2009, p. 192). O excelente desempenho quando visto individualmente é raro, mas não é surpreendente que em um grupo grande de administradores ao menos um deles apresente um excelente desempenho.

Ao se lançar um moeda há dois resultados possíveis, cara ou coroa, fazendo isso

¹¹ os cálculos foram feitos no *octave*.

15 vezes, tem-se um total de $2^{15} = 32.768$ permutações, mas apenas uma delas é sair apenas caras, logo, a probabilidade de uma pessoa jogar a moeda 15 vezes e somente sair caras é de $\frac{1}{32.768}$. Será calculada a chance de que algum dos mil administradores que começaram a jogar moedas em 1991 tenha tirado 15 caras seguidas. Para caracterizar o espaço amostral, considere o conjunto S de todas as matrizes $M = (m_{i,j})$ de ordem 15×1.000 tais que $m_{i,j} \in \{cara, coroa\}$. A j -ésima coluna da matriz M representa a sequência de 15 lançamentos do j -ésimo administrador.

Todos os possíveis cenários de lançamentos estão representados no conjunto¹² S e $|S| = 2^{15 \times 1.000}$. Sejam $A_j = \{M \in S, m_{k,j} = cara, k = 1, \dots, 15\}$. Isto é, é o conjunto de 15 lançamentos de 1.000 pessoas nos quais a j -ésima pessoa tirou 15 caras seguidas. Note que $|A| = (2^{15})^{1.000-1}$. No conjunto A , o resultado de uma delas já está definido. Assim, sobram 999 outros candidatos que podem apresentar qualquer outro resultado, pois basta que apenas um tenha atendido ao evento. A proeza de tirar apenas caras em 15 lançamentos de uma moeda.

A chance de algum dos mil administradores tirar 15 caras seguidas ao lançar uma moeda uma vez por ano durante 15 anos é dada por:

$$\frac{|\cup_{j=1}^{1.000} A_j|}{|S|}$$

Para calcular a quantidade de elementos do conjunto no numerador, será utilizado o princípio da inclusão-exclusão ($n = 1.000$):

$$\begin{aligned} |\cup_{j=1}^n A_j| &= \sum_{j=1}^n |A_j| - \sum_{1 \leq j_1 < j_2 \leq n} |A_{j_1} \cap A_{j_2}| + \sum_{1 \leq j_1 < j_2 < j_3 \leq n} |A_{j_1} \cap A_{j_2} \cap A_{j_3}| \\ &\quad + \dots + (-1)^{n-1} |A_{j_1} \cap A_{j_2} \cap \dots \cap A_n|. \end{aligned}$$

Note que

$$|A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_k}| = (2^{15})^{n-k}.$$

A interseção entre estes conjuntos é composta dos resultados que são iguais para as j -ésimas pessoas. Dessa forma, há indecisão para os demais $j - k$ jogadores.

¹² $|S|$ representa o número de elementos do conjunto S

Assim

$$\begin{aligned}
|\cup_{j=1}^n A_j| &= \sum_{j=1}^n (2^{15})^{n-1} - \sum_{1 \leq j_1 < j_2 \leq n} (2^{15})^{n-2} + \sum_{1 \leq j_1 < j_2 < j_3 \leq n} (2^{15})^{n-3} + \dots + (-1)^{n-1} \\
&= \binom{n}{1} (2^{15})^{n-1} - \binom{n}{2} (2^{15})^{n-2} + \binom{n}{3} (2^{15})^{n-3} - \dots + \binom{n}{n} (-1)^{n-1} (2^{15})^{n-n} \\
&= - \left[\sum_{k=1}^n \binom{n}{k} (-1)^k (2^{15})^{n-k} - \binom{n}{0} (-1)^0 (2^{15})^n + (2^{15})^n \right] \\
&= (2^{15})^n - (2^{15} - 1)^n.
\end{aligned}$$

Vale observar que na penúltima linha da demonstração há o desenvolvimento do binômio de Newton $(2^{15} - 1)^n$. Logo, a chance de que algum dos mil administradores tire 15 caras seguidas ao lançar uma moeda uma vez por ano durante 15 anos é dada por:

$$\frac{|\cup_{j=1}^{1.000} A_j|}{|S|} = \frac{(2^{15})^{1.000} - (2^{15} - 1)^{1.000}}{(2^{15})^{1.000}} = 1 - \frac{(2^{15} - 1)^{1.000}}{(2^{15})^{1.000}} = 1 - \left(1 - \frac{1}{2^{15}}\right)^{1.000}.$$

Que é cerca¹³ de 3%.

Vamos recapitular alguns detalhes do cálculo acima. Foi usada uma matriz para relacionar dois aspectos do experimento. Cada administrador foi associado a uma coluna da matriz A . Cada linha da matriz A corresponde a um lançamento de moeda por cada um dos 1.000 administradores. A matriz A_j contém uma sequência de 15 caras, ou seja, o j -ésimo administrador foi bem sucedido. Para cada administrador, há 2^{15} possíveis resultados distintos ao jogarem uma moeda 15 vezes. Como são 1.000 indivíduos, tem-se $2^{15 \times 1.000}$ resultados diferentes. O sucesso do j -ésimo administrador aparece combinado a 999 dos demais. Assim, há variação em $2^{15 \times 1.000 - 1}$. Seria 2^{15} para cada um dos demais 999 participantes. Cada um destes possíveis resultados está representado por um elemento do conjunto de matrizes A_j .

Considere $n = 1.000$. O conjunto $\cup_{j=1}^n A_j$ reúne todas as possibilidades de administradores bem sucedidos. Note que os conjuntos A_j não são disjuntos pois cada um deles reúne no conjunto de 15 anos, os casos em que há pelo menos um (o j -ésimo) administrador bem sucedido. Mas podem haver dois, três, ..., até mil administradores bem sucedidos. Foram determinados quantos elementos há na união de todos os A_j . Pelo Princípio da Inclusão-Exclusão, para calcular $|\cup_{j=1}^n A_j|$,

¹³ Foi utilizado o *octave* para efetuar os cálculos:

```

octave:1> 1-(1-1/2^(15))^1000
ans = 0.0300570695428242

```

- somou-se a quantidade de elementos de cada um dos A_j (conjunto em que há pelo menos um administrador bem sucedido, o j -ésimo) $-\sum_{j=1}^n |A_j|$
- foram subtraídas todas as possibilidades em que há pelo menos dois administradores bem sucedidos $-\sum_{1 \leq j_1 < j_2 \leq n} |A_{j_1} \cap A_{j_2}|$
- somou-se a quantidade de casos em que há pelo menos três administradores bem sucedidos e assim por diante até chegar a uma possibilidade, todos os administradores serem bem sucedidos: conseguirem tirar somente caras em 15 jogadas.

Na igualdade $\sum_{1 \leq j_1 < j_2 \leq n} (2^{15})^{n-2} = \binom{n}{2} (2^{15})^{n-2}$, vê-se que para todas as duplas possíveis, tem-se $\binom{n}{2}$ formas de escolhê-las. Para estes administradores, haverá resultados definidos, apenas caras, mas para os demais 998, serão $(2^{15})^{1.000-2} \cdot 1 \cdot 1 = 2^{15 \times 998} \cdot 1 \cdot 1$ possíveis resultados. O argumento para as demais parcelas é análogo.

4 PROBABILIDADE CONJUNTA E CONDICIONAL

A pior forma de desigualdade é tentar fazer duas coisas diferentes iguais.

Aristóteles

Este capítulo trata de exemplos nos quais os eventos se relacionam ou não. Há casos de probabilidade condicional, teorema do produto e até de eventos independentes. Os exemplos de Mlodinow (2009) não são como os convencionais de probabilidade. Mais uma vez, o autor omite dados que uma nota de rodapé poderia orientar na resolução dos mesmos. No entanto, são criativos, fogem do trivial utilizado nos livros que apresentam este conteúdo. O objetivo deste capítulo é analisar cada exemplo. Dessa forma, conseguir preencher as lacunas entre eles e as respostas que o autor afirma corresponderem às indagações.

Steyer e Nagel (2017) salientam a importância em distinguirmos entre o significado matemático ou formal dos termos probabilísticos e o significado desses termos quando usados em uma aplicação da teoria da probabilidade para um fenômeno concreto do mundo real. O conceito intuitivo de um evento geralmente implica que os eventos são ordenados em relação ao tempo. Ou seja, na linguagem comum, um “evento” é anterior, simultâneo ou posterior a outro “evento”. No entanto, eventos no contexto da teoria da probabilidade não são necessariamente ordenados em relação ao tempo. Ainda assim, nas aplicações um ordenamento é importante e até vital para haver sentido do fenômeno descrito (STEYER; NAGEL, 2017).

Definição 1. *Dizemos que dois eventos A e B definidos no mesmo espaço amostral são independentes se e somente se*

$$P(A \cap B) = P(A) \cdot P(B). \quad (1)$$

Se dois eventos são independentes, a probabilidade de ambos ocorrerem será o produto da probabilidade de cada um calculado de forma isolada. Quando aplicado a eventos, o termo “independente” é usado para designar eventos que satisfazem (1). Por vezes, apenas somos capazes de decidir se dois eventos são ou não independentes após o cálculo das três probabilidades em (1) (KALBFLEISCH, 1985).

Exemplo 1. *Uma carta é selecionada de um baralho de 52 cartas. Considere o evento A , a carta é um ás e o evento B , a carta é do naipe de espadas.*

Supondo equiprobabilidade das 52 cartas, temos

$$P(A \cap B) = \frac{1}{52}, \quad P(A) = \frac{4}{52}, \quad P(B) = \frac{13}{52}.$$

Portanto, (1) é satisfeita. Vemos que quando as 52 cartas são equiprováveis, os eventos A e B são independentes.

Suponha que a carta do ás de espadas tenha sido marcada e que agora o jogador que distribui as cartas seja capaz de fazer com que o ás de espadas saia duas vezes mais que qualquer outra carta do baralho. Assim a probabilidade do ás de espadas sair é $\frac{2}{53}$ enquanto que a de qualquer outra carta é $\frac{1}{53}$. Temos

$$P(A \cap B) = \frac{2}{53}, \quad P(A) = \frac{5}{53}, \quad P(B) = \frac{14}{53}.$$

Neste caso, $P(A \cap B) \neq P(A)P(B)$ e os eventos A e B não são independentes.

Ainda considerando eventos A e B definidos no mesmo espaço amostral, queremos agora definir a probabilidade do evento B ao restringirmos nossa atenção ao subconjunto do espaço amostral definido pelo evento A .

Definição 2. *Sejam A e B eventos definidos no mesmo espaço amostral tais que $P(A) > 0$. Definimos a probabilidade condicional de B dado A , $P(B|A)$ por*

$$P(B|A) = \frac{P(A \cap B)}{P(A)}. \quad (2)$$

Quando $P(A) = 0$, $P(B|A)$ não está definida. Ilustraremos com exemplos a relação entre $P(B|A)$ e $P(B)$. Note que se A e B são eventos independentes, segue de (1) e (2) que $P(B|A) = P(B)$. Isto é, a ocorrência de A não afeta a probabilidade de B .

Exemplo 2. *Considere um dado no formato de um octaedro com faces numeradas de 1 a 8. Associe a ele o espaço amostral equiprovável*

$$\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

Sejam A o evento: sair uma face ímpar e o evento B : sair ao menos 7.

Temos

$$P(A) = \frac{4}{8} = \frac{1}{2}, \quad P(B) = \frac{2}{8} = \frac{1}{4}, \quad A \cap B = \{7\}.$$

Logo

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{2} = P(A).$$

Observe também que

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1}{4} = P(B).$$

Exemplo 3. No mesmo espaço amostral Ω , considere os eventos

$$A+ = \{1, 2, 3, 5, 7\} \quad e \quad A- = \{3, 5, 7\}$$

Temos

$$P(B|A+) = \frac{P(A+ \cap B)}{P(A+)} = \frac{1}{5} < P(B)$$

e

$$P(B|A-) = \frac{P(A- \cap B)}{P(A-)} = \frac{1}{3} > P(B).$$

Estes exemplos nos mostram que as probabilidades $P(B)$ e $P(B|A)$ podem ser iguais ou qualquer uma delas maior do que a outra.

4.1 Policial Divorciado e Morto

Suponha que uma pessoa casada tenha, em média, uma chance de aproximadamente $\frac{1}{50}$ de se divorciar a cada ano. Por outro lado, um policial tem uma chance de aproximadamente $\frac{1}{5.000}$ de morrer em serviço. Qual é a probabilidade de que um policial casado se divorcie e morra no mesmo ano? Segundo o princípio acima, se tais eventos forem independentes, a probabilidade será de aproximadamente $\frac{1}{50} \cdot \frac{1}{5.000}$, que equivale a $\frac{1}{250.000}$. É claro que os eventos não são independentes; na verdade, estão ligados: depois de morrer, droga, o policial não pode mais se divorciar. Assim, a chance de que ele tenha tanto azar é um pouco menor que $\frac{1}{250.000}$. (MLODINOW, 2009, p. 42)

Considere os eventos $D = \text{se divorciar}$ e $M = \text{policial morrer}$. Como Mlodinow (2009) informa a não independência dos eventos, vamos calcular a probabilidade da interseção usando a probabilidade condicional (2):

$$P(M \cap D) = P(M|D) \cdot P(D).$$

Ao afirmar que essa probabilidade é menor do que $P(M) \cdot P(D)$, Mlodinow (2009) nos permite concluir que¹⁴

$$P(M|D) < P(M).$$

¹⁴ $P(M|D) \cdot P(D) < P(D) \cdot P(M)$

No entanto, não é possível entender de imediato o que permite que ele faça esta afirmação.

Antes de prosseguir, serão introduzidas algumas noções. Diz-se que um evento A de um espaço amostral Ω é binomial se ele tem probabilidade p de ocorrer e a probabilidade do evento complementar, $\Omega - A$, tem probabilidade $1 - p$. Isto é, temos apenas dois possíveis resultados. Observe que os eventos M e D são binomiais. (IRETON-JONES; GOTTSCHLICH; BELL, 1998). O coeficiente de correlação entre dois eventos binomiais A e B é definido por

$$\rho_{AB} = \frac{P(A \cap B) - P(A)P(B)}{\sqrt{P(A)(1 - P(A))}\sqrt{P(B)(1 - P(B))}}. \quad (3)$$

Ele mensura a força de associação entre os dois eventos e a direção dessa associação. Quando ele é positivo a ocorrência de um evento reforça a do outro; quando é negativo a ocorrência de um evento enfraquece a do outro.

Os eventos M e D estão negativamente correlacionados. Isto é, o numerador em (3) é negativo:

$$P(M \cap D) - P(M)P(D) < 0.$$

4.2 Caso dos Collins

No ano de 1964 na cidade de Los Angeles, a sr.(a) Juanita Brooks foi empurrada pelas costas e teve a bolsa roubada. Ela conseguiu se levantar e viu uma mulher loira correndo. O vizinho Jonh Bass olhou na direção dos gritos. Disse ter visto uma mulher que entrou correndo em um carro amarelo estacionado. O carro partiu e passou ao lado do sr. John. Ele conseguiu ver que o motorista era um homem negro com barba e bigode. No entanto, a sr.(a) Juanita não viu o homem e o vizinho não viu a mulher. Além disso, outras testemunhas apresentaram informações desencontradas sobre a cor do veículo. Uns diziam que era amarelo, outros que era amarelo com teto cinzento e até amarelo com teto branco. Dias depois um policial encontrou um carro amarelo com teto cinza estacionado em frente à uma casa do bairro. Verificou que os donos se encaixavam na descrição dos suspeitos. O casal Collins foi preso e levado ao tribunal. Como não havia provas contundentes contra o casal, a acusação consultou um professor de matemática que apresentou algumas probabilidades, são elas (MLODINOW, 2009, p. 46):

O professor afirmou que a regra do produto valia para essa situação. Assim, a probabilidade de encontrar um casal que atendesse a descrição era de $\frac{1}{12.000.000}$. No entanto, essas características apresentam interseções não vazias.

Um homem que tem barba, em geral, tem bigode também. Se um homem negro tem barba, a probabilidade de ele ter bigode não é de $\frac{1}{4}$, mas sim de quase 100%. Neste

Tabela 1 - Caso dos Collins

CARACTERÍSTICAS	PROBABILIDADE INDIVIDUAL
Automóvel parcialmente amarelo	1/10
Homem com bigode	1/4
Homem negro com barba	11/10
Garota com rabo de cavalo	1/10
Garota loira	1/3
Casal inter-racial num carro	1/1.000

Fonte: A autora, 2017.

caso, pode-se retirar a probabilidade de um “homem negro com barba”, dessa forma, o valor final cairia para $\frac{1}{1.200.000}$. O livro não cita, mas a ré é descrita usando um penteado “rabo de cavalo”, por isso a probabilidade desta característica entra no cálculo.¹⁵ Além disso, não fica clara a origem das probabilidades destas características.

O que é relevante para o julgamento não é o cálculo da probabilidade de um casal reunir essas características. O importante é dado que um casal apresente essas características, qual a chance de ser culpado? No bairro onde ocorreu o crime, havia alguns milhares de habitantes, Assumindo a probabilidade citada pela acusação, $\frac{1}{1.200.000}$, poderia-se encontrar dois, ou talvez três casais com as mesmas características dos Collins. Assim, a probabilidade de o casal Collins ser culpado é de $\frac{1}{2}$ ou $\frac{1}{3}$ (MLODINOW, 2009, p. 49).

Pode-se considerar dois eventos: A é um casal se enquadrar na descrição dos suspeitos e B o casal ser culpado. $P(A/B) = 1$, pois dados que o casal é culpado, possui as características citadas. 100% de chance. No entanto, por estimativa, $P(B/A) = \frac{1}{2}$ ou $P(B/A) = \frac{1}{3}$. A acusação tentou confundir o júri entre estas probabilidades condicionais. Probabilidades muito significativas para deixar o júri em dúvida para acusar um casal sem provas concretas.

4.3 Duas Meninas

Se uma mãe está grávida de gêmeos, qual a probabilidade de nascerem duas meninas? Seria $\frac{1}{4}$. Poderia-se mostrar isso apresentando todos os elementos do espaço amostral: (menina, menina), (menina, menino), (menino, menina) e (menino, menino). Se a pergunta fosse, dado que um bebê é menina, qual a probabilidade de nascerem duas meninas? É importante analisar que não se menciona qual deles é menina. Se é o primeiro

¹⁵ <http://www.law.harvard.edu/publications/evidenceiii/cases/people.htm>

ou o segundo bebê. Para fazer isso pode-se utilizar dois métodos (MLODINOW, 2009, p. 61):

- **Espaço Amostral**

Nesta situação, somente um caso não atende a descrição, (menino,menino), logo, a probabilidade é de $\frac{1}{3}$.

- **Probabilidade Condicional**

Neste método, usaremos a fórmula (2):

Evento A = **nascer uma menina**

Evento B = **nascerem duas meninas**

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

No entanto, a pergunta poderia ser, dado que o **primeiro** bebê é menina, qual a probabilidade de ambos serem meninas? (MLODINOW, 2009, p. 61) Serão utilizados dois métodos:

- **Espaço Amostral**

Nesta situação, só há dois casos do espaço amostral que satisfazem o evento: (menina, menino) ou (menina, menina). A probabilidade é de $\frac{1}{2}$.

- **Probabilidade Condicional**

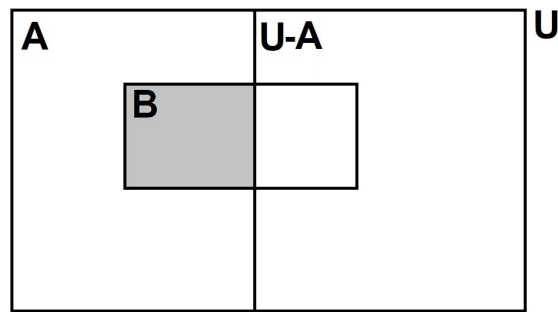
Neste método, usaremos a fórmula já citada:

Evento A = **a primeira ser menina**

Evento B = **nascerem duas meninas**

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2}$$

Figura 4 - Relação entre conjuntos



Fonte: A autora, 2017.

4.4 Ler Mentes

Mlodinow (2009) cita um caso de um curioso estudante. Nos anos 70 do século XX, em Harvard, um aluno de psicologia relatou ao seu professor algumas situações inusitadas. Entre elas, que sua esposa conseguia dizer as palavras em que ele estava pensando. Ele acreditava que havia uma conspiração contra ele.

A probabilidade de se escolher uma pessoa de forma aleatória e ela ter problemas psiquiátricos é pequena. Da mesma forma, é pouco provável que uma pessoa escolhida ao acaso acredite que sua esposa pode ler mentes. No entanto, a probabilidade de uma pessoa acreditar que sua esposa lê mentes uma vez que essa pessoa tem problemas psiquiátricos é alta. Assim como, ter problemas psiquiátricos, dado que pensa que a esposa pode ler mentes.

Seja o evento A acreditar que a esposa leia mentes. O evento B é ter problemas psiquiátricos. Como são eventos positivamente correlacionados, a probabilidade de ambos ocorrerem é maior do que o produto das probabilidades de A e de B . A Figura 4 ilustra esse fato. Ela ilustra também a relação entre $P(B|A)$ e $P(B)$ quando A e B estão positivamente relacionados. Quando se reduz o espaço amostral a pessoas com problemas psiquiátricos está se aplicando a probabilidade condicional. É possível concluir o que a intuição já definiu. Encontrar alguém que acredite em leitura de mentes fica mais provável. Ao calcularmos $P(B|A)$, os casos favoráveis são todos os elementos do conjunto A que estão no conjunto B e o espaço amostral é o conjunto A , visto que este evento já ocorreu.

4.5 Menos é mais

Os cientistas Daniel Kahneman e Amós Tversky desenvolveram experimentos para identificar a noção que as pessoas têm de probabilidade. Apresentaram a elas asserções sobre uma personagem fictícia e pediram que indicassem a mais provável. Curiosamente,

as asserções indicadas como mais prováveis eram as que relacionavam duas afirmações. Por exemplo, as pessoas afirmaram que era mais provável que “Linda fosse bancária e participasse do movimento feminista” do que a afirmação “Linda participa do movimento feminista”. Esse erro de julgamento em pequena escala ilustra um risco maior: aferir probabilidade sobre eventos desconhecidos pode gerar julgamentos errados (MLODINOW, 2009, p. 35). Neste caso, a intuição não é confiável. Em geral, as pessoas pensam que as afirmações com mais dados são mais prováveis. Acreditam que dois eventos associados apresentam probabilidade maior do que apenas um deles isolado. A maior quantidade de dados dá uma falsa ideia de veracidade. No entanto, é o evento isolado que é mais provável.

O que é mais provável: que um réu, depois de encontrar um corpo, deixe a cena do crime, ou que um réu, depois de encontrar um corpo, deixe a cena do crime porque teme ser acusado pelo macabro assassinato? É mais provável que o presidente aumente os gastos federais com educação ou que aumente os gastos federais com educação utilizando fundos obtidos pelo corte de outros gastos dirigidos aos estados? É mais provável que uma empresa aumente suas vendas no ano que vem ou que aumente as vendas no ano que vem porque a economia em geral passará por um bom ano? Em todos os casos, embora a segunda opção seja menos provável que a primeira, pode parecer mais provável (MLODINOW, 2009, p. 33).

Em todos esses exemplos, há a mesma situação. Comparam-se as probabilidades de um evento isolado com a associação dele a outro. Como consequência de (2), vemos que a probabilidade conjunta $P(A \cap B)$ é sempre menor do que $P(A)$, $P(B)$, $P(B|A)$ ou $P(A|B)$.

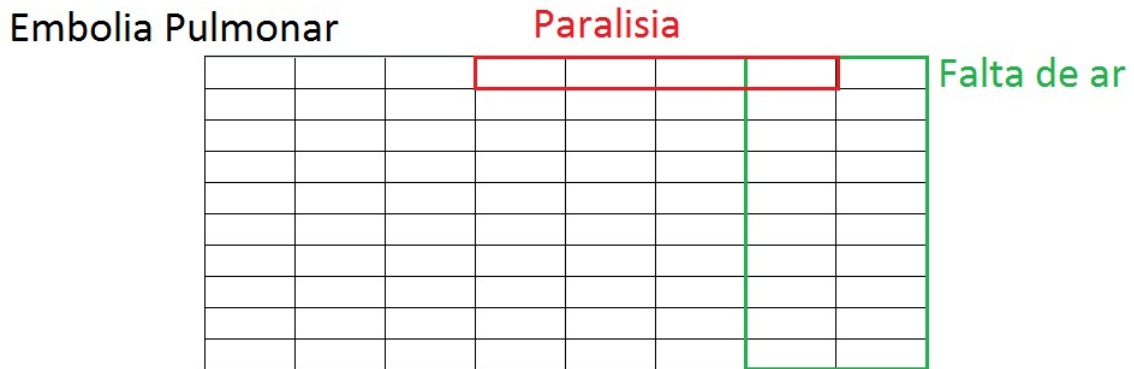
Sobre o caso do réu, o primeiro evento é deixar a cena do crime o segundo, ser culpado. São eventos independentes. Eventos independentes são eventos que podem ocorrer sem que um influencie o outro. Ao se lançar dois dados, o resultado de um “não afeta” o número que sairá no outro. O réu pode deixar a cena por diversos motivos, por estar sendo ameaçado, para chamar ajuda ou até mesmo, por ser culpado.

Outro experimento feito por Daniel Kahneman e Amós Tversky consistiu em apresentar uma situação de um paciente hipotético a um grupo de médicos. O suposto enfermo tinha embolia pulmonar e possíveis sintomas desenvolvidos. Um deles é raro, paralisia parcial e outro comum, falta de ar. Os médicos acreditaram que era mais provável o paciente apresentar ambos do que apenas a paralisia (MLODINOW, 2009, p. 34). Considere a paralisia parcial como o evento A e a falta de ar como o evento B . A e B são eventos independentes.

Suponha $P(A) = \frac{1}{20}$ e $P(B) = \frac{1}{4}$. A probabilidade de manifestação apenas do raro é de $\frac{1}{20}$. Pelo teorema do produto, a probabilidade de ambos ocorrerem é de $\frac{1}{20} \cdot \frac{1}{4} = \frac{1}{80}$. Dessa forma, pode-se concluir que é menos provável que ambos os sintomas ocorram.

Pode-se analisar o teorema do produto pelo ponto de vista da geometria. Na figura 5, há os conjuntos dos casos de embolia analisados e de cada evento. Utiliza-se

Figura 5 - Abordagem geométrica



Fonte: A autora, 2017.

cada quadrado como uma unidade. Assim, as probabilidades supostas são representadas na figura. Verifica-se que a interseção entre os dois sintomas é menor do que a região do evento A , “Paralisia”.

Ainda sobre eventos independentes, pode-se citar o processo judicial de Paula Jones contra Bill Clinton. Na época, ele era presidente dos Estados Unidos. Ela o acusava de assédio sexual.¹⁶ 200 advogados foram convidados a estimar se o litígio terminaria brevemente. Para um grupo foi perguntado apenas isso. Para outro, foram apresentadas hipóteses para tal fato. Acordo, retirada da queixa ou encerramento por parte do juiz foram algumas das justificativas. O segundo grupo ficou mais inclinado a acreditar no fim prematuro (MLODINOW, 2009, p. 34).

Pode-se separar cada situação em dois eventos: A e B . O primeiro seria o processo acabar rapidamente. O segundo qualquer uma das situações que ocorreriam junto. B poderia ser, haver acordo, por exemplo. Além das hipóteses de evento B , poderia haver a morte de uma das partes, o surgimento de um fato novo entre outras. Dizer que o processo irá acabar por acordo é escolher uma das muitas possibilidades. O evento não ficará mais provável de acontecer porque foi apresentada uma justificativa. A interseção de dois eventos é menos provável do que apenas um deles ocorrer.

Mlodinow (2009) questiona por que os gregos antigos não definiram a probabilidade. Alguns fatores contribuíram para isso, um deles é a religião. Acreditavam que o que acontecia em suas vidas era a vontade dos deuses. Não havia a noção de aleatório. Também não sabiam aritmética. Apesar de terem desenvolvido a base da Matemática que temos até hoje, não podiam calcular resultados mais prováveis em simples jogos.

A probabilidade de, ao jogar um dado, sair o número 6 é de $\frac{1}{6}$. Ao se arremessarem

¹⁶ <http://www1.folha.uol.com.br/fsp/mundo/ft30079904.htm>

dois dados, a probabilidade de sair 6 em ambos é de $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} \cong 0,03$, ou seja, 3% de chance de conseguir.

Diante do exposto, conclui-se que é mais provável que A ocorra do que a interseção $A \cap B$. O que é apresentado como justificativa, na verdade é um segundo evento. Ele dá a falsa sensação de veracidade ao primeiro. Assim, ambos ocorrerem é a interseção de eventos independentes.

5 DIANTE DO ACASO

*Tudo é número.
Pitágoras*

Os exemplos apresentados neste capítulo mostram como as pessoas lidam com o conceito de aleatório. A ideia é mostrar que, apesar dos experimentos serem randômicos, ainda sim, as pessoas buscam padrões nos resultados. Deseja-se exibir soluções lógicas de cada exemplo, mas também como conceitos da psicologia interferem na opinião das pessoas. Mesmo aquelas que possuem conhecimentos matemáticos.

A psicologia tem interesse em compreender o comportamento humano diante do acaso. Ela tem especial interesse pelo chamado “matching”. Não há consenso entre os psicólogos em caracterizar essa estratégia como um equívoco cometido por falta de compreensão da teoria da probabilidade ou como um traço inerente do comportamento humano.

Koehler e James (2014) descrevem uma experiência na qual havia 10 pares de copos, em cada par, um verde e um vermelho. Os pesquisadores usaram um dado de 10 faces, sendo 7 verdes e 3 vermelhas. Lançavam-no e escondiam uma moeda embaixo do copo do par cuja cor correspondia a qual saíra no dado. Sendo assim, a sequência de cores para o esconderijo das moedas foi aleatória. Os participantes tinham que adivinhar em qual copo de cada dupla a moeda estava escondida. Foi combinado que quem encontrasse moedas poderia ficar com elas. Obviamente que a ideia dos jogadores era maximizar sua quantidade de acertos.

Para alcançar seus objetivos, as pessoas optaram por duas estratégias. Uma optavam sempre pela cor cuja probabilidade era maior de sair. Neste caso, a verde. Outras preferiram corresponder suas escolhas às probabilidades de cada cor (“probability matching”). Seus palpites seguiram 70% para o verde e 30% para o vermelho. Esses comportamentos eram recorrentes em experiências nas quais havia escolhas binárias com predominância da segunda estratégia.

A estratégia de optar sempre pelo verde é a estratégia ótima, ela maximiza a quantidade de acertos. Ela permite acertar, em média, 70% dos palpites. No caso do “matching”, o participante precisa optar por uma cor e de fato ela ser a certa. Nesta estratégia, o participante opta pelo verde cerca de 70% das vezes e há 70% de chance de a cor correta ser essa. Já para o vermelho, ele é escolhido cerca de 30% das vezes e a chance desta escolha estar certa é de 30%. No “matching”, há em média, 58% de chance de acerto, pois $0,7 \cdot 0,7 + 0,3 \cdot 0,3 = 0,58$.

Do ponto de vista probabilístico, o “matching” é uma escolha equivocada, pois estamos diante um evento aleatório. Não há um padrão, por isso, não é possível definir

em qual jogada seria o momento de se optar pela cor menos provável. Uma vez que o experimento é randômico, a melhor opção é a de maximizar.

Koehler e James (2014) afirmam que pessoas que tinham um melhor desenvolvimento cognitivo optaram pela maximização em experimentos de mesma natureza. O fato de terem conhecimentos matemáticos contribuía para a escolha, mas não era o fator principal. No entanto, em experiências feitas com crianças pequenas, animais e pessoas que estavam com índice de glicose baixo, observaram que a maximização era a estratégia preferida. Esses resultados contradiziam a ideia de que a maximização era uma estratégia complexa. Indicavam que era uma ação da intuição.

Pesquisadores verificaram que, na maioria dos casos, se a pessoa tivesse que dar um único palpite, optava pela estratégia de maximização. No exemplo dos copos, escolheria a cor verde, por ser a mais provável. No entanto, em sequências de palpites, optavam pelo “matching”.

Koehler e James (2014) afirmam que os pesquisadores Gaissmaier e Schooler (2008 apud KOEHLER; JAMES, 2014) fizeram uma experiência de previsão binária em duas etapas. Na primeira, apresentaram sequências aleatórias. Nesta fase, identificaram os participantes que eram adeptos da maximização e os que preferiam o “matching”. Na segunda, as sequências não eram aleatórias, apresentavam algum padrão. Os que utilizavam o “matching” conseguiram explorar melhor o padrão dos que os demais.

Ainda não é consenso entre os pesquisadores qual dos dois métodos é mais elaborado em relação a ganho cognitivo. Alguns acreditam que o “matching” é uma estratégia superior. Estão embasados no argumento da maximização ser uma reação intuitiva. Além disso, houve experimentos que demonstraram existir um maior esforço cognitivo ao se aplicar o “matching”. No entanto, apesar de ser uma estratégia superior, é usado de forma incorreta. Em parte, se deve ao fato das pessoas não acreditarem que lidam com sequências randômicas.

Mlodinow (2009) afirma que muitos têm dificuldade de lidar com o que é aleatório. Isso é verdade, mas não é a única causa para escolherem o “matching”. O autor não deixa claro qual o mecanismo que exibia as cores aos participantes. Admitindo que fosse um computador que acendesse as lâmpadas, as pessoas poderiam ter dúvida se de fato seria um experimento aleatório. Elas poderiam ficar tentadas a entender que a programação da máquina iria influenciar em algum tipo de padrão. Não estariam de todo erradas.

É muito difícil recriar experimentos usando computadores cujos resultados são aleatórios.¹⁷ Os programas computacionais conseguem gerar números “pseudo-aleatórios”. Já o site <random.org> garante que pode gerar números realmente aleatórios. De forma mais simples, pode-se garantir a randomicidade do experimento com bolas em uma caixa.

¹⁷ <https://www.ime.usp.br/pf/algorithmos/aulas/random.html>

Ao invés de se acender lâmpadas, bolas seriam sorteadas. Neste caso, para que a probabilidade de sair cada cor fosse a mesma, seria necessário haver a mesma quantidade de cada cor. Admitindo que a pessoa que as sorteie não tenha como vê-las ou diferenciá-las de qualquer forma a não ser pela cor após o sorteio.

5.1 Estratégias diante do acaso

Newell e Schulze (2017) descrevem um experimento como os anteriores. Há duas escolhas (cores branca e preta) e os participantes precisam dar palpite sobre qual delas irá sair. As pessoas conhecem as probabilidades de cada cor, mas costumam utilizar estratégias diferentes¹⁸:

Considere um dado com 4 faces pretas e 2 brancas que será lançado 60 vezes. Antes de cada jogada, uma pessoa será perguntada sobre qual

¹⁸ Consider the following problem (referred to hereafter as the “die problem”): A die with 4 black faces and 2 white faces will be rolled 60 times. Before each roll you will be asked to predict which color (black or white) will show up once the die is rolled. You will be given 1\$ for each correct prediction. Assume that you want to make as much money as possible. What strategy would you use in order to make as much money as possible by making the most correct predictions?

1. Go by intuition, switching when there have been too many of one color or the other.
2. Predict the more likely color (black) on most of the rolls but occasionally, after a long run of blacks, predict white.
3. Make predictions according to the frequency of occurrence (4 of 6 for black and 2 of 6 for white). That is, predict twice as many blacks as whites.
4. Predict the more likely color (black) on all of the 60 rolls.
5. Predict more black than white, but switching back and forth depending on “runs” of one color or the other.

The description of this problem tells you everything you need to know in order to make the rational choice. The relative pay-off probabilities are known, the payoff amount is known, and assuming you believe that the die is fair and will not be altered throughout the 60 rolls, then you know that pay-off probabilities will remain stationary.

Yet, people seem to find this problem quite difficult: Newell and Rakow (2007), for example, found that a majority of their participants chose Strategy (c). That is to say, they chose the strategy which exemplifies a “probability matching” – and hence incorrect – response. Strategy (d) is the maximizing strategy, the correct answer in the sense that it will yield more correct predictions than any of the other strategies. It will yield around 40 correct predictions ($4/6 \times 60$), a level that cannot be exceeded given the randomness of the outcomes. Strategy c – the matching strategy – is expected to yield only around $(4/6 \times 4/6 \times 60) + (2/6 \times 2/6 \times 60) = 33$ correct predictions, because whenever a white outcome is predicted there is only a 1 in 3 probability that it will occur. Such predictions are “wasted” because a black outcome is twice as likely. Why are people tempted to match under such transparent conditions?

cor (preta ou branca) irá aparecer. O participante irá receber um dólar por cada resultado que acertar. Assume-se que a pessoa tentará ganhar a maior quantidade de dinheiro possível. Qual estratégia irá garantir o maior número de acertos?

1. Seguir a intuição, trocando quando tiver saído uma ou outra cor muitas vezes.
2. Optar mais pela cor preta na maioria dos resultados, mas, após ter caído muitas vezes preta, chutar branca.
3. Dar palpites de acordo com a proporção de cada cor, nesse caso, (4 de 6 para a cor preta e 2 de 6 para a branca). Isso significa optar por duas vezes mais pretas do que brancas.
4. Chutar sempre a cor preta para todos os 60 lançamentos do dado.
5. Optar mais pela preta do que pela branca, mas, dependendo da quantidade de cores que saia, trocar pela outra.

A descrição do problema diz tudo que é preciso saber para dar palpites racionais. As probabilidades são conhecidas. Admitindo que o dado é honesto e continuará sendo após os 60 lançamentos, conclui-se que as probabilidades serão mantidas.

No entanto, a maioria das pessoas parecem achar esse problema difícil: Newell e Rakow (2007), por exemplo, descobriram que a maioria dos participantes escolhe a estratégia (3). Isso significa que eles escolhem a probabilidade de “matching”. O método que amplia a quantidade de acertos mais do que qualquer outro é o do item (4). Ele garante 40 acertos ($4/6 \times 60$), um nível que não pode ser superado em experimentos aleatórios.

A estratégia de “matching” só resultaria em torno de 33 acertos, ($4/6 \times 4/6 \times 60$) + ($2/6 \times 2/6 \times 60$) = 33, pois o palpite na cor branca só possui uma chance em três de aparecer. Esse “chute” é perdido, uma vez que a cor preta tem o dobro de probabilidade de sair. Por que as pessoas tentam o método de “matching” diante de condições tão óbvias? (NEWELL; SCHULZE, 2017, p. 64)

O cálculo de “matching” desse exemplo é baseado no teorema do produto. Antes de defini-lo é preciso considerar a probabilidade condicional. Segundo Carvalho et al. (2016, p. 133), dados dois eventos A e B , a probabilidade condicional de B dado A é o número

$$P(A \cap B)/P(A).$$

Será utilizada a simbologia:

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

Aplicando para o exemplo apresentado:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{2}{10}}{\frac{5}{10}} = \frac{2}{5}$$

A partir da definição de probabilidade condicional, pode-se enunciar o teorema do produto, segundo Fonseca e Martins (1996), “A probabilidade da ocorrência simultânea de dois eventos, A e B , do mesmo espaço amostral, é igual ao produto da probabilidade de um deles pela probabilidade condicional do outro, dado o primeiro.”

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

Na experiência dos copos, o fato da moeda estar embaixo de um copo verde, não interfere onde estará na próxima dupla. Trata-se do teorema do produto para eventos independentes. Neste caso, sejam os eventos A e B independentes, $P(A|B) = P(A)$ e $P(B|A) = P(B)$ (FONSECA; MARTINS, 1996, p. 18), ou seja, o fato de ter ocorrido A antes, não afeta a probabilidade do evento B ocorrer e vice-versa.

Em geral, as pessoas escolhem o método do item 3 a fim de ampliar a quantidade de acertos, porém, o que aumenta o número de palpites corretos é o item 4. Neste último caso, em 60 lançamentos do dado, tem-se $\frac{4}{6} \cdot 60 = 40$, ou seja, a quantidade de acertos é em torno de 40 deles, mas na estratégia de “matching”, é preciso ocorrer duas situações. Para haver acerto, a cor sorteada e o palpite do participante devem coincidir. Para calcular o número esperado de acertos no “matching”, vamos considerar o espaço amostral formado por pares ordenados em que a primeira coordenada indique a cor da face sorteada e a segunda coordenada o palpite dado pelo participante. Vamos indicar por b_1, b_2, b_3, b_4 , as faces pretas e por w_1, w_2 , as faces brancas do dado. Analogamente, como estamos admitindo que o participante utilizará a estratégia de “matching”, o conjunto de seus possíveis palpites irá refletir o desequilíbrio de ocorrência das duas cores. Vamos denotar por u_1, u_2, u_3, u_4 , palpites na cor preta e por v_1, v_2 , palpites na cor branca.

$$\Omega = \{(f, g), f \in \{b_1, b_2, b_3, b_4, w_1, w_2\}, g \in \{u_1, u_2, u_3, u_4, v_1, v_2\}\}.$$

Os eventos serão subconjuntos de Ω . Para estimar os acertos do participante, vamos

considerar os eventos

$$A_1 = \{\text{uma face preta é sorteada}\} = \{(f, g), f \in \{b_1, b_2, b_3, b_4\}, g \in \{u_1, u_2, u_3, u_4, v_1, v_2\}\},$$

$$A_2 = \{\text{uma face branca é sorteada}\} = \{(f, g), f \in \{w_1, w_2\}, g \in \{u_1, u_2, u_3, u_4, v_1, v_2\}\},$$

$$B_1 = \{\text{um palpite na cor preta}\} = \{(f, g), f \in \{b_1, b_2, b_3, b_4, w_1, w_2\}, g \in \{u_1, u_2, u_3, u_4\}\},$$

$$B_2 = \{\text{um palpite na cor branca}\} = \{(f, g), f \in \{b_1, b_2, b_3, b_4, w_1, w_2\}, g \in \{v_1, v_2\}\}.$$

Observe que a probabilidade de sair a cor preta é

$$P(A_1) = \frac{\#A_1}{\#\Omega} = \frac{24}{36} = \frac{4}{6}.$$

Em 60 lançamentos do dado, tem-se em torno de $\frac{4}{6} \cdot 60 = 40$ lançamentos em que a cor da face sorteada é preta. Note que os eventos são A_i e B_i , $i = 1, 2$ são independentes. Desse modo, a probabilidade de um palpite no preto, dado que a face sorteada foi preta é dada por

$$P(B_1|A_1) = P(B_1) = \frac{\#B_1}{\#\Omega} = \frac{24}{36} = \frac{4}{6}.$$

Assim, o número esperado de acertos no preto em 60 lançamentos é dado por

$$60 \cdot P(A_1 \cap B_1) = 60 \cdot P(A_1) \cdot P(B_1) = \frac{4}{6} \cdot \frac{4}{6} \cdot 60 = \frac{80}{3}.$$

Analogamente, o número esperado de acertos na cor branca em 60 lançamentos é dado por

$$60 \cdot P(A_2 \cap B_2) = \frac{2}{6} \cdot \frac{2}{6} \cdot 60 = \frac{20}{3}$$

Assim a **média de acertos** na estratégia de “matching” é dada por

$$\frac{80}{3} + \frac{20}{3} = \frac{100}{3} \cong 33.$$

Observe que este valor é inferior ao obtido na estratégia do item 4.

5.2 Exemplo das lâmpadas

Mlodinow (2009) trata da estratégia de “matching” ao descrever um experimento que utiliza escolha binária que consistiu em acender uma lâmpada vermelha ou uma verde de forma aleatória. O participante devia observar uma sequência de acendimentos e depois

tentar adivinhar a cor que seria acesa. A partir de um determinado número de tentativas, mesmo que seja um processo randômico, a pessoa que participa da experiência pode entender que existe um suposto padrão para a sequência de cores e descobri-lo significaria 100% de acerto nos demais palpites. O participante pode entender que deveria usar a estratégia de “matching”.

Considere que a máquina das lâmpadas de fato gere sequências aleatórias. Se a luz verde apareceu em 75% dos resultados anteriores, pode-se pensar em duas estratégias para acertar qual será a próxima a acender. A primeira seria sempre responder verde, assim, haveria uma média de 75% de acertos. Esse método é usado por alguns animais. Outra opção é o “matching”, significa descobrir um suposto padrão para ampliar a quantidade de palpites certos. Os seres humanos costumam optar pelo “matching”, mas no entanto, isso resultaria apenas cerca de 6 acertos de 10 tentativas. (MLODINOW, 2009, p. 13).

Admitindo que o experimento é aleatório, o fato da luz verde ter aparecido em 75% das vezes na fase de observação, não define que essa proporção será seguida nos próximos acendimentos. A chance da verde acender é sempre 50% para cada tentativa, pois foi dito que é um processo randômico e não se determinou que era um espaço amostral não-equiprovável. Não se definiu que uma cor tem maior probabilidade do que outra. Sobre o desempenho do uso do “matching” nessa experiência:

Os experimentos que investigavam a probabilidade de “matching” utilizam processos que geram resultados aleatórios (exemplo das luzes verde e vermelha que são acesas). Isso significa que a probabilidade de aparecer a cor verde é sempre a mesma independente do que ocorreu anteriormente. Dessa forma, é fácil mostrar que o “matching” é inferior à estratégia de maximização¹⁹. (KOEHLER; JAMES, 2014, p. 104)

Considere uma pessoa que deseja aumentar ao máximo seus acertos na experiência das lâmpadas. Supondo que já houve um determinado número de acendimentos e que o participante já percebeu a probabilidade de sair cada cor. Verificou-se que o verde saiu em 75%. A partir disso, considere uma sequência de n acendimentos. Admitindo que o participante irá adotar a estratégia de “matching”, então ele irá escolher verde $0,75 \times n$ vezes e vermelho $0,25 \times n$ vezes. Em cada um desses casos, a chance de ele estar certo é respectivamente dada por 0,75 e 0,25. Desse modo é provável que ele tenha

$$(0,75 \times n) \times (0,75) + (0,25 \times n) \times 0,25 = 0,5625n + 0,0625n = 0,625n$$

¹⁹ In the experiments investigating probability matching that are of interest here, the outcomes being predicted (e.g., green vs. red light) are determined by a random process that is serially independent and stationary, which means that probability of, say, the green light illuminating is the same on every trial, regardless of what occurred on the previous trial or how many trials have elapsed. Under such circumstances, it is easy to show that, if one’s goal is to maximize the number of correct predictions, probability matching is inferior to an alternative strategy in which the higher probability outcome is predicted on every trial.

acertos. São 62,5% de acertos em n acendimentos. Isto é, acerto em média em $(0,75 \times 0,75) + (0,25 \times 0,25)$ das vezes, ou seja, em torno de 60% das vezes.

Vale lembrar que o fato do experimento ser aleatório, não garante que as proporções apresentadas nos primeiros acendimentos serão mantidas nos próximos. Os cálculos apresentados referem-se a uma média e foram feitos admitindo que o participante de alguma forma capturou as probabilidades subjacentes às cores verde e vermelho. Mais precisamente, a quantidade esperada de acertos do participante em n acendimentos seriam dados por

$$(0,75 \times n) \times p + (0,25 \times n) \times (1 - p),$$

onde p é a probabilidade por ele atribuída ao verde. O que é importante observar é que na estratégia de “matching” temos em geral $0 < p < 1$. Portanto,

$$(0,75 \times n) \times p + (0,25 \times n) \times (1 - p) < (0,75 \times n) \times p + (0,75 \times n) \times (1 - p) = 0,75 \times n.$$

Isto é, a estratégia de “matching” nunca supera em desempenho a de maximização – aquela em que sempre apostamos na cor que é percebida como mais provável.

De certa forma, um tipo de estratégia de “matching” já era perseguido por Pitágoras. Muitos matemáticos buscaram enquadrar o mundo nas leis matemáticas de forma a encontrar padrões. Boyer e Merzbach (2012) afirma que os pitagóricos veneravam a Matemática, a tomavam não somente como ciência, mas como religião. Quando Pitágoras diz que “tudo é número” se refere a esse fato, mas também à tentativa de procurar padrões a sua volta.

Até hoje as pessoas buscam leis para prever situações. Isso se dá até mesmo em sistemas aleatórios, como os números que serão sorteados na Mega Sena ou os resultados das ações na Bolsa de Valores²⁰. Talvez não entendam o conceito da aleatoriedade ou simplesmente acreditam que exista um padrão que se repete após centenas de resultados.

Aparentemente, não é apenas o “matching” responsável por isso. Grandes descobertas foram feitas com o objetivo de se buscar um padrão para um determinado experimento. Foi assim que a Física foi criada.

²⁰ <http://www.infomoney.com.br/acoes/noticia/416800/numeros-fibonacci-forma-simples-para-tentar-prever-reverso-es-mercado>

6 LEI DE BENFORD

O aumento da canalhice é o resultado da má distribuição de renda.

Millôr Fernandes

O livro de Mlodinow (2009) não preenche as lacunas sobre como conseguiu encontrar muitas das soluções que apresenta para seus exemplos. Por outro lado, há uma gama de casos interessantes associados à probabilidade entre outros assuntos. Este capítulo se propõe a apresentar a Lei de Benford. Uma das ferramentas nas auditoriais de empresas.

6.1 Apresentação da Lei

Em 1881, o astrônomo Simon Newcomb observou que as páginas das tabelas de logaritmos correspondentes aos primeiros dígitos 1 e 2 eram mais gastas do que as dos dígitos 8 e 9. Ele descobriu que alguns algarismos aparecem com mais frequência como primeiros dígitos significativos do que outros em alguns conjuntos aleatórios. O cientista enunciou a lei do Primeiro Dígito Significativo, também conhecida como Lei de Benford. Muitos anos depois, em 1938, a lei foi redescoberta por Frank Benford, mas nenhum deles a provou, somente Theodore Hill o fez em 1995. Rousseau (2010)

A Lei de Benford diz que em um conjunto de números positivos aleatórios a frequência dos primeiros dígitos significativos de números é desigual. O primeiro dígito significativo de um número positivo é o dígito não nulo mais à esquerda em sua representação decimal. Para descrever a lei, é conveniente escrever todo número em função de sua mantissa. Todo número real positivo x , pode ser escrito como

$$x = y \cdot 10^n, \quad y \in [1, 10), \quad n \in \mathbb{Z},$$

y é a mantissa de x e sua parte inteira nos dá o primeiro dígito significativo de x . A partir desta representação, Hill (1995) descreveu os conjuntos $\{D_1 = i\}$ dos números reais positivos cujo primeiro dígito significativo era igual a i da seguinte maneira

$$\{D_1 = i\} = \cup_{n=-\infty}^{+\infty} [i, i + 1) \cdot 10^n, \quad i = 1, 2, \dots, 9.$$

Esta descrição permitiu que Hill (1995) definisse e caracterizasse a σ -álgebra em que a função probabilidade associada a lei de Benford estivesse bem definida.

Seja $i \in \{1, 2, \dots, 9\}$. A Lei de Benford nos diz que a frequência de i será dada

Tabela 2 - Frequência dos primeiros dígitos significativos

i	1	2	3	4	5	6	7	8	9
$B(i)$	0,3010	0,1761	0,1249	0,0969	0,0792	0,0669	0,0580	0,0511	0,0458

Fonte: Rousseau (2010)

por

$$\log_{10} \left(1 + \frac{1}{i} \right).$$

Segundo a Lei de Benford, o algarismo 1 aparece como primeiro dígito significativo cerca de 30%, o 2, aproximadamente em 18% dos casos e assim por diante até o algarismo 9, em torno de 5% Mlodinow (2009, p. 92). Esses valores podem ser calculados assim:

- **Para** $i = 1$

$$\log_{10} \left(1 + \frac{1}{i} \right) = \log_{10} \left(1 + \frac{1}{1} \right) = \log_{10} 2 \cong 0,3010,$$

aproximadamente 30%.

- **Para** $i = 2$

$$\log_{10} \left(1 + \frac{1}{i} \right) = \log_{10} \left(1 + \frac{1}{2} \right) = \log_{10} \frac{3}{2} = \log_{10} 3 - \log_{10} 2$$

$$\cong 0,4771 - 0,3010 = 0,1761,$$

aproximadamente 18%.

- **Para** $i = 9$

$$\log_{10} \left(1 + \frac{1}{i} \right) = \log_{10} \left(1 + \frac{1}{9} \right) = \log_{10} \frac{10}{9} = \log_{10} 10 - \log_{10} 9$$

$$= 1 - \log_{10} 3^2 = 1 - 2 \cdot \log_{10} 3 \cong 1 - 2 \cdot 0,4771 = 0,0458,$$

aproximadamente 5%

A frequência de cada algarismo pode ser verificada na Tabela2:

Tabela 3 - Frequência dos primeiros dígitos significativos nos balancetes

DÍGITO	BENFORD %	FREQUÊNCIA RELATIVA
1	30,1 %	30,74 %
7	5,8 %	5,92 %
8	5,1 %	5,33 %

Fonte: Gava e Vitiello (2007)

6.2 Aplicações da Lei de Benford

Uma aplicação para a Lei de Benford é descobrir fraudes financeiras. Essa lei foi utilizada no caso de Kevin Lawrence. Ele foi acusado de desviar o dinheiro de seus investidores. O contador forense, Darrell Dorrell, compilou cerca de 70 mil dados bancários referentes a transferências e cheques emitidos pelo réu e utilizou a Lei de Benford Mlodinow (2009, p. 92). Os dados não passaram no teste, e a investigação prosseguiu e desmascarou o esquema fraudulento de Lawrence e seus sócios.

Para explicar como é feita a primeira análise, pode-se utilizar o caso de Wayne James Nelson, gerente do escritório do Tesoureiro do Estado do Arizona. A maioria dos cheques que ele passou seguiu o oposto da Lei de Benford, o primeiro algarismo de cada cheque começava com 7, 8 ou 9 que são justamente os algarismos que menos aparecem na distribuição. O gerente emitiu os cheques abaixo de 100 mil dólares, possivelmente para tentar encobrir a fraude ou por serem valores que não chamariam a atenção do governo, mas ao criar os valores que deveriam ser aleatórios, criou um padrão²¹.

Outro exemplo de aplicação da lei foi descrito por Gava e Vitiello (2007). O estudo utilizou cerca de 2 mil balancetes de 44 empresas públicas brasileiras no período de 1986 a 2002. O objetivo era verificar se houve indício de fraude e associar a mesma ao período de alta ou baixa inflação Gava e Vitiello (2007). Os autores registraram a frequência dos algarismos de 1 a 9 em todos os 71.799 balancetes. Alguns dos dados encontrados como primeiro dígito estão na Tabela 3 Gava e Vitiello (2007):

Para verificar se a variação em relação à frequência esperada era aceitável, os autores utilizaram alguns testes estatísticos.

Em nosso estudo, a Lei de Benford foi aplicada ao balancete de abril de 2017 do governo do Estado do Rio de Janeiro²² referente às despesas liquidadas em programas. Foram, analisados 105 valores, apresentados na Tabela:

²¹ <http://www.journalofaccountancy.com/issues/1999/may/nigrini.html>

²² <http://www.fazenda.rj.gov.br>

Tabela 4 - Frequência dos primeiros dígitos significativos no balancete de abril de 2017

DÍGITO	BENFORD %	FREQUÊNCIA	F. RELATIVA
1	30,1 %	36	34,29 %
2	17,6 %	22	20,95 %
3	12,5 %	15	14,29 %
4	9,7 %	6	5,71 %
5	7,9 %	5	4,76 %
6	6,7 %	5	4,76 %
7	5,8 %	5	4,76 %
8	5,1 %	3	2,86 %
9	4,6 %	8	7,62 %

Fonte: A autora, 2017.

Demos um tratamento análogo ao proposto por Gava e Vitiello (2007) à discrepância e, aparentemente, não há indício de fraude, porém, seria necessária uma perícia técnica a fim de determinar se de fato não houve.

Embora seja uma grande descoberta, o fato de um grupo de números não seguir a distribuição dessa lei, não quer dizer que houve fraude, apenas demonstra a necessidade de haver uma averiguação mais criteriosa para a amostra Rousseau (2010, p. 15)

7 O ANDAR DO BÊBADO

A imaginação é mais importante do que o conhecimento.

Albert Einstein

Este capítulo apresenta um experimento de Mlodinow (2009) aparentemente simples. Desejou-se mostrar a relação do andar do bêbado com o que é aleatório, um caminhar sem rumo, a sorte o define. O foco deste capítulo é associar o rigor matemático da enunciação de teoremas a situações práticas. Isso foi uma preocupação em toda a dissertação, mas aqui houve um cuidado maior. O objetivo é que o leitor não tenha lacunas no entendimento.

Na Matemática, há uma ideia recorrente de utilizar estruturas simples para resolver problemas complexos. Esta estratégia será usada para discutir o exemplo dos filmes do George Lucas.

Imagine que George Lucas produza um novo filme da série *Guerra nas estrelas* e, em um teste de mercado, resolva realizar um experimento insano. Ele lança dois filmes idênticos, mas com títulos diferentes: *Guerra nas estrelas: episódio A* e *Guerra nas estrelas: episódio B*. Cada filme tem sua própria campanha publicitária e seu programa de distribuição, idênticos em cada detalhe, a não ser pelo fato de que os trailers e propagandas de um filme dizem *Episódio A* e os do outro dizem *Episódio B*. Agora fazemos uma competição entre os dois. Qual filme fará mais sucesso? Suponha que observemos os primeiros 20 mil espectadores e registremos o filme que resolveram assistir (ignorando os fãs obstinados que assistirão a ambos e insistirão na ideia de que existem diferenças sutis, mas significativas, entre os dois). (MLODINOW, 2009, p. 21)

Como os filmes e suas campanhas publicitárias são idênticos, podemos assumir a probabilidade de 50% de cada um ser escolhido. Assim, pode-se utilizar o modelo matemático no qual a escolha de cada espectador pode ser simulada pelo lançamento de uma moeda. É intuitivo acreditar que cada filme permanecerá na liderança durante a metade do tempo. No entanto, é 88 vezes mais provável que um dos filmes fique na liderança do que a mesma passe de um para outro continuamente (MLODINOW, 2009, p. 21).

Feller (1957) descreve uma disputa entre duas pessoas. O jogo se baseia em lançar uma moeda. Uma simples partida de cara ou coroa. O exemplo de Mlodinow (2009) é inspirado no de Feller (1957) e descreve uma situação similar. Apesar de serem dois filmes, ambos são iguais. Assim, sua decisão pode ser simulada através do lançamento de uma moeda. Pode-se associar cada face da moeda com a preferência de cada espectador por um dos filmes. Imagina-se que há uma fila de pessoas na porta do cinema. Uma moeda é lançada a fim de simular qual filme o espectador verá.

O título deste capítulo é também o do livro de Mlodinow (2009). Refere-se a uma caminhada aleatória. Em relação aos filmes, a randomicidade rege as decisões dos espectadores. O assunto aqui tratado foi o mais desafiador nesta dissertação. Será utilizado o artigo de Border (2017) por apresentar uma linguagem de mais fácil compreensão. Vamos combinar as abordagens e notações de Border (2017) e Feller (1968).

Ao invés de utilizar as faces da moeda conforme a abordagem de Feller (1957), os votos em cada filme serão associados a dois números. Se o escolhido for o A , o voto será contado como $+1$. Se for o B , -1 . Cada espectador t , sendo $t = 1, 2, \dots$ tem seu voto representado por X_t . Assim, pode-se realizar a soma desses votos:

$$S_t = X_1 + X_2 + \dots + X_t$$

Se essa soma for um número positivo, tem-se que o filme A teve mais votos. Se for um número negativo, B foi o ganhador da bilheteria. No caso da soma ser nula, houve empate. Neste caso, Feller (1957, p. 67) diz “convencionaremos afirmar que Peter lidera no n -ésimo lançamento se $S_n > 0$ ou $S_n = 0$, mas $S_{n-1} > 0$ (isto é, em caso de empate será declarado líder o competidor liderou na penúltima jogada).”²³ Utilizando a analogia da disputa dos filmes, em caso de empate no voto t , o *Episódio A* ainda estará na liderança se o voto anterior foi para A . Caso contrário, B será o líder.

7.1 O que é um passeio aleatório?

Quando uma função está definida em um espaço amostral, ela é chamada de variável aleatória. Uma variável de Rademacher é uma variável aleatória que assume apenas dois estados, cada um deles com probabilidade $1/2$. Isto é, $X : \Omega \rightarrow \{-1, 1\}$ tal que

$$P(X = 1) = P(X = -1) = \frac{1}{2}.$$

Uma sequência de Rademacher é uma sequência (X_t) de variáveis de Rademacher independentes. O índice t indica um instante no tempo, que tal como Feller (1968), chamaremos de época. O conjunto de épocas coincide com o de inteiros não negativos:

$$\mathbb{Z}_+ = \{0, 1, 2, \dots\}.$$

A época 0 é o instante que antecede o primeiro lançamento da moeda (ou voto).

²³ We shall therefore agree to say that as the n th trial Peter leads if either $S_t > 0$ or $S_t = 0$, but $S_{t-1} > 0$, (i.e., in case of a tie that player leads who led at the preceding trial).

Para cada t , definimos a soma acumulada

$$S_t = X_1 + \cdots + X_t.$$

Convencionamos que $S_0 = 0$. A sequência $S_0, S_1, S_2, \dots, S_t, \dots$ é chamada de passeio aleatório simples no conjunto dos números inteiros.

7.2 Caminhos

Feller (1957) associa o voto do t -ésimo espectador e a soma acumulada S_t dos votos até o momento ao ponto (t, S_t) . A preferência de cada espectador define uma sequência de pontos. Isto é, ele identifica o passeio aleatório (S_t) a um conjunto de pontos do plano (t, S_t) . Seguir estes pontos significa percorrer um caminho.

Definição 3. *Sejam x, y inteiros, $x > 0$. Um caminho $s = (s_0, s_1, s_2, s_3, \dots, s_x)$ da origem ao ponto (x, y) é a linha poligonal cujos vértices têm abscissas $0, 1, 2, \dots, x$ e ordenadas $s_0, s_1, s_2, \dots, s_x$ satisfazendo:*

$$s_j - s_{j-1} = \pm 1, \quad s_0 = 0, \quad s_x = y. \quad (4)$$

Tal como em Border (2017) e Feller (1968), quando nos referirmos ao caminho, usaremos o s_j para indicar a soma acumulada S_j .

Em conformidade com Feller (1957), no contexto da disputa de episódios a abcissa $x = p + m$ do ponto final de um caminho representa o total de votos, admitindo que cada pessoa vota em algum dos dois episódios e que o episódio A tenha recebido p votos e o episódio B , m votos. Quando uma pessoa votar no episódio A será somada uma unidade positiva ao montante. Se o voto for para o episódio B , será subtraída uma unidade. Se a soma parcial $s_k = X_1 + X_2 + \cdots + X_k$ é positiva, o resultado parcial até o k -ésimo voto indica a vitória de A . Quando ela é negativa, indica a vitória de B . Ao final da votação, temos $s_x = p - m$ é o saldo da soma de todos os votos. Se s_x for positivo, A vence a disputa. Para a soma igual a zero, há empate. Se a soma for negativa, B vence. Em (4), temos

$$s_j - s_{j-1} = X_j = \pm 1, \quad \text{para } j = 1, 2, \dots, x.$$

A soma atual menos a anterior resulta em um determinado voto X_j durante a apuração. Se for para o episódio A , $X_j = +1$, para o B , $X_j = -1$. Ou seja, sempre é possível verificar para qual episódio foi o j -ésimo voto X_j , uma vez que $s_{j-1} + X_j = s_j$. Existe uma bijeção entre a sequência de votos e as somas acumuladas.

7.2.1 Pontos “alcançáveis”

Quando $s_t = k$, para $k \in \mathbb{Z}$, dizemos que o caminho s visita k na época t . Se existe um caminho s tal que $s_t = k$, dizemos que o caminho s atinge o ponto (t, k) ou que o ponto (t, k) é alcançável a partir da origem.

Quais pontos do plano pertencentes a $\mathbb{Z}_+ \times \mathbb{Z}$ são alcançáveis? Voltemos à disputa de episódios para analisar esta questão. Considere que t pessoas assistirão ao *Episódio A* ou ao *Episódio B*. Assim, a soma dos votos de cada episódio corresponde ao total de votos t . Sendo p o número de votos do episódio *A* e m , do *B*, então $t = p + m$. Assumindo que para cada voto em *A* foi registrado como $+1$ e para o *B*, como -1 , obtemos a soma acumulada de votos $S_t = p - m = k$. Se associamos a esta votação o caminho $s = (s_0, s_1, \dots, s_t)$, vemos que o ponto (t, k) é alcançável. Note que todo caminho que atende à Definição 3 pode ser associado a uma votação entre os dois episódios.

Proposição 2. *Para que o ponto $(t, k) \in \mathbb{Z}_+ \times \mathbb{Z}$ seja alcançável, devem existir inteiros não negativos p e m , tais que*

$$p + m = t \quad e \quad p - m = k. \quad (5)$$

Corolário 1. *Se $(t, k) \in \mathbb{Z}_+ \times \mathbb{Z}$ é alcançável, então as coordenadas t e k têm a mesma paridade. Além disso, $t \geq |k|$*

Demonstração. Seja $(t, k) \in \mathbb{Z}_+ \times \mathbb{Z}$ alcançável. Como $t \in \mathbb{Z}_+$, tem-se duas possibilidades para t :

- i) t é par. Isto é, $t = 2\nu$ para algum $\nu \in \mathbb{Z}_+$. Sabemos que existe uma solução inteira não negativa de (5):

$$\begin{cases} 2\nu = p + m \\ k = p - m \end{cases}.$$

Logo

$$k = 2(\nu + p) = 2(\nu - m).$$

Portanto, se t é par, k também deve ser par para que (5) tenha solução inteira não negativa. Neste caso

$$p = \frac{t + k}{2} \quad e \quad m = \frac{t - k}{2}.$$

- ii) t é ímpar. Isto é, $t = 2\nu + 1$ para algum $\nu \in \mathbb{Z}_+$. Sabemos que existe uma solução

inteira não negativa de (5):

$$\begin{cases} 2\nu + 1 = p + m \\ k = p - m \end{cases}.$$

Logo

$$k = 2(\nu + p) + 1 = 2(\nu - m) + 1.$$

Portanto, se t é ímpar, k também deve ser ímpar para que (5) tenha solução inteira não negativa. Neste caso

$$p = \frac{t+k}{2} \quad \text{e} \quad m = \frac{t-k}{2}.$$

Como $t = p + m$, temos $p, m \leq t$. Como $k = p - m$, temos

$$-t = 0 - t \leq k \leq t - 0 = t.$$

□

O ponto $(5, 4)$ não é alcançável. De fato, para $(t, k) = (5, 4)$, o sistema (5) não tem solução (p, m) em $\mathbb{Z}_+ \times \mathbb{Z}_+$.

De maneira análoga a Feller (1957), Border (2017) define a quantidade de caminhos possíveis até um ponto alcançável.

Definição 4. O número de caminhos até um ponto $(t, k) \in \mathbb{Z}_+ \times \mathbb{Z}$ é denotado por $N_{t,k}$. Se (t, k) não for alcançável, então $N_{t,k} = 0$.

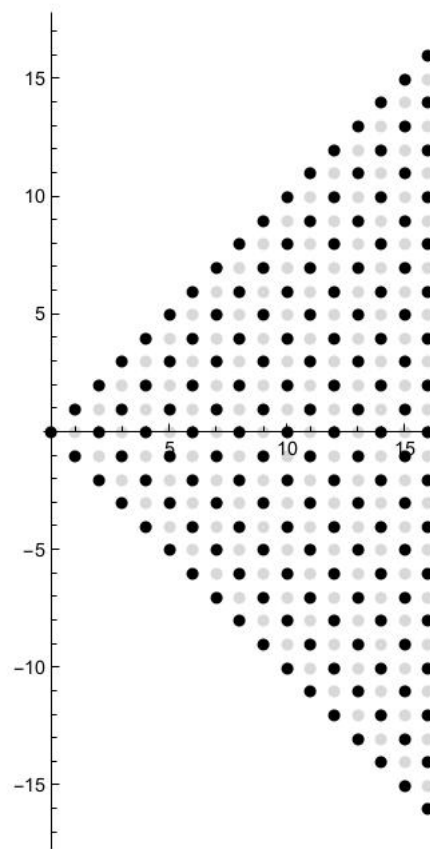
Na Proposição 3, Border (2017) nos mostra como determinar o número de caminhos até um ponto alcançável (t, k) .

Proposição 3. Se (t, k) é um ponto alcançável, então

$$N_{t,k} = \binom{t}{\frac{t+k}{2}} = \binom{t}{\frac{t-k}{2}}. \quad (6)$$

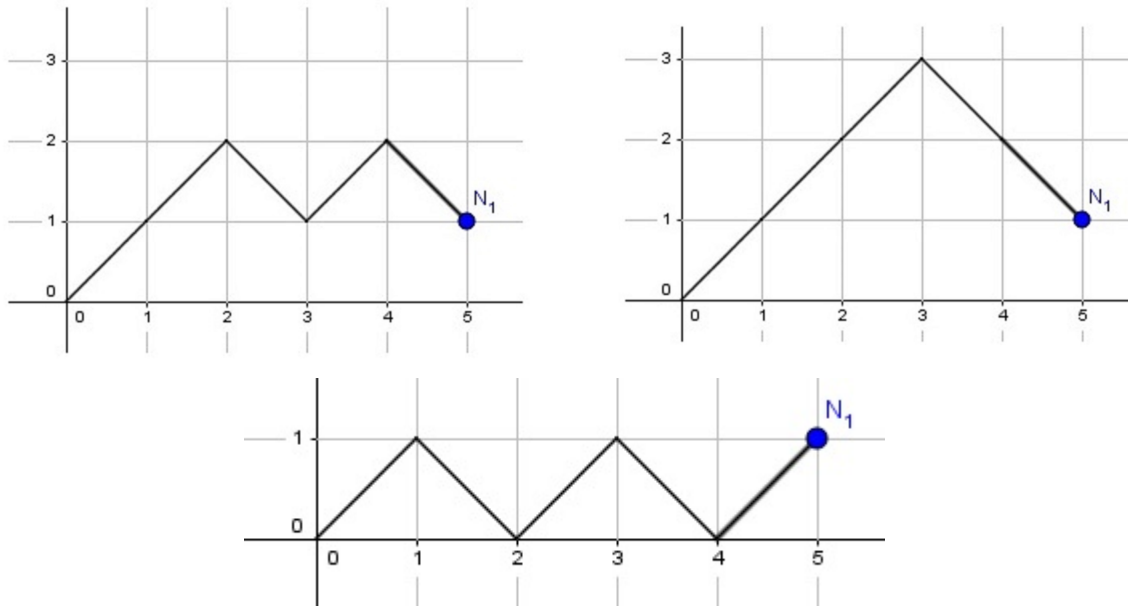
Prova. Pela Proposição 2, há inteiros não negativos p e m tais que (5) é satisfeita. Utilizando a analogia dos filmes, p é a quantidade de votos no filme A . É o número de vezes que aparece o $+1$; m é a quantidade de vezes que aparece -1 . Na soma dos votos, $+1$ e -1 podem aparecer em qualquer ordem, assim, o total de organizações possíveis é uma

Figura 6 - Os pontos alcançáveis estão em preto.



Fonte: Border (2017).

Figura 7 - Alguns caminhos até (5, 1)



Fonte: A autora, 2017.

permutação com repetição de $t = p + m$. Dessa forma:

$$N_{t,k} = \binom{p+m}{p} = \frac{(p+m)!}{p!m!} = \binom{p+m}{m} = \binom{t}{\frac{t+k}{2}} = \binom{t}{\frac{t-k}{2}}.$$

□

Então existem exatamente $N_{t,k}$ caminhos diferentes da origem ao ponto (t, k) .

A Figura 7 ilustra alguns caminhos da origem até o ponto alcançável $N_1 = (5, 1)$. Há dez caminhos possíveis ligando a origem ao ponto N_1 . Pode-se pensar em uma determinada sequência de passos e depois utilizar permutação com repetição. Pode-se traçar qualquer poligonal da origem ao ponto escolhido. Usando a Figura 7, tem-se que há três subidas e duas descidas. Associando aos episódios, A foi escolhido três vezes e B duas. São cinco votos, assim a permutação $P_5^{3,2} = 10$ caminhos.

Dizemos que um caminho da origem até (t, k) tem comprimento t . Feller (1957) explica que se a ordenada do ponto final (t, k) não for determinada e tivermos fixado apenas o número de espectadores t , tem-se 2^t caminhos distintos de comprimento t que partem da origem. De fato, para cada espectador há duas possibilidades de escolha. Assim o voto de cada um deles pode ser associado ao lançamento de uma moeda. Assim, para cada um dos t votos, tem-se duas opções, resultando em 2^t caminhos ligando a origem a um ponto (t, k) para algum $k \in \mathbb{Z}$. Se $k > 0$, o episódio A vence, se $k < 0$, B vence e se $k = 0$, houve empate. Um ponto qualquer (t, k) pode ser ligado à origem por um caminho, se t e k forem da forma (6). Sempre se pode traçar poligonais de um ponto qualquer do

plano à origem, mas aqui há condições que devem ser satisfeitas (4). $p, m \in \mathbb{N}$ pois representam a quantidade de votos em cada um dos episódios. Além disso, na definição, $t \in \mathbb{N}$, de fato, pois representa a quantidade total de votos. Pela definição de t e k , na situação em que episódio B fosse o mais votado, k assumiria valor negativo.

Em um caminho, a ordenada s_t também admite uma outra interpretação: ele representa a posição de uma partícula que se move ao longo do eixo vertical partindo da origem; a cada época ela se move uma unidade para cima se o episódio A recebeu um voto, ou para baixo, se o voto foi para o episódio B . Na época n , a posição da partícula é o ponto s_n do eixo vertical. O caminho $s = (s_0, s_1, \dots, s_n)$ representa o diagrama espaço-tempo do passeio aleatório (S_t) , o eixo horizontal faz o papel do eixo temporal. (FELLER, 1957, p. 74).

Assumiremos que todos os caminhos são igualmente prováveis. A descrição dos eventos associados a passeios aleatórios é feita basicamente através da imposição de condições sobre as somas S_j . O evento “na época n , a partícula está na posição r ” será denotado por $\{S_n = r\}$ e sua probabilidade $P(S_n = r)$, por $p_{n,r}$.

Border (2017) denota a probabilidade de se alcançar um determinado ponto (t, k) por $p_{t,k} = P(S_t = k)$.

Corolário 2. *Se (t, k) é um ponto alcançável, então*

$$p_{t,k} = \binom{t}{\frac{t+k}{2}} \cdot 2^{-t}. \quad (7)$$

Os caminhos que foram contados em (6) sempre partem da origem $(0, 0)$ do sistema de coordenadas. A partir disso, define-se um recurso que será utilizado algumas vezes no futuro. Considere a nova origem no ponto (t_0, k_0) , tem-se:

Corolário 3. *Se (t_1, k_1) é um ponto alcançável a partir de (t_0, k_0) , o número de caminhos que ligam esses pontos é*

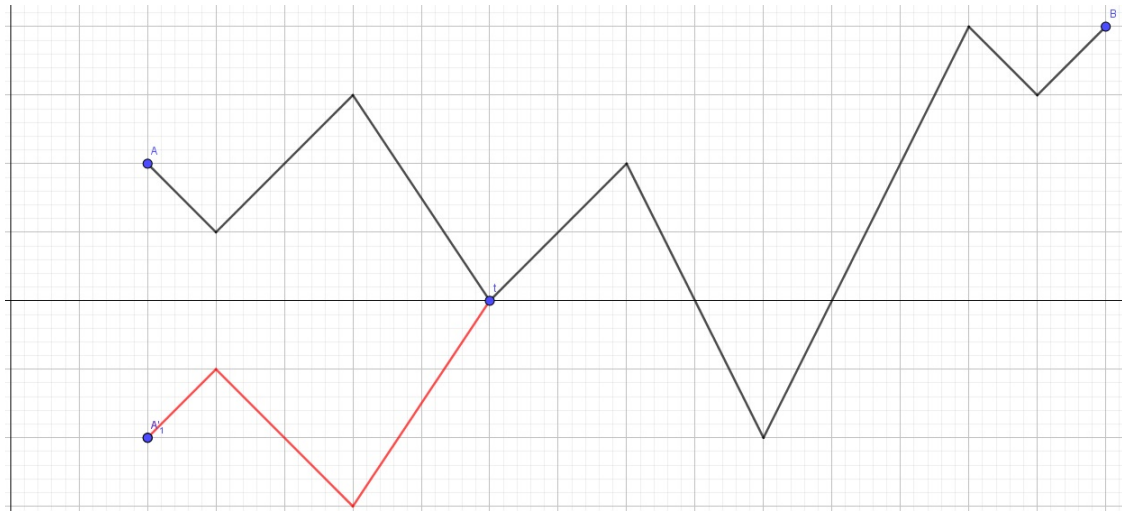
$$N_{t_1-t_0, k_1-k_0} \quad (8)$$

Neste caso, (t_0, k_0) se tornou a nova origem, e utilizamos a fórmula (6).

7.3 Princípio da Reflexão

Feller (1957 apud BORDER, 2017) usa os caminhos como ferramenta central para análise de passeios aleatórios e de suas aplicações. Assume que é possível manipular trechos deles. Cortar, transladar, rotacionar são algumas das técnicas utilizadas. Assim, busca uma bijeção entre os caminhos que deseja contar com outros. A ideia é facilitar

Figura 8 - Exemplo do Princípio de Reflexão



Fonte: A autora, 2017.

a contagem para definir a probabilidade desejada. O primeiro método utilizado foi o Princípio da Reflexão.

Lema 1 (Princípio da Reflexão). *Sejam (t_1, k_1) um ponto alcançável a partir de (t_0, k_0) e ambos do mesmo lado do eixo x . Então há uma correspondência biunívoca entre o conjunto dos caminhos de (t_0, k_0) a (t_1, k_1) que tocam ou cruzam o eixo x com o conjunto de todos os caminhos que vão de $(t_0, -k_0)$ a (t_1, k_1) .*

Prova. Sem perda de generalidade, assuma que as coordenadas de $A = (t_0, k_0)$ e $B = (t_1, k_1)$ são números inteiros e positivos, ou seja, $t_1 > t_0 \geq 0, k_0, k_1 > 0$. Por reflexão de A em relação ao eixo x , é encontrado o ponto $A' = (t_0, -k_0)$ (Figura 8). Um caminho de A para B é definido como na Definição 3, com A sendo a origem.

Considere um caminho $\{s_{t_0} = k_0, s_{t_0+1}, \dots, s_{t_1-t_0} = k_1\}$ de A para B tendo um ou mais vértices no eixo x . Seja t^* a abscissa do primeiro ponto em que o caminho toca o eixo x (Figura 8); ou seja, escolha t^* tal que $s_a > 0, \dots, s_{t^*-1} > 0, s_{t^*} = 0$. Então $\{-s_{t_0}, -s_{t_0+1}, \dots, -s_{t^*-1}, s_{t^*} = 0, s_{t^*+1}, s_{t^*+2}, \dots, s_{t_1-t_0}\}$ é um caminho ligando A' a B e tendo $T = (t^*, 0)$ como o primeiro vértice do eixo x . As seções AT e $A'T$ sendo reflexões uma da outra, garantem a existência uma correspondência um a um entre os caminhos A' para B e os caminhos de A para B . \square

Para visualizar o Princípio da Reflexão, considere na Figura 8, os pontos $A = (t_0, k_0)$, $B = (t_1, k_1)$ e $A_1 = (t_0, -k_0)$. O caminho vermelho é a reflexão do trecho que encontra pela primeira vez o eixo x no ponto t^* . Com este procedimento, é possível estabelecer uma correspondência biunívoca entre os caminhos de A até B que tocam ou cruzam o eixo e os caminhos de A' até B .

7.4 Apuração parcial da votação

O entendimento do comportamento da disputa de episódios será feito basicamente através da imposição de condições sobre as somas S_j . O evento “na época n , a apuração parcial dos votos é igual a r ” será denotado por $\{S_n = r\}$ e sua probabilidade $P(S_n = r)$, por $p_{n,r}$. Veremos nesta seção algumas situações especiais que podem acontecer durante a votação e que serão úteis na análise da disputa de episódios.

7.4.1 Contando caminhos

O primeiro resultado desta seção é consequência do Princípio da Reflexão (Lema 1). Apresentaremos um detalhamento da prova feita por Border (2017).

Teorema 1 (Teorema da Eleição). *Se $k > 0$, então há exatamente*

$$\frac{k}{n} \cdot N_{n,k}$$

caminhos da origem ao ponto (n, k) que satisfazem $s_t > 0$, $t = 1, 2, \dots, n$.

Prova. Se $s_t > 0$ para todo t , $t = 1, 2, \dots, n$, então $s_1 = 1$, ou seja, o primeiro voto não foi para o episódio B , caso em que teríamos $s_1 = -1$, uma vez que $s_t > 0$, $\forall t$.

Como todo caminho considerado tem $s_1 = 1$, para contar todos os possíveis caminhos até (n, k) , basta tomar o ponto $(1, 1)$ como ponto de partida. Pelo Corolário 3, o número total de caminhos de $(1, 1)$ para (n, k) é $N_{n-1, k-1}$. No entanto, nem todos os caminhos que partem de $(1, 1)$ nos interessam. Alguns destes caminhos tocam o eixo x , assim, não satisfazem $s_t > 0$. Quantos destes caminhos tocam o eixo x ? Pelo Princípio da Reflexão (Lema 1) essa quantidade coincide com a dos caminhos que ligam $(1, -1)$ a (n, k) . O Corolário 3 aplicado ao ponto $(1, -1)$ como ponto de partida nos dá $N_{n-1, k+1}$. Vale lembrar que dentre os caminhos contados em $N_{n-1, k+1}$, também considera caminhos que cruzam o eixo x . O objetivo é contar quantos tocam, assim, se cruzou é porque tocou antes.

Assim, o número de caminhos de $(1, 1)$ a (n, k) que não tocam o eixo x é:

$$N_{n-1, k-1} - N_{n-1, k+1}$$

Considere p e m como em (5). Isto é,

$$p = \frac{n+k}{2} \quad \text{e} \quad m = \frac{n-k}{2}.$$

Então²⁴ $p + m = n$, $p - m = k$, $n + k = 2p$. Assim:

$$N_{n-1,k-1} - N_{n-1,k+1} = \binom{n-1}{\frac{n-1+k-1}{2}} - \binom{n-1}{\frac{n-1+k+1}{2}} = \binom{n-1}{\frac{n+k-2}{2}} - \binom{n-1}{\frac{n+k}{2}}$$

Substituindo $n + k = 2p$ e $n = p + m$, tem-se:

$$\begin{aligned} &= \binom{m+p-1}{p-1} - \binom{m+p-1}{p} \\ &= \frac{(m+p-1)!}{(p-1)!(m+p-1-(p-1))!} - \frac{(m+p-1)!}{p!(m+p-1-p)!} \\ &= \frac{(m+p-1)!}{(p-1)!m!} \cdot \frac{p}{p} - \frac{(m+p-1)!}{p!(m-1)!} \cdot \frac{m}{m} \\ &= \frac{(p-m)(p+m-1)!}{p!m!} \cdot \frac{p+m}{p+m} \\ &= \frac{p-m}{p+m} \cdot \frac{(p+m)!}{p!m!} \end{aligned}$$

Substituindo $p + m = n$, $p = \frac{n+k}{2}$ e $m = \frac{n-k}{2}$, tem-se:

$$= \frac{k}{n} \cdot \frac{n!}{(n+k/2)! \cdot (n-k/2)!} = \frac{k}{n} \cdot N_{n,k}.$$

□

Segundo Feller (1957), há várias interpretações para o problema tratado no Teorema 1. Originalmente, ele foi enunciado em 1887 por Bertrand. Em uma delas, considera-se uma votação com dois candidatos P e Q , com p e q votos respectivamente. Admitindo $p > q$, a probabilidade de não haver mudança na liderança durante a contagem de votos é $p^{-q}/p+q$. Não podemos utilizar o Teorema 1 para resolver a disputa de episódios pois ele exige o conhecimento da quantidade de votos recebido por cada concorrente.

Denotaremos

$$L_{2n} = \frac{1}{n+1} \cdot \binom{2n}{n}. \quad (9)$$

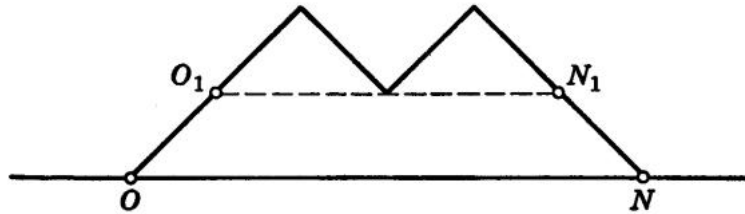
Teorema 2. *Entre os $N_{2n,0}$ caminhos ligando a origem ao ponto $2n$ do eixo x , há:*

(a) *exatamente L_{2n-2} caminhos tais que:*

$$s_1 > 0, s_2 > 0, \dots, s_{2n-1} > 0, s_{2n} = 0 \quad (10)$$

²⁴ basta somar as duas equações primeiras para encontrar a última igualdade

Figura 9 - Caminho típico



Fonte: Feller (1957).

(b) exatamente L_{2n} caminhos tais que:

$$s_1 \geq 0, s_2 \geq 0, \dots, s_{2n-1} \geq 0, s_{2n} = 0 \quad (11)$$

(c) exatamente L_{2n} caminhos tais que:

$$s_1 \geq 0, s_2 \geq 0, \dots, s_{2n-1} = 1, s_{2n} = 2 \quad (12)$$

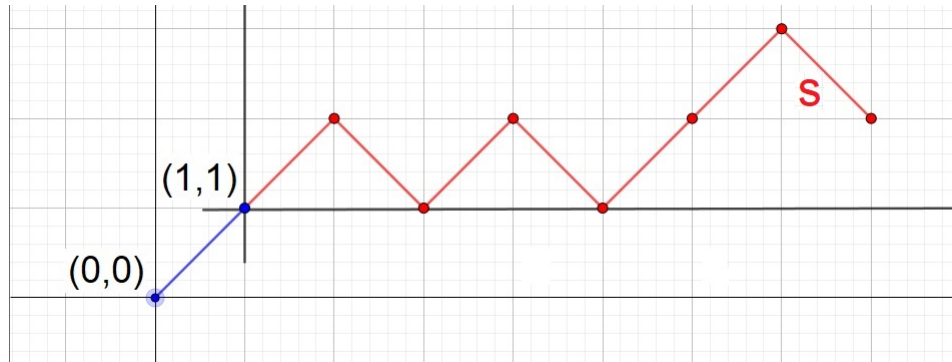
Prova. Cada caminho que satisfaz a condição (10) passa pelo ponto $N_1 = (2n-1, 1)$ e pelo Teorema 1, o número de caminhos da origem até N_1 tais que $s_1 > 0, s_2 > 0, \dots, s_{2n-2} > 0$ é igual a

$$\frac{1}{2n-1} \cdot \binom{2n-1}{n-1} = \frac{1}{2n-1} \cdot \frac{(2n-2)!}{n!(n-1)!} = \frac{1}{n} \cdot \binom{2n-2}{n-1} = L_{2n-2}. \quad (13)$$

Isto prova (a).

Considere um caminho s de comprimento $2n+2$ que satisfaz a condição (10). Isto é, s vai da origem $O = (0, 0)$ até $N = (2n+2, 0)$ e $s_t > 0$ para $t = 1, 2, \dots, 2n$. Omitindo o primeiro e o último lados, temos um caminho s' que liga o ponto $O_1 = (1, 1)$ para $N_1 = (2n-1, 1)$ e ao mesmo tempo é tal que todos seus vértices estão em cima ou acima da reta $y = 1$ (veja Figura 9). Considerando a origem em O_1 , temos um caminho da nova origem ao ponto N_1 (que tem as novas coordenadas $(2n, 0)$), nenhum desses vértices estão abaixo do novo eixo x . Temos então estabelecido uma correspondência um a um entre caminhos que satisfazem (11) e todos aqueles de comprimento $2n+2$ que satisfazem (10), e o teorema fica provado.

Sejam $\mathcal{A} = \{s, s_1 \geq 0, s_2 \geq 0, \dots, s_{2n-1} \geq 0, s_{2n} = 0\}$ e $\mathcal{B} = \{s, s_1 \geq 0, s_2 \geq 0, \dots, s_{2n-1} = 1, s_{2n} = 2\}$ e $F : \mathcal{A} \rightarrow \mathcal{B}$ tal que para cada $s \in \mathcal{A}$ associa $s' = F(s) = (s_1, \dots, s_{2n-1}, 2)$. F é uma bijeção. Logo, \mathcal{A} e \mathcal{B} têm a mesma quantidade de elementos. \square

Figura 10 - Caminho s 

Fonte: A autora, 2017.

Como

$$|\{s, s_1 > 0, s_2 > 0, \dots, s_{2n-1} > 0, s_{2n} = 0\}| = L_{2n-2}$$

e

$$|\{s_1 \geq 0, s_2 \geq 0, \dots, s_{2n-3} \geq 0, s_{2n-2} = 0\}| = L_{2n-2},$$

o Teorema 2 nos diz que há tantos caminhos para $(2n, 0)$ com todos os vértices interiores acima do eixo x como há para $(2n - 2, 0)$ sem vértices abaixo do eixo x .

Considere

- \mathcal{Z}_t é conjunto dos caminhos que satisfazem²⁵ $s_t = 0$.
- \mathcal{P}_t é o conjunto dos caminhos que satisfazem $s_1 > 0, \dots, s_t > 0$.
- \mathcal{N}_t é o conjunto dos caminhos que satisfazem $s_1 \geq 0, \dots, s_t \geq 0$.

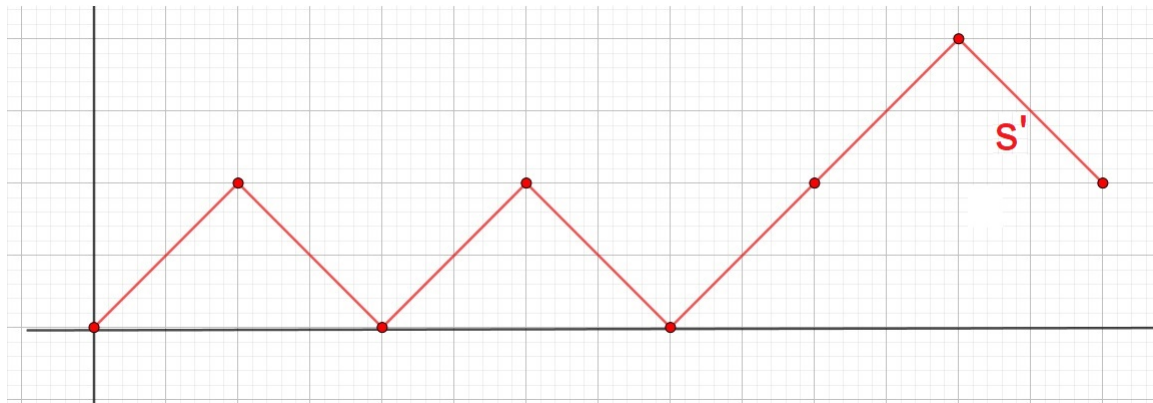
Mostraremos agora algumas relações úteis entre os conjuntos $\mathcal{Z}_t, \mathcal{P}_t$ e \mathcal{N}_t

Lema 2. *Há uma correspondência um para um entre \mathcal{P}_{2m} e \mathcal{N}_{2m-1} .*

Prova. Todo caminho s de \mathcal{P}_{2m} passa por $(1, 1)$ e satisfaz $s_j \geq 1$ para $j = 1, \dots, 2m$, pois em \mathcal{P}_{2m} , todas as somas parciais são positivas. (No contexto da disputa de episódios, como o primeiro voto foi 1, vemos que ele foi para o Episódio A.)

Vamos considerar o ponto $(1, 1)$ como nova origem do plano cartesiano (Figura 10). Deste modo, cria-se a partir de s , um caminho s' de comprimento $2m - 1$ (Figura 11). O

²⁵ Quando Border (2017) utiliza S_t , ele se refere aos somatório dos votos; quando usa s_t se refere aos caminhos, necessariamente ao ponto gerado pelo voto t apurado que gerou o somatório S_t .

Figura 11 - Caminho s' partindo da nova origem $(1, 1)$ 

Fonte: A autora, 2017.

primeiro voto foi contabilizado, sobraram $2m - 1$. Precisamente, $s' = (s'_0, s'_1, \dots, s'_{2m-1}) = (s_1 - 1, s_2 - 1, \dots, s_{2m} - 1)$.

Dessa forma, $s' \in \Gamma_{2m-1}$. □

Lema 3 (Lema de Nelson). *Há uma correspondência um para um entre \mathcal{Z}_{2m} e \mathcal{N}_{2m} . Mais ainda, a cada caminho em \mathcal{Z}_{2m} que possui um valor mínimo $-k$, corresponde um caminho em \mathcal{N}_{2m} que termina em $(2m, 2k)$.*

Prova. Para provar o lema, indicaremos como construir uma função bijetiva $F : \mathcal{Z}_{2m} \rightarrow \mathcal{N}_{2m}$.

Considere um caminho s em \mathcal{Z}_{2m} . Ele assume um valor mínimo $-k^* \leq 0$ em algum ponto $t \leq 2m$, possivelmente $-k^*$ é assumido mais de uma vez. Seja t^* o menor t para o qual $s_t = -k^*$.

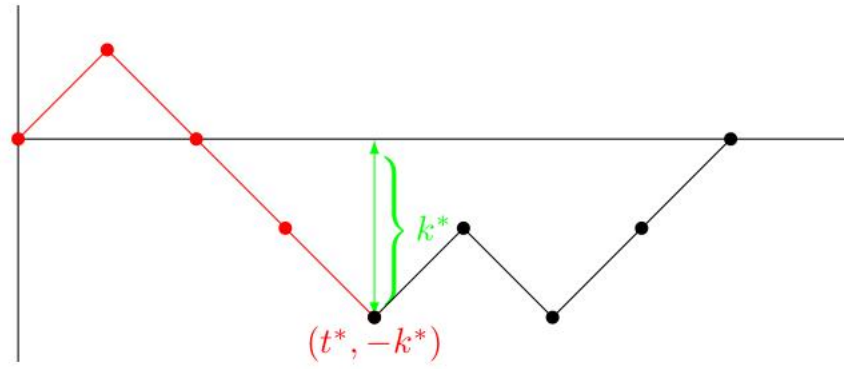
Observe que se o caminho s já é um elemento de \mathcal{N}_{2m} , temos $s_t \geq 0$ para $t = 0, \dots, 2m$. Consequentemente, $k^* = 0$ e $t^* = 0$. Neste caso, definimos $F(s) = s$. Caso s não pertença a \mathcal{N}_{2m} , temos necessariamente $s_t < 0$ para algum $0 < t < 2m$. Logo, $k^* > 0$ e $0 < t^* < 2m$ (Figura 12). Vamos construir um caminho s' em \mathcal{N}_{2m} da seguinte maneira (veja Figura 13):

- Tome o trecho do caminho s que vai de $(0, 0)$ até $(t^*, -k^*)$.
- Reflita este trecho em relação à reta vertical $t = t^*$.
- Deslize o trecho refletido até que o antigo ponto final $(t^*, -k^*)$ coincida com o ponto $(2m, 0)$.
- Considere $(t^*, -k^*)$ como a nova origem do plano cartesiano.

O caminho s' pertence a \mathcal{N}_{2m} e definimos $s' = F(s)$ (veja Figura 14)

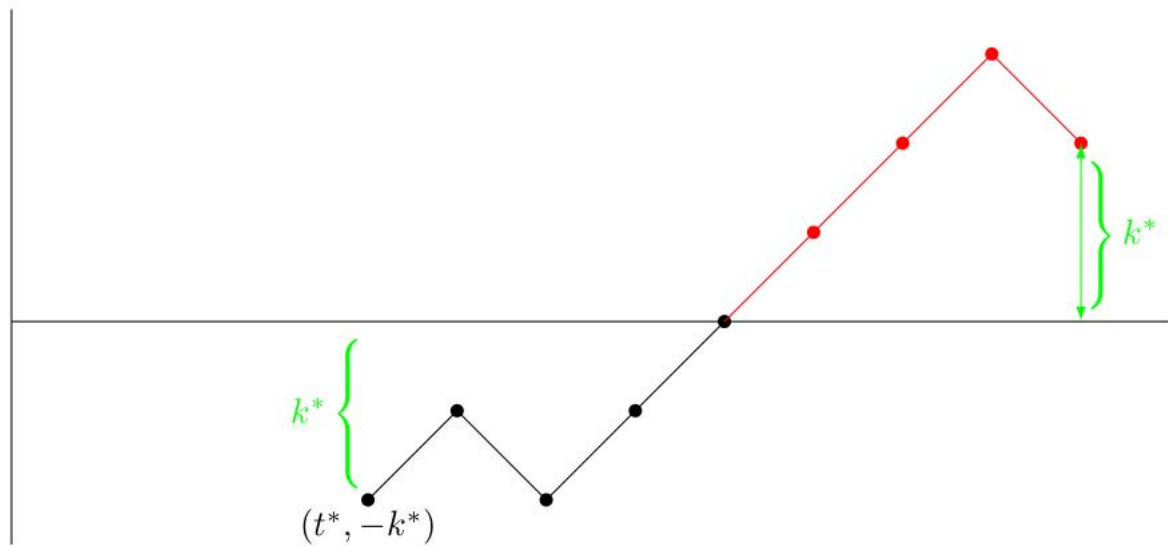
$$s' = (s_{t^*} + k^*, s_{t^*+1} + k^*, \dots, s_{2m} + k^*, s_{t^*-1} + 2k^*, \dots, s_1 + 2k^*, 2k^*).$$

Figura 12 - O caminho s não pertence a \mathcal{N}_{2m}

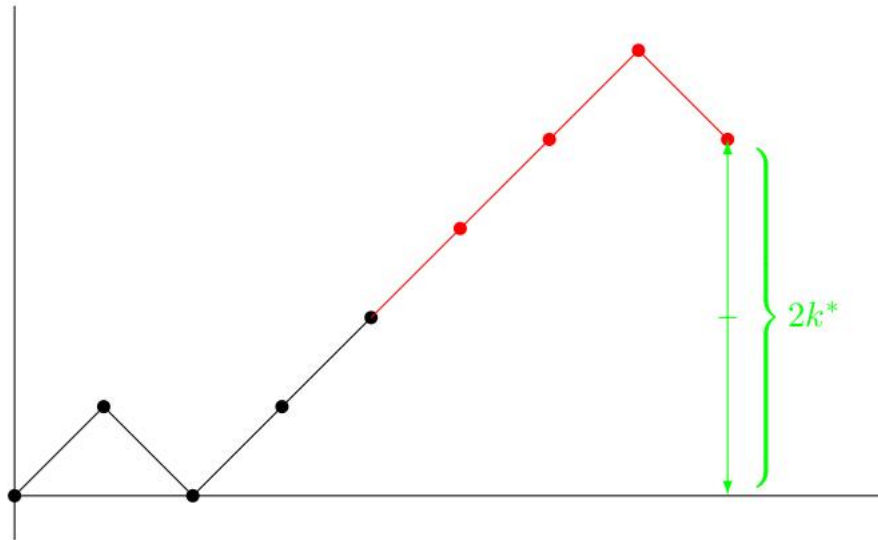


Fonte: Border (2017).

Figura 13 - Construção do caminho s'



Fonte: Border (2017).

Figura 14 - O caminho s' 

Fonte: Border (2017).

Mostraremos agora que F é inversível: seja s um caminho em \mathcal{N}_{2m} . Se $s_{2m} = 0$, temos $F^{-1}(s) = s$. Se $s_{2m} > 0$, sabemos pelo Corolário 1 que s_{2m} é par. Isto é, $s_{2m} = 2\bar{k}$, para algum inteiro $\bar{k} > 0$. Considere \bar{t} a última época em que $s_t = \bar{k}$ (veja Figura 15) Vamos construir um caminho s' em \mathcal{Z}_{2m} da seguinte maneira:

- Tome o trecho do caminho s que vai de (\bar{t}, \bar{k}) até $(2m, 2\bar{k})$.
- Reflita este trecho em relação à reta vertical $t = \bar{t}$.
- Deslize o trecho refletido até que o antigo ponto final (\bar{t}, \bar{k}) coincida com a origem (veja Figura 16).
- Considere o ponto inicial como a nova origem do plano cartesiano (veja Figura 17).

O caminho s' pertence a \mathcal{Z}_{2m} :

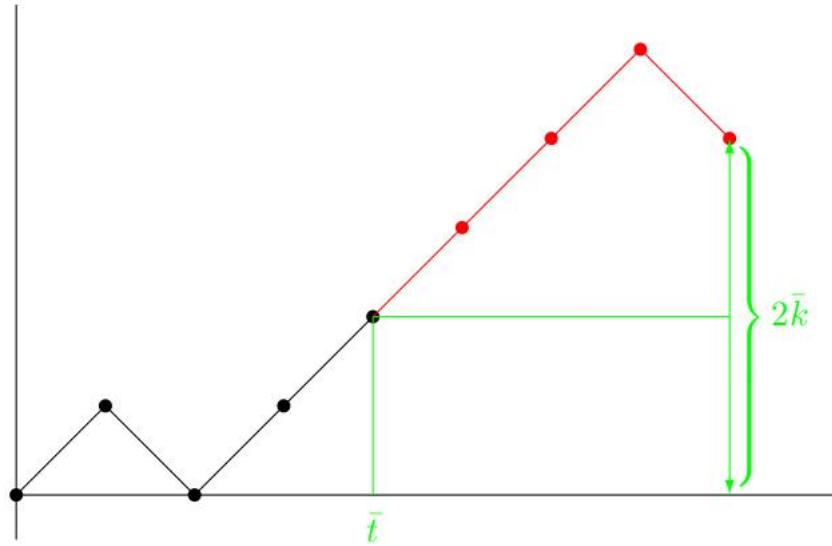
$$s' = (s_{2m} - 2\bar{k}, s_{2m-1} - 2\bar{k}, \dots, s_{\bar{t}} - 2\bar{k}, s_1 - \bar{k}, \dots, s_{\bar{t}-1} - \bar{k}, 0).$$

O procedimento descrito para construção de s' inverte o descrito na definição de F . Provamos assim a correspondência um a um entre \mathcal{Z}_{2m} e \mathcal{N}_{2m} . \square

7.4.2 Lema principal

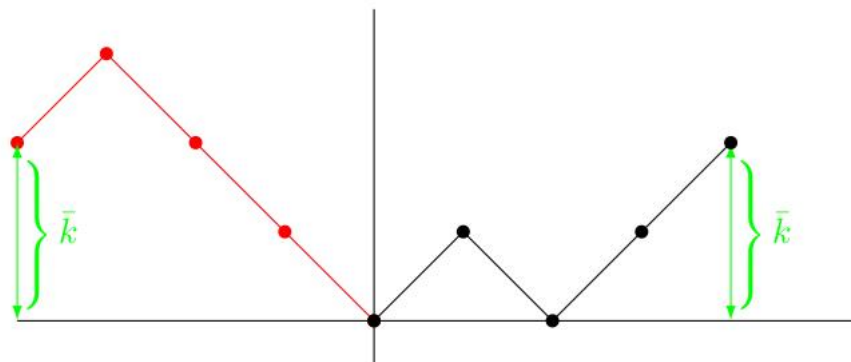
Border (2017) diz que a abordagem geométrica será utilizada para provar o Lema a seguir.

Figura 15 - Última época em que $s_t = \bar{k}$

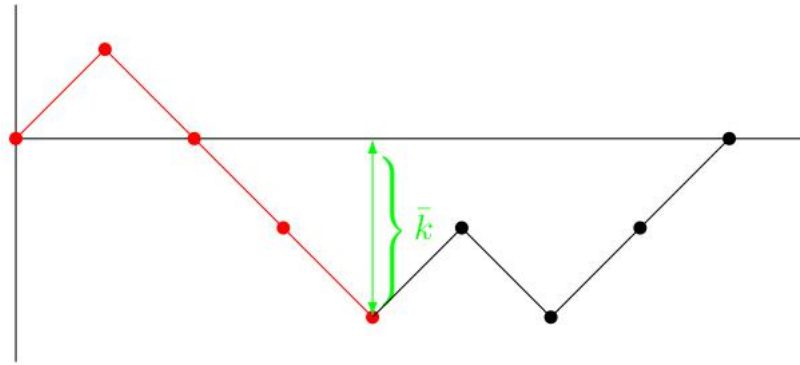


Fonte: Border (2017).

Figura 16 - Deslize o trecho refletido até a origem



Fonte: Border (2017).

Figura 17 - O caminho $s' = F^{-1}(s)$ 

Fonte: Border (2017).

Lema 4. *São idênticas as probabilidades:*

$$P(S_{2m} = 0), \quad (14)$$

$$P(S_1 \neq 0, \dots, S_{2m} \neq 0), \quad (15)$$

$$P(S_1 \geq 0, \dots, S_{2m} \geq 0), \quad (16)$$

$$P(S_1 \leq 0, \dots, S_{2m} \leq 0), \quad (17)$$

$$2P(S_1 > 0, \dots, S_{2m} > 0), \quad (18)$$

$$2P(S_1 < 0, \dots, S_{2m} < 0). \quad (19)$$

Prova. Observe que todos os eventos citados podem ser associados a caminhos de comprimento $2m$. Para calcular a probabilidade $P(E)$ de um evento E associado a caminhos de comprimento $2m$, é preciso determinar a quantidade de caminhos associadas ao evento E e dividir por 2^{2m} .

A probabilidade em (16) está associada à situação em que o Episódio A sempre vence. Podem ocorrer empates, mas A sempre está na liderança, pois não há $S_t < 0$. O caminho não cruza o eixo x . Assim, não há troca de liderança. Na probabilidade (17), é a mesma situação, porém quem sempre lidera é B . Nos dois casos, por simetria²⁶, temos

$$|\{s, s_1 \geq 0, \dots, s_{2m} \geq 0\}| = |\{s, s_1 \leq 0, \dots, s_{2m} \leq 0\}|.$$

Portanto, $P(S_1 \geq 0, \dots, S_{2m} \geq 0) = P(S_1 \leq 0, \dots, S_{2m} \leq 0)$.

Analogamente, (18) e (19) correspondem a vitórias, sem empate, para A e B

²⁶ Basta definir $F : \{s, s_1 \geq 0, \dots, s_{2m} \geq 0\} \rightarrow \{s, s_1 \leq 0, \dots, s_{2m} \leq 0\}$, $F(s) = -s$.

respectivamente. Por simetria, temos

$$|\{s, s_1 > 0, \dots, s_{2m} > 0\}| = |\{s, s_1 < 0, \dots, s_{2m} < 0\}|.$$

Portanto, $P(S_1 > 0, \dots, S_{2m} > 0) = P(S_1 < 0, \dots, S_{2m} < 0)$.

Vamos agora calcular $P(S_1 \neq 0, \dots, S_{2m} \neq 0)$. Note que

$$\{s, s_1 \neq 0, \dots, s_{2m} \neq 0\} = \{s, s_1 > 0, \dots, s_{2m} > 0\} \cup \{s, s_1 < 0, \dots, s_{2m} < 0\}.$$

Portanto

$$\begin{aligned} P(S_1 \neq 0, \dots, S_{2m} \neq 0) &= \frac{|\{s, s_1 \neq 0, \dots, s_{2m} \neq 0\}|}{2^{2m}} \\ &= \frac{|\{s, s_1 > 0, \dots, s_{2m} > 0\}| + |\{s, s_1 < 0, \dots, s_{2m} < 0\}|}{2^{2m}} \\ &\stackrel{(18)}{=} 2P(S_1 > 0, \dots, S_{2m} > 0) \stackrel{(19)}{=} 2P(S_1 < 0, \dots, S_{2m} < 0). \end{aligned}$$

Será demonstrada a igualdade entre (16) e (18). O Lema 2 mostra que há uma bijeção entre os conjuntos \mathcal{P}_{2m} e \mathcal{N}_{2m-1} . Logo estes conjuntos possuem a mesma quantidade de elementos.

$$|\mathcal{P}_{2m}| = |\mathcal{N}_{2m-1}|.$$

Como $2m - 1$ é ímpar, sabe-se pelo Corolário 1 que $s'_{2m-1} > 0$ para todo $s' \in \mathcal{N}_{2m-1}$. Somente para uma quantidade par de votos seria possível obter um empate no último voto. Com a bijeção, cada caminho s em \mathcal{P}_{2m} pode ser associado a um único caminho s' em \mathcal{N}_{2m-1} . Como queremos comparar as probabilidades $P(S_1 > 0, \dots, S_{2m} > 0)$ e $P(S_1 \geq 0, \dots, S_{2m} \geq 0)$, vamos construir a partir de s' um elemento s'' de \mathcal{N}_{2m}

$$s'' = (s'_1, \dots, s'_{2m-1}, s''_{2m})$$

Só há duas possibilidades para o último voto. Se for para o Episódio A , será adicionado $+1$ a s'_{2m-1} , assim $s''_{2m} > 0$. Se o filme escolhido for o B , será somado -1 , assim, $s''_{2m} \geq 0$. Assim, o número de caminhos do conjunto \mathcal{N}_{2m} é o dobro de \mathcal{N}_{2m-1} . Isto é,

$$|\mathcal{N}_{2m}| = 2|\mathcal{N}_{2m-1}|.$$

Então

$$P(S_1 \geq 0, \dots, S_{2m} \geq 0) = \frac{|\mathcal{N}_{2m}|}{2^{2m}} = \frac{2|\mathcal{N}_{2m-1}|}{2^{2m}} = \frac{2|\mathcal{P}_{2m}|}{2^{2m}} = 2P(S_1 > 0, \dots, S_{2m} > 0).$$

Isto é

$$P(S_1 \geq 0, \dots, S_{2m} \geq 0) = P(S_1 \geq 0, \dots, S_{2m-1} \geq 0) = 2P(S_1 > 0, \dots, S_{2m} > 0).$$

Concluí-se que as probabilidades em (16) e (18) são iguais. Para concluir a prova do lema, vamos mostrar que as probabilidades em (14) e (16) são iguais. Pelo Lema de Nelson (Lema 3), temos

$$|\mathcal{Z}_{2m}| = |\mathcal{N}_{2m}|.$$

Então

$$P(S_{2m} = 0) = \frac{|\mathcal{Z}_{2m}|}{2^{2m}} = \frac{|\mathcal{N}_{2m}|}{2^{2m}} = P(S_1 \geq 0, \dots, S_{2m} \geq 0).$$

□

7.4.3 Empates

Deseja-se saber se houve troca na liderança de um dos filmes. Assim, é preciso contar o número de vezes que os caminhos cruzaram o eixo x . Primeiramente, é preciso contar quantas vezes eles tocaram o eixo mencionado. Estudaremos agora os caminhos que unem a origem a um ponto N do eixo x (observe que uma abscissa ímpar no eixo x é impossível como ponto final de caminhos deste tipo). Temos $N = (2n, 0)$.

Definição 5 (Retornos ao zero). *Quando um caminho toca o eixo x , diz-se que retornou a zero ou à origem. Neste caso, $s_t = 0$ para algum voto t .*

Segue da Proposição 3 que t é par. Podemos também entender esta condição considerando um caso particular. Suponha que os primeiros a votos foram para o Episódio A . Assim, a partir da origem, $(0, 0)$, foram adicionados a segmentos de comprimento 1 com coeficiente angular $+1$. Para que s_t seja zero, é preciso que se tenha a mesma quantidade de votos no Episódio B . Dessa forma, $s_t = 0$ após $t = 2a$ votos.

Considere $t = 2n$. O número de caminhos da origem a $(2n, 0)$ é $N_{2n,0}$, então pelo Corolário 2, a probabilidade u_{2n} de um caminho de comprimento $2n$ terminar no ponto $(2n, 0)$ é dada por:

$$u_{2n} = \binom{2n}{n} \cdot 2^{-2n}. \tag{20}$$

$u_{2n} = P(S_{2n} = 0)$ é a probabilidade de ocorrer um empate²⁷ na época $t = 2n$.

Definição 6 (Primeiro retorno ao zero). *Quando um caminho toca o eixo x na época $2m$ e $s_1 \neq 0, s_2 \neq 0, \dots, s_{2m-1} \neq 0$, diz-se que aconteceu o primeiro retorno a zero. Denotamos por f_{2m} a probabilidade de ocorrer o primeiro retorno a zero na época $2m$. Isto é,*

$$f_{2m} = P(S_1 \neq 0, S_2 \neq 0, \dots, S_{2m-1} \neq 0, S_{2m} = 0).$$

Como consequência do Lema 4, temos

Corolário 4. *Uma fórmula explícita para f_{2m} é*

$$f_{2m} = u_{2m-2} - u_{2m} = \frac{1}{2m-1} u_{2m} = \frac{1}{2m-1} \binom{2m}{m} \frac{1}{2^{2m}}, \quad (m = 1, 2, \dots). \quad (21)$$

Prova. Como $2m - 1$ é ímpar, pelo Corolário 1, temos necessariamente $s_{2m-1} \neq 0$. Logo, podemos escrever o evento primeiro retorno na época $2m$ como $(S_1 \neq 0, S_2 \neq 0, \dots, S_{2m-2} \neq 0, S_{2m} = 0)$. Os caminhos associados a este evento pertencem à diferença entre dois conjuntos: o conjunto dos eventos em que s_t nunca se anula até a época $2m - 2$ menos o dos eventos em que s_t nunca se anula até a época $2m$. Isto é,

$$\begin{aligned} \{s, s_1 \neq 0, s_2 \neq 0, \dots, s_{2m-2} \neq 0, s_{2m} = 0\} = \\ = \{s, s_1 \neq 0, s_2 \neq 0, \dots, s_{2m-2} \neq 0\} \setminus \{s, s_1 \neq 0, s_2 \neq 0, \dots, s_{2m} \neq 0\}. \end{aligned}$$

Observe que para caminhos de comprimento $2m$, temos

$$\{s, s_1 \neq 0, s_2 \neq 0, \dots, s_{2m} \neq 0\} \subset \{s, s_1 \neq 0, s_2 \neq 0, \dots, s_{2m-2} \neq 0\}.$$

Além, disso a quantidade de caminhos de comprimento $2m$ tais que $s_1 \neq 0, s_2 \neq 0, \dots, s_{2m-2} \neq 0$ é igual a quatro vezes a a quantidade de caminhos de comprimento $2m - 2$ tais que $s_1 \neq 0, s_2 \neq 0, \dots, s_{2m-2} \neq 0$. temos

$$\begin{aligned} P((S_1 \neq 0, S_2 \neq 0, \dots, S_{2m-2} \neq 0) \setminus (S_1 \neq 0, S_2 \neq 0, \dots, S_{2m} \neq 0)) \\ = \frac{4 \cdot |\{s, s_1 \neq 0, s_2 \neq 0, \dots, s_{2m-2} \neq 0\}| - |\{s, s_1 \neq 0, s_2 \neq 0, \dots, s_{2m} \neq 0\}|}{2^{2m}} \\ = P(S_1 \neq 0, S_2 \neq 0, \dots, S_{2m-2} \neq 0) - P(S_1 \neq 0, S_2 \neq 0, \dots, S_{2m} \neq 0) \end{aligned}$$

que pelo Lema 4 é igual a

$$= u_{2m-2} - u_{2m}.$$

²⁷ Nada se sabe sobre a quantidade de vezes que isso ocorreu até $t = 2n$.

Por (20), temos

$$\begin{aligned}
 u_{2m-2} - u_{2m} &= \binom{2m-2}{m-1} \cdot 2^{-2(m-1)} - \binom{2m}{m} \cdot 2^{-2m} \\
 &= \frac{(2m-2)!}{(m-1)!(m-1)!} \frac{1}{2^{2m-2}} - u_{2m} = \frac{4m^2}{2m(2m-1)} u_{2m} - u_{2m} \\
 &= \left(\frac{2m}{2m-1} - 1 \right) u_{2m} = \frac{1}{2m-1} u_{2m} = \frac{1}{2m-1} \binom{2m}{m} \frac{1}{2^{2m}}
 \end{aligned}$$

□

Podemos obter também uma fórmula recursiva para o retorno ao zero envolvendo o primeiros retornos.

Corolário 5. *Para $m \geq 1$, temos*

$$u_{2m} = \sum_{r=1}^m f_{2r} u_{2m-2r}. \quad (22)$$

Prova. Se um retorno à origem acontece na época $2m$, então o *primeiro* retorno ao zero acontece em uma época $2r \leq 2m$. Neste caso, todo caminho s de comprimento $2m$ da origem até o ponto $(2m, 0)$ cujo primeiro retorno ao zero aconteceu no ponto $(2r, 0)$ pode ser decomposto em um trecho de comprimento $2r$ em que $s_1 \neq 0, s_2 \neq 0, \dots, s_{2r-1} \neq 0$ e um trecho de comprimento $2m - 2r$ em que $s_{2m} = 0$. Portanto, a quantidade de caminhos de comprimento $2m$ da origem até o ponto $(2m, 0)$ cujo primeiro retorno ao zero aconteceu no ponto $(2r, 0)$ é dada por

$$2^{2r} \cdot f_{2r} \cdot 2^{2m-2r} \cdot u_{2m-2r}.$$

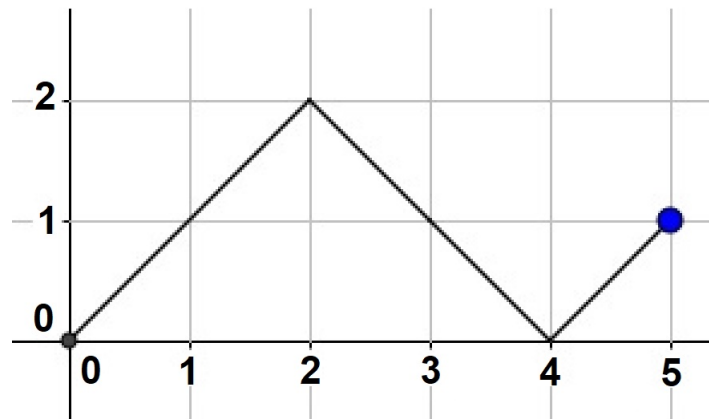
Somando sobre r , obtemos (22). □

7.5 Na liderança

O foco desta seção é analisar a probabilidade de um dos filmes permanecer na liderança. A probabilidade ter sido escolhido pela maioria dos espectadores. Feller (1957) afirma que a probabilidade de haver troca na liderança é próxima de zero. Em caso de empate na época t , Feller (1957 apud MLODINOW, 2009) define que a liderança é do competidor que ganhava até o voto $t - 1$. É uma hipótese razoável pois o empate não caracteriza troca na liderança.

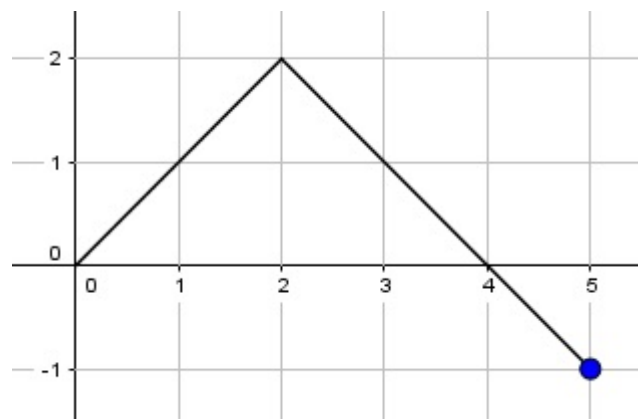
Feller (1957) descreve um apostador fictício Peter. A cada jogada de uma moeda honesta ele ganha ou perde uma unidade. A sequência de vitórias e derrotas é associada a um caminho, uma linha poligonal em que cada vitória é associada a um segmento de

Figura 18 - Vitória de Peter



Fonte: A autora, 2017.

Figura 19 - Derrota de Peter



Fonte: A autora, 2017.

comprimento 1 e de coeficiente angular $+1$. Quando Peter perde, pode-se associar a um segmento de comprimento 1 e de coeficiente angular -1 . Quando a linha poligonal tocar o eixo horizontal, a quantidade de sucessos e o número de fracassos estarão empatados. Se a linha cruzar o eixo, isso implica mudança na sorte de Peter, troca da liderança. Se o último segmento estiver acima do eixo x , Peter venceu. Se o segmento estiver abaixo do eixo, ele perdeu.

Pode-se pensar nesta estrutura com dois participantes. Considere os jogadores A e B . O segmento com coeficiente angular positivo é a representação da vitória de A em uma tentativa. Já o segmento com coeficiente angular negativo, é o sucesso de B em uma jogada. Cada tentativa irá corresponder a um segmento formando uma linha poligonal. As ordenadas dos vértices dessa linha representam o saldo dos pontos acumulados pelos

dois jogadores ao longo de uma sequência de rodadas. Cada poligonal determina um caminho diferente. Todos os possíveis caminhos caracterizam todos os possíveis resultados de partidas sucessivas entre os jogadores A e B . É preciso analisar todos os caminhos possíveis a fim de calcular a probabilidade de mudar a liderança. Não é factível listar todos os casos possíveis. Serão utilizados teoremas.

Analogamente, podemos adaptar esta perspectiva para a disputa entre os episódios. Toda vez que o Episódio A receber um voto, será acrescentado um segmento de comprimento 1 com coeficiente angular positivo. Quando o Episódio B receber um voto, vamos associar um segmento de comprimento 1 e de coeficiente angular -1 . Assim, a sequência de votos dados aos episódios corresponderá a uma poligonal. Sempre que ela estiver acima do eixo x , significa que A está com mais votos acumulados do que B . Analogamente, para o caso de B estar vencendo, a poligonal estará abaixo do eixo x . Dessa forma, o Episódio A está na liderança enquanto a poligonal estiver acima do eixo x . Quando estiver abaixo do eixo, B estará a frente da disputa. Quando a poligonal encosta no eixo x ao se contar o k -ésimo voto, há empate na soma acumulada de pontos de cada episódio. Neste caso, ainda não houve a mudança na liderança. Para verificar quem tem vantagem até este momento, é preciso verificar qual episódio era o líder até o voto $k - 1$.

É importante observar, que a análise de liderança feita por Feller (1957) e adaptada por Mlodinow (2009) para a disputa de episódios não está interessada na liderança ininterrupta e sim na análise dos períodos em que cada episódio se manteve na liderança. Para esclarecer este aspecto, vamos retomar Feller (1957). A apuração parcial da votação nos episódios A e B até o t -ésimo voto pode ser acompanhada em S_t ou s_t .

Em conformidade com Feller (1957)²⁸, diremos que o episódio A está na liderança entre a época $k - 1$ e k se o segmento que liga os pontos $(k - 1, s_{k-1})$ e (k, s_k) encontra-se acima do eixo x , isto é, se ao menos um dos vértices s_{k-1} ou s_k é positivo (neste caso, o outro vértice é positivo ou nulo). Analogamente, diremos que o episódio B está na liderança entre a época $k - 1$ e k se o segmento que liga os pontos $(k - 1, s_{k-1})$ e (k, s_k) encontra-se abaixo do eixo x , isto é, se ao menos um dos vértices s_{k-1} ou s_k é negativo (neste caso, o outro vértice é negativo ou nulo).

Feller (1968, p. 78) observa que nossa intuição nos leva a acreditar que em uma longa disputa de lançamentos de uma moeda entre dois jogadores, cada um deles permaneceria cerca de metade do tempo na liderança, e que a troca de liderança não seria um pouco frequente. No entanto, o próximo resultado nos mostra que neste caso, somos traídos por nossa intuição. O próximo teorema analisa a fração do tempo total em que a partícula permanece acima do eixo x . No contexto da disputa dos episódios, o teorema

²⁸ We shall say that *the particle spends the time from $k - 1$ to k on the positive side if the k th side of its path lies above the x -axis*, that is, if at least one of the two vertices s_{k-1} and s_k is positive (in which case the other is positive or zero). (FELLER, 1957, p. 77)

nos dirá a probabilidade do episódio A permanecer na liderança ao longo de uma fração do total de votos. Por vezes, esperamos que a fração $1/2$ (isto é, permanecer na liderança durante metade dos votos) seja a mais provável. No entanto, acontece exatamente o contrário. As frações próximas a $1/2$ são as menos prováveis. Enunciaremos o teorema²⁹ no contexto da disputa de episódios.

Teorema 3. *Seja $\alpha_{2k,2n}$ a probabilidade de que durante o intervalo de 0 a $2n$ o episódio A permaneça durante $2k$ votos na liderança e o episódio B permaneça durante $2n - 2k$ votos na liderança. Então*

$$\alpha_{2k,2n} = u_{2k}u_{2n-2k}, \quad k = 0, 1, \dots, n. \quad (23)$$

Prova. A probabilidade de que o episódio A permaneça na liderança durante todo o intervalo de 0 a $2n$ é dada por (16) e coincide, pelo Lema 4, com (14) que foi calculada em (20):

$$P(S_1 \geq 0, \dots, S_{2n} \geq 0) = P(S_{2n} = 0) = u_{2n} = \binom{2n}{n} \cdot 2^{-2n}.$$

Portanto, $\alpha_{2n,2n} = u_{2n}u_0 = u_{2n}$. Analogamente, usando (17) do Lema 4, obtemos $\alpha_{0,2n} = u_0u_{2n} = u_{2n}$. Isto é, cada episódio tem a mesma probabilidade de permanecer na liderança durante todo o intervalo de 0 a $2n$.

Resta provar (23) para $1 \leq k \leq n - 1$. A prova será feita por indução sobre n . Antes da indução, vamos deduzir uma expressão para $\alpha_{2k,2n}$ recursiva que envolve também probabilidades de primeiro retorno ao zero.

Se o episódio A permaneça durante $2k$ votos na liderança e o episódio B permaneça durante $2n - 2k$ votos na liderança, em algum momento houve um empate. Seja $2r$ a época em que ocorreu o primeiro retorno ao zero. Como $1 \leq k \leq n - 1$, temos $2r < 2n$. Isto é, o primeiro empate deve acontecer antes do término da votação. Do contrário, apenas um dos episódios permaneceria o tempo todo na liderança. Assim, como o primeiro empate acontece na época $2r$, o caminho s associado à disputa dos episódios pertence a uma das duas classes:

- Na primeira classe, o episódio A liderou durante todo o intervalo de 0 a $2r$.

²⁹ Em Feller (1957, p. 77), Theorem 1: Let $\alpha_{2k,2n}$ be the probability that in the time interval from 0 to $2n$ the particle spends $2k$ time units on the positive side and $2n - 2k$ time units on the positive side and $2n - 2k$ time units on the negative side. Then

$$\alpha_{2k,2n} = u_{2k}u_{2n-2k}$$

- Na segunda classe, o episódio B liderou durante todo o intervalo de 0 a $2r$.

Na primeira classe, temos necessariamente $r \leq k \leq n - 1$, e no trecho do caminho s posterior ao vértice $(2r, 0)$, o episódio A permanece na liderança exatamente por mais $2k - 2r$ votos. Vamos calcular quantos caminhos de comprimento $2n$ há na primeira classe. Note que temos $2^{2r} \cdot f_{2r}$ caminhos de comprimento $2r$ cujo primeiro retorno acontece na época $2r$. Por simetria, em metade deles o episódio A lidera todo tempo. Além disso, pela definição de $\alpha_{2k, 2n}$, há $2^{2n-2r} \cdot \alpha_{2k-2r, 2n-2r}$ caminhos de comprimento $2n - 2r$ em que episódio A permanece durante $2k - 2r$ votos na liderança e o episódio B permaneça durante $2n - 2k$ votos na liderança. Consequentemente, o número total de caminhos de comprimento $2n$ na primeira classe é dado por

$$\frac{1}{2} \cdot 2^{2r} \cdot f_{2r} 2^{2n-2r} \cdot \alpha_{2k-2r, 2n-2r} = 2^{2n-1} \cdot f_{2r} \cdot \alpha_{2k-2r, 2n-2r}$$

Na segunda classe, temos necessariamente $2k \leq 2n - 2r$, pois todo o tempo em que o episódio A permanece na liderança ocorrerá após o vértice $(2r, 0)$. Além disso, o episódio B liderou durante todo o intervalo de 0 a $2r$. Vamos calcular quantos caminhos de comprimento $2n$ há na segunda classe. Note que temos $2^{2r} \cdot f_{2r}$ caminhos de comprimento $2r$ cujo primeiro retorno acontece na época $2r$. Por simetria, em metade deles o episódio B lidera todo tempo. Além disso, pela definição de $\alpha_{2k, 2n}$, há $2^{2n-2r} \cdot \alpha_{2k, 2n-2r}$ caminhos de comprimento $2n - 2r$ em que episódio A permanece durante $2k$ votos na liderança. Consequentemente, o número total de caminhos de comprimento $2n$ na primeira classe é dado por

$$\frac{1}{2} \cdot 2^{2r} \cdot f_{2r} 2^{2n-2r} \cdot \alpha_{2k, 2n-2r} = 2^{2n-1} \cdot f_{2r} \cdot \alpha_{2k, 2n-2r}.$$

Observe que as classes são disjuntas entre si e também são disjuntas com as classes correspondentes a diferentes valores de r . Desse modo

$$\alpha_{2k, 2n} = \frac{1}{2} \sum_{r=1}^k f_{2r} \cdot \alpha_{2k-2r, 2n-2r} + \frac{1}{2} \sum_{r=1}^{n-k} f_{2r} \cdot \alpha_{2k, 2n-2r}. \quad (24)$$

De posse de (24) – expressão recursiva para $\alpha_{2k, 2n}$ que envolve também probabilidades de primeiro retorno ao zero – e pelo argumento apresentado inicialmente para $k = 0$ e $k = 2n$, temos

$$\alpha_{2k, 2n} = \begin{cases} u_0 u_{2n}, & k = 0 \\ \frac{1}{2} \sum_{r=1}^k f_{2r} \cdot \alpha_{2k-2r, 2n-2r} + \frac{1}{2} \sum_{r=1}^{n-k} f_{2r} \cdot \alpha_{2k, 2n-2r}, & 1 \leq k \leq n - 1 \\ u_{2n} u_0, & k = n \end{cases}$$

Vamos agora provar (23), por indução sobre n . Para $n = 1$, só temos dois possíveis valores

para k : $k = 0$ ou $k = 1$. Vimos na parte inicial da demonstração que (23) se verifica para $n = 1$. Nossa hipótese de indução é

$$\alpha_{2k,2\nu} = u_{2k}u_{2\nu-2k}, \quad \nu = 1, 2, \dots, n-1.$$

Observe que no primeiro somatório em (24), temos $n-k \leq n-r \leq n-1$. No segundo somatório, $k \leq n-r \leq n-1$. Usando a hipótese de indução em (24), obtemos

$$\begin{aligned} \alpha_{2k,2n} &= \frac{1}{2} \sum_{r=1}^k f_{2r} \cdot \alpha_{2k-2r,2n-2r} + \frac{1}{2} \sum_{r=1}^{n-k} f_{2r} \cdot \alpha_{2k,2n-2r} \\ &= \frac{1}{2} \sum_{r=1}^k f_{2r} \cdot u_{2k-2r}u_{2n-2k} + \frac{1}{2} \sum_{r=1}^{n-k} f_{2r} \cdot u_{2k}u_{2n-2r-2k} \\ &= \frac{1}{2} \cdot u_{2n-2k} \sum_{r=1}^k f_{2r} \cdot u_{2k-2r} + \frac{1}{2} \cdot u_{2k} \sum_{r=1}^{n-k} f_{2r} \cdot u_{2n-2r-2k} \end{aligned}$$

Por (22)

$$= \frac{1}{2} \cdot u_{2n-2k}u_{2k} + \frac{1}{2} \cdot u_{2k}u_{2n-2k} = u_{2k}u_{2n-2k}.$$

Portanto,

$$\alpha_{2k,2n} = \frac{1}{2} \cdot u_{2n-2k} \sum_{r=1}^k f_{2r} \cdot u_{2k-2r} + \frac{1}{2} \cdot u_{2k} \sum_{r=1}^{n-k} f_{2r} \cdot u_{2n-2r-2k} = u_{2k}u_{2n-2k}. \quad (25)$$

□

O evento para a situação dos episódios de George Lucas é que um dos filmes esteja na liderança. Pretende-se verificar se é mais provável que um deles esteja a frente ou que alternem na preferência dos espectadores. Deve-se lembrar que a escolha de cada pessoa está associada ao lançamento de uma moeda. Sem perda de generalidade, pode-se definir que ao sair cara, o voto é para o episódio A , quando sair coroa para o B . Por convenção, quando o voto for para o episódio A , será adicionado $+1$ à soma acumulada. Se o voto for para B , -1 . Pode-se fazer somas acumuladas a cada voto, S_1, S_2, \dots, S_x . Neste caso, são 20.000 espectadores, então o vetor formado por essas somas parciais tem 20.000 coordenadas, é uma 20.000-upla. Para saber para quem foi um determinado voto ε_i , é preciso utilizar as somas S_i e S_{i-1} como visto.

Vamos usar o Teorema 3 para determinar a probabilidade do episódio A permanecer na liderança ao longo de uma fração do total de votos. Sendo que $2k$ é o número de votos ao longo dos quais o episódio A permanece na liderança. Lembramos que

$$u_{2n} = \binom{2n}{n} \cdot 2^{-2n}.$$

Seja $2n = 20.000$ e $k = 0$, ou seja, o episódio A nunca assumirá a liderança. Substituindo na fórmula (23), tem-se:

$$\begin{aligned}\alpha_{0,20.000} &= \binom{0}{0} \cdot 2^{-2 \cdot 0} \cdot \binom{20.000}{10.000} \cdot 2^{-20.000} = 1 \cdot \binom{20.000}{10.000} \cdot 2^{-20.000} \\ &= \frac{20.000!}{10.000! \cdot 10.000!} \cdot \frac{1}{2^{20.000}} = \frac{20.000!}{(10.000!)^2 \cdot 2^{20.000}}\end{aligned}$$

Para $2k = 10.000$, ou seja, ambos os filmes permanecem na liderança pelo mesmo período, tem-se:

$$\begin{aligned}\alpha_{10.000,10.000} &= \binom{10.000}{5.000} \cdot 2^{-10.000} \cdot \binom{10.000}{5.000} \cdot 2^{-10.000} \\ &= \frac{10.000!}{(5.000!)^2} \cdot \frac{1}{2^{10.000}} \cdot \frac{10.000!}{(5.000!)^2} \cdot \frac{1}{2^{10.000}} = \frac{(10.000!)^2}{(5.000!)^4 \cdot 2^{20.000}}\end{aligned}$$

Para verificar que uma probabilidade é cerca de 88 vezes maior do que a outra, será feita a divisão³⁰:

$$\begin{aligned}\frac{20.000!}{(10.000!)^2 \cdot 2^{20.000}} \div \frac{(10.000!)^2}{(5.000!)^4 \cdot 2^{20.000}} &= \frac{20.000!}{(10.000!)^2 \cdot 2^{20.000}} \cdot \frac{(5.000!)^4 \cdot 2^{20.000}}{(10.000!)^2} \\ &= 20.000! \cdot \left(\frac{5.000!}{10.000!} \right)^4 \\ &\cong 1,8192 \cdot 10^{77.337} \cdot \left(\frac{4,2285 \cdot 10^{16.325}}{2,8462 \cdot 10^{35.659}} \right)^4 \\ &\cong 8,8625 \cdot 10 = 88,625\end{aligned}$$

A probabilidade de um dos filmes permanecer na liderança por todo o período da escolha dos 20.000 espectadores é praticamente 88 vezes maior do que a de haver igual períodos de liderança para ambos os filmes.

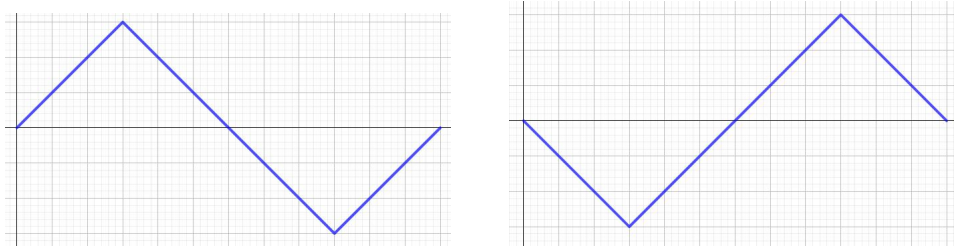
7.6 Troca de liderança

Possivelmente, quando olhamos para o problema da disputa dos episódios pela primeira vez e pensamos no cenário “permanecer na liderança por metade do tempo”, temos duas possibilidades em mente (Figura 20)

No entanto, no contexto do Teorema 3 os caminhos ilustrados na Figura 21 também

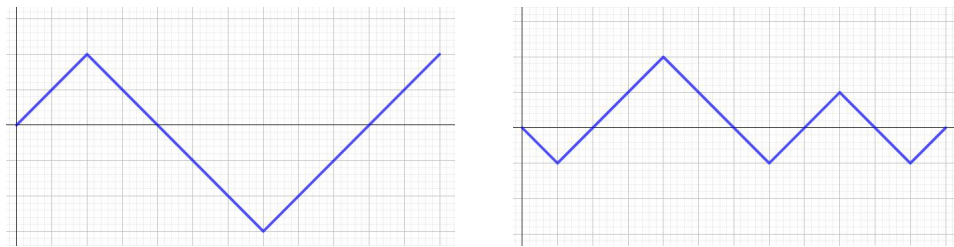
³⁰ Os fatoriais foram calculados por <https://www.calculatorsoup.com/calculators/discretemathematics/factorials.php>

Figura 20 - “Permanecer na liderança por metade do tempo”



Fonte: A autora, 2017.

Figura 21 - “Permanecer na liderança por metade do tempo”



Fonte: A autora, 2017.

correspondem a “permanecer na liderança por metade do tempo”. Observe que a diferença entre eles é a quantidade de trocas de liderança. O número de vezes que a poligonal cruzar o eixo x é a quantidade de mudanças na liderança.

Não importa o quão longa seja a série de jogadas, o número mais provável de mudança de liderança é zero; exatamente uma troca é mais provável do que duas, duas são mais prováveis do que três e assim sucessivamente. (FELLER, 1957, p. 68)³¹

Há uma mudança de sinal na época t se s_{t-1} e s_{t+1} tem sinais opostos. Isso requer que $s_t = 0$. Conseqüentemente, pelo Corolário 1, t deve ser par. Precisaremos de alguns lemas para estabelecer a probabilidade de trocas de liderança.

Considere $t = 2m + 1$ e

$$\mathcal{U}_c = \{s, s_1 = 1 \text{ e } s \text{ troca de sinal exatamente } c \text{ vezes até } t\} \quad (26)$$

$$\mathcal{F}_c = \{s, s_t = 2c + 1\}. \quad (27)$$

Lema 5. *Para todo ímpar $2m+1$, existe uma correspondência um a um entre os conjuntos \mathcal{U}_0 e \mathcal{F}_0 .*

Prova. Seja $s \in \mathcal{U}_0$. Seja s' o caminho correspondente ao trecho de s a partir de $(1, 1)$. Ao considerarmos $(1, 1)$ como a origem de um novo sistema de coordenadas, não cruzar o

³¹ No matter how long the series of tossings, the most probable number of changes of lead is zero; exactly one change of lead is more probable than two, two changes are more probable than three, etc.

eixo horizontal no antigo sistema de coordenadas corresponde a não cruzar a reta $y = -1$. Note que há uma correspondência um a um entre os caminhos em \mathcal{U}_0 e os caminhos de comprimento $2m$ que não cruzam a reta $y = -1$.

Dizer que a reta $y = -1$ nunca foi cruzada é o mesmo que analisar a situação em que a reta $y = -2$ nunca foi tocada (ou cruzada). De fato, note que para cruzar a reta $y = -1$, é preciso que na época anterior ou posterior a reta $y = -2$ tenha sido tocada ou cruzada. Com $2m$ épocas, pelo Corolário 1, o caminho s' termina necessariamente em um ponto de ordenada par: $(2m, 2k)$, $k \in \mathbb{Z}$. Como o caminho nunca toca (ou cruza) a reta $y = -2$, s'_{2m} é um inteiro par e não-negativo. Isto é, os caminhos considerados devem terminar em pontos da forma $(2m, 2k)$ com $k \geq 0$.

Vamos contar quantos elementos há em \mathcal{U}_0 da seguinte maneira: para cada $k \geq 0$,

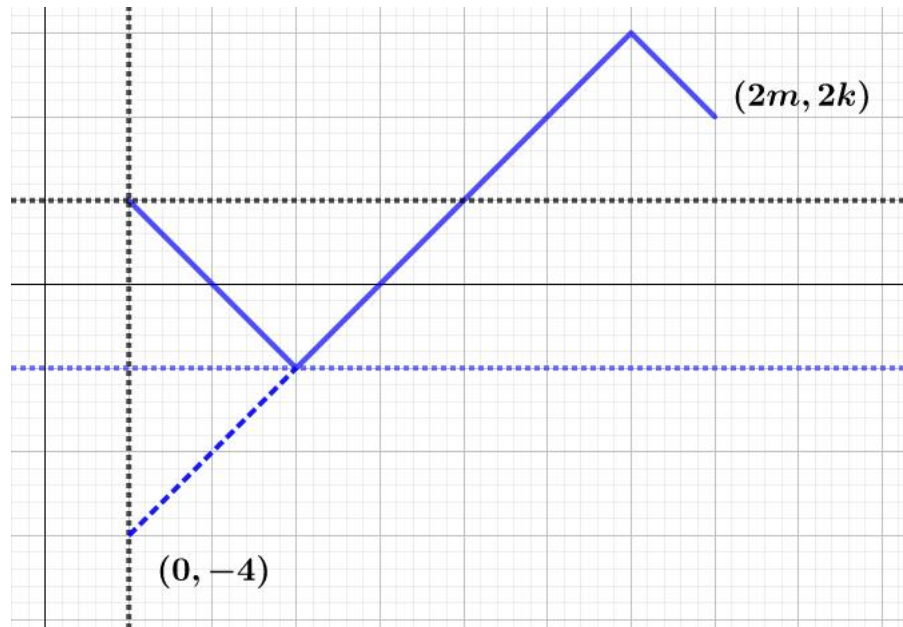
- vamos contar quantos caminhos de comprimento $2m$ partem da origem e chegam ao ponto $(2m, 2k)$;
- vamos contar quantos caminhos de comprimento $2m$ partem da origem, chegam ao ponto $(2m, 2k)$ e tocam a reta $y = -2$;
- a diferença entre os dois valores obtidos corresponde à quantidade de caminhos em \mathcal{U}_0 que terminam em $(2m, 2k)$
- para obter o número de elementos em \mathcal{U}_0 vamos somar em k .

Pela Proposição 2, há $N_{2m, 2k}$ caminhos que partem da origem e chegam ao ponto $(2m, 2k)$. A quantidade de caminhos que partem de $(0, 0)$ e chegam a $(2m, 2k)$ inclui os caminhos que tocam ou cruzam a reta $y = -2$. Pelo princípio da reflexão (Lema 1), a quantidade de caminhos que tocam ou cruzam a reta $y = -2$ coincide com a quantidade de caminhos que partem de $(0, -4)$ e chegam a $(2m, 2k)$ com $k \geq 0$. Que por sua vez, coincide com a quantidade de caminhos que partem de $(0, 0)$ e chegam a $(2m, 2k + 4)$. Para obter os caminhos que pertencem a \mathcal{U}_0 , devemos retirar de todos os caminhos de comprimento $2m$ aqueles que tocam ou passam pela reta $y = -2$. Logo, há $N_{2m, 2k} - N_{2m, 2k+4}$ caminhos em \mathcal{U}_0 que terminam em $(2m, 2k)$, com $k \geq 0$. Portanto

$$\begin{aligned} |\mathcal{U}_0| &= \sum_{k=0}^{\infty} (N_{2m, 2k} - N_{2m, 2k+4}) = \sum_{k=0}^m (N_{2m, 2k} - N_{2m, 2k+4}) \\ &= N_{2m, 0} + N_{2m, 2} + N_{2m, 4} + \cdots + N_{2m, 2m} - N_{2m, 4} - N_{2m, 6} - N_{2m, 8} - \cdots - N_{2m, 2m} \\ &= N_{2m, 0} + N_{2m, 2} \end{aligned}$$

Pela Proposição 2, temos $|\mathcal{F}_0| = N_{2m+1, 1}$. Pela relação de Stifel:

$$|\mathcal{U}_0| = N_{2m, 0} + N_{2m, 2} = \binom{2m}{m} + \binom{2m}{m-1} = \binom{2m+1}{m} = N_{2m+1, 1} = |\mathcal{F}_0|.$$

Figura 22 - Princípio da reflexão, $y = -2$ 

Fonte: A autora, 2017.

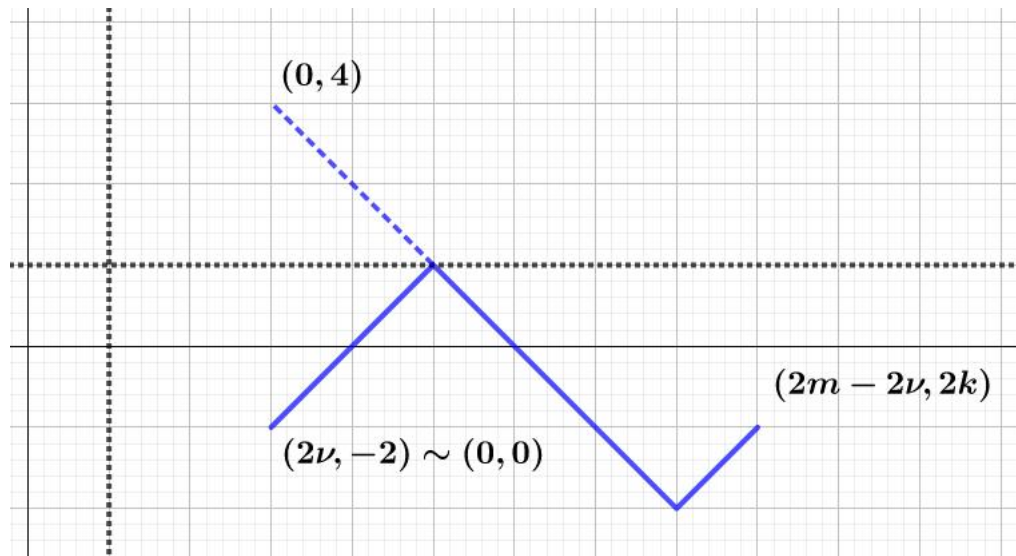
□

Lema 6. Para todo ímpar $2m+1$, existe uma correspondência um a um entre os conjuntos \mathcal{U}_1 e \mathcal{F}_1 .

Prova. Para determinar a quantidade de elementos de \mathcal{U}_1 , vamos considerar $(1, 1)$ como nova origem e analisar caminhos de comprimento $2m$ que cruzem $y = -1$ apenas uma vez. Como há apenas um cruzamento da reta $y = -1$ e o caminho parte da origem, a primeira ordenada após o cruzamento só pode ser negativa. Além disso, pelo Corolário 1, como a ordenada no cruzamento é ímpar, a abscissa também deve ser ímpar. Um caminho que cruza $y = -1$ na época $2\nu - 1$ pode ser decomposto em dois caminhos. Um que vai de $(0, 0)$ a $(2\nu, -2)$ e outro, de comprimento $2m - 2\nu$ que parte de $(2\nu, -2)$. Como $c = 1$, no caminho de comprimento $2m - 2\nu$ não há novos cruzamentos com a reta $y = -1$. Para contar os caminhos que partem de $(2\nu, -2)$, vamos argumentar como no caso $c = 0$.

Consideramos $(2\nu, -2)$ como nova origem. Queremos calcular quantos caminhos de comprimento $2m - 2\nu$ não cruzam com a reta $y = 1$. Isto é, não tocam nem cruzam a reta $y = 2$. Cada um destes caminhos termina em um ponto da forma $(2m - 2\nu, 2k)$ com $k \leq 0$. Para cada $k \leq 0$,

- vamos contar quantos caminhos de comprimento $2m - 2\nu$ partem da origem e chegam ao ponto $(2m - 2\nu, 2k)$;
- vamos contar quantos caminhos de comprimento $2m - 2\nu$ partem da origem, chegam ao ponto $(2m - 2\nu, 2k)$ e tocam a reta $y = 2$.

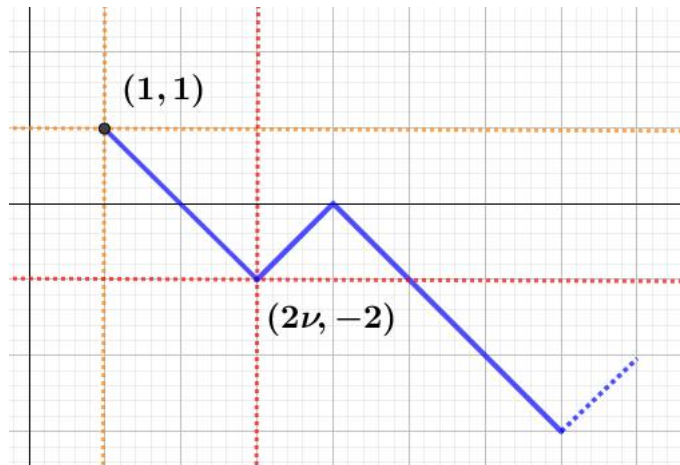
Figura 23 - Princípio da reflexão, $y = 2$ 

Fonte: A autora, 2017.

Pela Proposição 2, há $N_{2m-2\nu, 2k}$ caminhos da origem e chegam ao ponto $(2m - 2\nu, 2k)$. A quantidade de caminhos que partem da origem e chegam a $(2m - 2\nu, 2k)$ inclui os caminhos que tocam ou cruzam a reta $y = 2$. Pelo princípio da reflexão (Lema 1), a quantidade de caminhos que parte da origem, termina em $(2m - 2\nu, 2k)$ e toca a reta $y = 2$, coincide com a quantidade de caminhos que parte de $(0, 4)$ e termina em $(2m - 2\nu, 2k)$. Que por sua vez, coincide com a quantidade de caminhos que parte da origem e termina em $(2m - 2\nu, 2k - 4)$. Tal quantidade é dada por $N_{2m-2\nu, 2k-4}$. Para calcular quantos caminhos de comprimento $2m - 2\nu$ não tocam nem cruzam a reta $y = 2$, fazemos

$$\begin{aligned}
 & \sum_{k=0}^{-\infty} (N_{2m-2\nu, 2k} - N_{2m-2\nu, 2k-4}) \\
 &= \sum_{k=0}^{-m+\nu} (N_{2m-2\nu, 2k} - N_{2m-2\nu, 2k-4}) \\
 &= N_{2m-2\nu, 0} + N_{2m-2\nu, -2} = \binom{2m-2\nu}{m-\nu} + \binom{2m-2\nu}{m-\nu+1} \\
 &= \binom{2m-2\nu+1}{m-\nu+1} = N_{2m-2\nu+1, -1}
 \end{aligned}$$

Isto é, a quantidade de caminhos de comprimento $2m - 2\nu$ que não tocam nem cruzam a reta $y = 2$ coincide com a de caminhos que partem da origem e terminam em $(2m - 2\nu + 1, -1)$. Portanto, podemos estabelecer uma correspondência um a um entre caminhos de comprimento $2m - 2\nu$ que não tocam nem cruzam a reta $y = 2$ e caminhos que partem da origem e terminam em $(2m - 2\nu + 1, -1)$. Voltando ao sistema de coordenadas anterior (ver Figura 24), podemos estabelecer uma correspondência um a um entre caminhos de com

Figura 24 - Sistema de coordenadas anterior: origem em $(1, 1)$ 

Fonte: A autora, 2017.

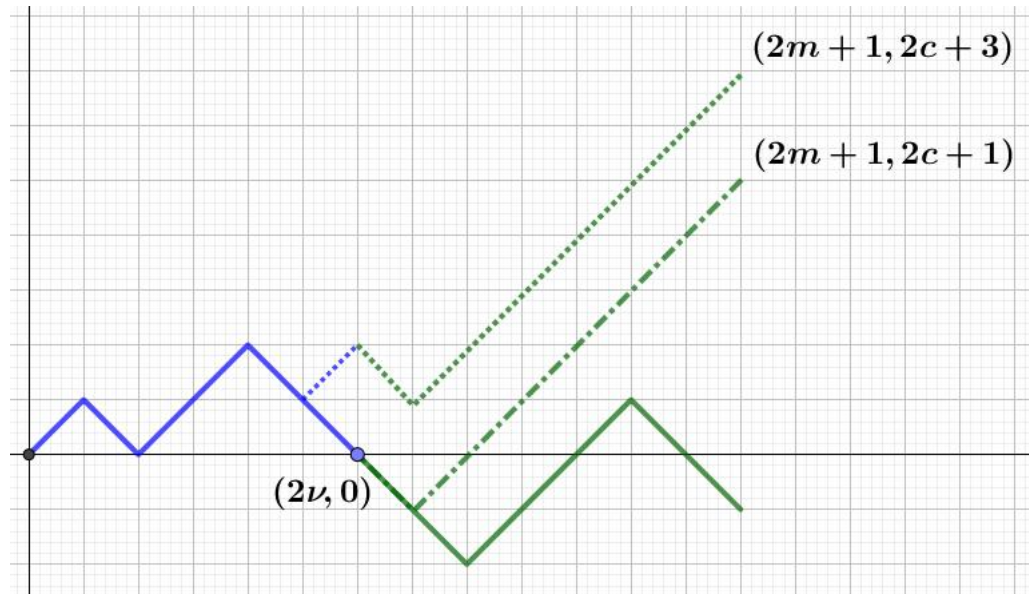
$2m - 2\nu$ que partem de $(2\nu, -2)$ e não tocam nem cruzam a reta $y = 0$ e caminhos que partem de $(2\nu, -2)$ e terminam em $(2m + 1, -3)$. Cada caminho que parte de $(2\nu, -2)$ e termina em $(2m + 1, -3)$ pode ser combinado com a porção inicial que vai de $(0, 0)$ a $(2\nu, -2)$ para formar um caminho da origem até $(2m + 1, -3)$. A reflexão em torno do eixo horizontal desses caminhos gera todos os elementos de \mathcal{F}_1 . Logo, $|\mathcal{U}_1| = |\mathcal{F}_1|$. \square

Lema 7. Para todo ímpar $t = 2m + 1$, existe uma correspondência um a um entre os conjuntos \mathcal{U}_c e \mathcal{F}_c .

Prova. Faremos a prova por indução em c . Os casos $c = 0$ e $c = 1$ já foram provados nos Lemas 5 e 6. Admita que $|\mathcal{U}_c| = |\mathcal{F}_c|$. Seja $s \in \mathcal{U}_{c+1}$. Seja $\bar{t} \geq 1$ a primeira época em que s cruza o eixo x . Sabemos que \bar{t} é par. Isto é, $\bar{t} = 2\nu$. Além disso, como $s_1 = 1$, sabemos que $s_{\bar{t}+1} = -1$. Observe ainda que no trecho s' de s que vai de $(0, 0)$ a $(2\nu, 0)$, todas as ordenadas são não-negativas. O trecho s'' de s que vai de $(2\nu, 0)$ a $(2m + 1, s_t)$ cruza o eixo x exatamente c vezes e $s''_1 = -1$. Seja s''' a reflexão de s'' em relação ao eixo x . Observe que s''' cruza o eixo x exatamente c vezes e $s'''_1 = 1$. Pela hipótese de indução, existe uma bijeção F definida de \mathcal{U}_c em \mathcal{F}_c . Considerando a origem em $(2\nu, 0)$, s''' pode ser visto como um elemento de \mathcal{U}_c . Seja $v''' = F(s''')$. Temos $v'''_{2m-2\nu+1} = 2c + 1$. Vimos no item (c) do Teorema 2 que podemos associar a s' um caminho v' de comprimento 2ν tal que $v'_{2\nu} = 2$. O caminho obtido pela junção de v' e v''' (transladado) pertence a \mathcal{F}_{c+1} (ver Figura 25). O processo descrito é inversível e concluímos que $|\mathcal{U}_{c+1}| = |\mathcal{F}_{c+1}|$. \square

Teorema 4. A probabilidade $\xi_{c,2m+1}$ de até a época $2m + 1$ ocorrerem exatamente c trocas de liderança é igual a $2p_{2m+1,2c+1}$. Isto é

$$\xi_{c,2m+1} = 2P\{S_{2m+1} = 2c + 1\}, \quad c = 0, 1, \dots \quad (28)$$

Figura 25 - Indução em c 

Fonte: A autora, 2017.

Prova. Seja $t = 2m + 1$. Considere

$\mathcal{C}_{t,c} = \{\text{caminhos que realizam exatamente } c \text{ trocas de sinais até a época } t\}$.

Vamos considerar também os conjuntos

$$\mathcal{C}_{t,c,1} = \mathcal{C}_{t,c} \cap \{s, s_1 = 1\} \quad \text{e} \quad \mathcal{C}_{t,c,-1} = \mathcal{C}_{t,c} \cap \{s, s_1 = -1\}$$

Note que $\mathcal{C}_{t,c,1}$ e $\mathcal{C}_{t,c,-1}$ são disjuntos e, por simetria, têm a mesma quantidade de elementos e $\mathcal{C}_{t,c} = \mathcal{C}_{t,c,1} \dot{\cup} \mathcal{C}_{t,c,-1}$.

Agora

$$\begin{aligned} \xi_{c,2m+1} &= \frac{|\mathcal{C}_{t,c}|}{2^{2m+1}} = \frac{|\mathcal{C}_{t,c,1}| + |\mathcal{C}_{t,c,-1}|}{2^{2m+1}} = \frac{1}{2} \frac{|\mathcal{C}_{t,c,1}|}{2^m} + \frac{1}{2} \frac{|\mathcal{C}_{t,c,-1}|}{2^m} \\ &= \frac{1}{2} \frac{|\mathcal{C}_{t,c,1}|}{2^m} + \frac{1}{2} \frac{|\mathcal{C}_{t,c,1}|}{2^m} = \frac{|\mathcal{C}_{t,c,1}|}{2^m} \\ &= P(\{c \text{ trocas de sinal em } 2m + 1 \text{ épocas dado que } S_1 = 1\}) \end{aligned}$$

Observe que $\mathcal{C}_{t,c,1}$ coincide com \mathcal{U}_c . Note que para calcular

$$P(\{c \text{ trocas de sinal em } 2m + 1 \text{ épocas dado que } S_1 = 1\}),$$

devemos dividir a quantidade de elementos de \mathcal{U}_c pela quantidade de caminhos de comprimento $2m + 1$ tais que $s_1 = 1$. Como tais caminhos passam por $(1, 1)$, há 2^{2m} caminhos

deste tipo. Pelo Lema 7

$$\xi_{c,2m+1} = \frac{|\mathcal{U}_c|}{2^m} = \frac{|\mathcal{F}_c|}{2^m} = 2 \cdot \frac{|\mathcal{F}_c|}{2^{m+1}} = 2P(S_{2m+1} = 2c + 1) = 2p_{2m+1,2c+1}$$

□

É importante lembrar que se estivermos interessados em controlar por quanto tempo um dos episódios é líder, mas também ter controle sobre troca de liderança será necessário combinar alguns resultados. Concluimos esta seção calculando a probabilidade P_e de cada episódio permanecer por metade do tempo na liderança e haver apenas exatamente uma troca de liderança quando consideramos 20.000 espectadores.

Calculamos anteriormente a probabilidade de um dos episódios permanecer metade do tempo na liderança

$$\alpha_{10.000,10.000} = \frac{(10.000!)^2}{(5.000!)^4 \cdot 2^{20.000}}$$

Pelo Teorema 4, a probabilidade³² de uma troca de liderança é igual a

$$\begin{aligned} \xi_{1,20.001} &= \frac{1}{2^{20.000}} \binom{20.001}{10.001} = \frac{1}{2^{20.000}} \cdot \frac{20.001!}{10.001! \cdot 10.000!} = \frac{20.001!}{10.001} \cdot \left(\frac{5.000!}{10.000!} \right)^4 \cdot \alpha_{10.000,10.000} \\ &= \frac{3.638594561 \cdot 10^{77341}}{10.001} \cdot \left(\frac{4.228577926 \cdot 10^{16.325}}{2.84625968 \cdot 10^{35.659}} \right)^4 \cdot \alpha_{10.000,10.000} \\ &= \frac{3.638594561 \cdot 10^5}{10.001} \cdot \left(\frac{4.228577926}{2.84625968} \right)^4 \cdot \alpha_{10.000,10.000} = 7,4681 \cdot \alpha_{10.000,10.000} \end{aligned}$$

Para calcular P_e , a troca de liderança deve acontecer na época 10.000. Devemos considerar dois casos: o episódio A lidera até a época 10.000 e o episódio B lidera até a época 20.000 e o episódio B lidera até a época 10.000 e o episódio A lidera até a época 20.000. Por simetria,

$$P_e = 2 \cdot \frac{|\{s_1 \geq 0, s_2 \geq 0, \dots, s_{9.999} \geq 0, s_{10.000} = 0\}|^2}{2^{20.000}}$$

Pelo item (b) do Teorema 2

$$\begin{aligned} &= 2 \cdot \frac{(L_{10.000})^2}{2^{20.000}} = \frac{2}{(5.001)^2 \cdot 2^{20.000}} \cdot \left(\frac{10.000}{5.000} \right)^2 = \frac{2}{(5.001)^2} \cdot \frac{(10.000!)^2}{(5.000!)^4 \cdot 2^{20.000}} \\ &= \frac{2}{(5.001)^2} \cdot \alpha_{10.000,10.000} \end{aligned}$$

³² Cálculos efetuados em <<https://www.calculatorsoup.com/calculators/discretemathematics/factorials.php>> e <<https://octave-online.net/>>.

CONCLUSÃO

O livro *O Andar do Bêbado* se dispõe a aproximar o público em geral da Matemática. Principalmente da probabilidade, assunto evitado por muitos. É rico em exemplos que não são encontrados em livros usuais sobre o tema. Suas comparações extrapolam os simples problemas de lançamentos de moedas. No entanto, alguns exemplos usam a estratégia das moedas com uma nova roupagem. Os exemplos apresentados no capítulo de Combinatória sobre executivos ou desempenho de alunos em provas chamam a atenção para situações em que a primeira intuição ilude. Faz com que as pessoas considerem como probabilidades raras ou pouco frequentes. No entanto, ao calculá-las, surpreendem e contrariam a primeira impressão. Por outro lado, o leitor não se pode deixar levar completamente por eles. Estes e muitos outros exemplos do livro deixam a sensação de que várias instâncias da vida cotidiana estão de tal modo reguladas pelo acaso que pouco controle ou interferências as pessoas têm sobre elas. Vale lembrar que há hipóteses feitas pelo autor que orientam o leitor para um caminho específico. Considerar que “todos os alunos estão despreparados e respondem às questões chutando as respostas” artificializam as situações descritas e relativizam o real significados das probabilidades encontradas. Da mesma forma, pode-se analisar o exemplo dos administradores de fundos que jogaram 15 moedas, uma por ano. Ao se considerar o evento saírem apenas caras, pode-se parecer algo pouco provável. Isso porque se considera que o experimento ainda não foi começado. No entanto, ao se analisar o mesmo evento em um contexto de probabilidade condicional, isso muda. Se a pergunta for: qual a probabilidade de saírem 15 caras, uma vez que já saíram 14, a probabilidade aumenta para 50%. De qualquer modo, é ainda interessante trazer o acaso para o nosso cotidiano. O entendimento da complexidade de nossa vida pode ser beneficiado se também a olharmos pelas lentes do acaso.

A obra é de fácil leitura em relação à dinâmica da escrita, mas em muitos exemplos o cálculo foi omitido. Mlodinow apenas apresenta respostas, dificultando o entendimento até mesmo de quem é da área de exatas. Ao tentar atingir um público amplo com exemplos “cotidianos” e pouca presença de termos técnicos e conceitos precisos, ele não deu boas condições para um leitor menos familiarizado com a teoria avançar e aprofundar seu conhecimento. Além disso, dificultou o entendimento preciso por parte de um público mais capacitado para compreender os detalhes do argumento. É o caso citado no capítulo Probabilidade Conjunta e Condicional sobre o policial divorciado e morto no mesmo ano. Aparentemente, Mlodinow deixa a pista da correlação negativa entre os eventos com a frase sobre a impossibilidade de se divorciar depois de morto. No entanto, ela foi introduzida no texto quase como uma piada e desviou a atenção do leitor para o seu papel “técnico”. Neste caso, especialmente, uma pequena nota de rodapé citando a correlação teria nos poupado bastante tempo e também indicado um caminho a ser seguido por um leitor

leigo mais interessado. Como fomentador de debate, o livro é rico em exemplos originais. Alguns aparentemente simples, mas que carregam conceitos matemáticos densos.

Além disso, esta pesquisa mostra que determinados conceitos usuais na Educação Básica não são consenso ou são pouco explorados. Definir espaço amostral como o conjunto de elementos equiprováveis não pode ser a primeira abordagem. Pode-se começar a explicar a probabilidade pelo conceito de Espaço Amostral. Só não é possível já utilizar a definição equiprovável para isso. Essa dissertação pôde concluir que o conceito de espaço amostral não é tão explorado. Um mesmo experimento pode exibir conjuntos de espaço amostral diferente. Não há um único, mas é preciso analisar a tática utilizada para calcular a probabilidade procurada. Espaços amostrais não equiprováveis não são muito abordados pelos livros didáticos. A estratégia utilizada aqui foi associá-los a conjuntos equiprováveis. Percebe-se, em muitos momentos, que a probabilidade não pode ser um conhecimento aplicado de forma mecânica. Quem aborda um problema precisa vivenciá-lo, se colocar na situação apresentada.

Outro fator existente é como as pessoas têm dificuldade em diferenciar qual situação seria mais provável. Pode-se verificar isso nos exemplos da seção Menos é Mais. Quanto mais informações eram apresentadas, mais prováveis se tornavam aos olhos dos participantes. Dois eventos associados pareciam ter mais chance de ocorrer do que apenas um isolado.

Além disso, Mlodinow exibe um curioso problema sobre lâmpadas. Considere uma máquina que acenda lâmpadas de duas cores, verde e vermelha, uma de cada vez. A informação de que a sequência de cores acendidas é aleatória passa despercebida por muitos. Em geral, as pessoas procuram um padrão. A princípio, pode-se pensar em que elas não acreditam em como uma máquina possa funcionar de forma aleatória. Pensou-se em utilizar uma caixa com bolas de mesmo peso e tamanho, mas cores diferentes para simular o acendimento das luzes. As bolas seriam sorteadas e depois repostas à caixa. Haveria a mesma quantidade de cada cor. No entanto, esse método não foi testado. Não foi possível verificar se o desempenho das pessoas seria o mesmo.

Ainda convém lembrar que o capítulo O Andar do Bêbado foi o mais desafiador desta dissertação. Mais uma vez se recorreu ao lançamento de moedas, porém diversos conteúdos foram aplicados para resolvê-lo. O problema poderia ter sido apresentado apenas como um simples lançamento de moedas. No entanto, Mlodinow preferiu investir em uma linguagem atual muito interessante. Fez a comparação a uma disputa por audiência entre filmes de George Lucas. Ao se pensar na analogia das moedas, pode-se entender que é um problema simples. No entanto, pôde-se verificar que não é tão simples contar o número de vezes em um dos filme está na liderança. Mlodinow não citou exatamente o exemplo de Feller. Feller afirma que é mais provável não haver troca nenhuma, em relação a ocorrer uma troca, esta em relação a duas, e assim por diante. Outro foco de confusão foi Mlodinow não precisar o que seria troca de liderança. Feller explica que é uma questão

de unidades de tempo, ou seja, o caminho percorrido pode ser de votos alternados.

Finalizamos este trabalho com algumas das lacunas deixadas por Mlodinow preenchidas. Tentamos facilitar o entendimento de seu exemplos. Sem sombra de dúvida, pusemos à prova nossa intuição diante do acaso.

REFERÊNCIAS

- AZEVEDO, C. O que é a probabilidade? interpretações da probabilidade. *DMAT – Comunicações*, p. 1–14, 2004.
- BORDER, K. C. *Simple Random Walk*. 2017. Disponível em: <<http://www.math.caltech.edu/~2016-17/2term/ma003/Notes/Lecture16.pdf>>. Acesso em: 14 de agosto de 2018.
- BOYER, C. B.; MERZBACH, U. C. *História da Matemática*. [S.l.]: Blucher, 2012.
- CARVALHO, P. C. P. et al. *Análise Combinatória e Probabilidade*. [S.l.]: SBM, 2016.
- CORRY, L. David hilbert and the axiomatization of physics (1894-1905). *Archive for History of Exact Sciences*, v. 2, n. 51, p. 83–198, 1997.
- FELLER, W. *An Introduction to Probability Theory and its Applications*. [S.l.]: John Wiley & Sons Inc., 1957.
- FELLER, W. *An Introduction to Probability Theory and its Applications*. [S.l.]: John Wiley & Sons Inc., 1968.
- FONSECA, J. da; MARTINS, G. de A. *Curso de estatística*. [S.l.]: Atlas, 1996.
- FOSSA, J. *Introdução às técnicas de demonstração na Matemática*. [S.l.: s.n.], 2009. (Contextos da ciência).
- GAISSMAIER, W.; SCHOOLER, L. J. The smart potential behind probability matching. *Cognition*, v. 109, n. 3, p. 416–422, 2008.
- GAVA, A. M.; VITIELLO, L. R. de S. Inflation, quarterly financial statements and fraud: Benford’s law and the brazilian case. In: *Anais do XXXI Encontro da ANPAD*. Rio de Janeiro, Brasil: [s.n.], 2007. p. 1–13.
- GORROOCHURN, P. *Classic Topics on the History of Modern Mathematical Statistics: From Laplace to More Recent Times*. [S.l.]: Wiley, 2016.
- HILL, T. P. The significant-digit phenomenon. *The American Mathematical Monthly*, AMS, v. 102, n. 4, p. 322–327, 1995.
- HOWIE, D. *Interpreting Probability Controversies and Developments in the Early Twentieth Century*. [S.l.]: Cambridge University Press, 2002.
- IRETON-JONES, C. S.; GOTTSCHLICH, M.; BELL, S. *Practice-oriented Nutrition Research: An Outcomes Measurement Approach*. [S.l.]: Aspen Publishers, 1998. (An Aspen publication).
- JEFFREYS, H. *The Theory of Probability*. Oxford: OUP, 1950.
- KALBFLEISCH, J. G. *Probability and Statistical Inference: Probability. vol. 1*. [S.l.]: Springer, 1985. (Springer Texts in Statistics).

- KOEHLER, D. J.; JAMES, G. Probability matching, fast and slow. *Psychology of Learning and Motivation*, v. 61, p. 103–131, 2014.
- KOLMOGOROV, A. N. *Foundations of the Theory of Probability*. [S.l.]: Chelsea Publishing Company, 1956.
- LIMA, E. L. *Análise Real*. 12. ed. [S.l.]: IMPA, 2016. v. 1.
- MEYER, P. *Probabilidade: aplicações à estatística*. [S.l.]: LTC, 2006.
- MLODINOW, L. *O andar do bêbado: Como o acaso determina nossas vidas*. [S.l.]: Zahar, 2009.
- NARENS, L. *Theories of Probability: an Examination of Logical and Qualitative Foundations*. [S.l.]: World Scientific, 2007.
- NEWELL, B. R.; RAKOW, T. The role of experience in decisions from description. *Psychonomic Bulletin & Review*, n. 14, p. 1133–1139, 2007.
- NEWELL, B. R.; SCHULZE, C. Probability matching. In: _____. *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory*. [S.l.]: Routledge/Taylor & Francis Group, 2017. cap. 5, p. 62–78.
- ROUSSEAU, C. Apprendre à frauder ou à détecter les fraudes? *Accromath*, v. 5, p. 2–7, 2010.
- SINGH, J. Theories of probability. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, Springer, v. 7, n. 3, p. 257–262, 1946.
- STEYER, R.; NAGEL, W. *Probability and Conditional Expectation: Fundamentals for the Empirical Sciences*. [S.l.]: Wiley, 2017. (Wiley Series in Probability).
- TUNALA, N. Determinação de probabilidades por métodos geométricos. *Revista do Professor de Matemática*, v. 20, p. 16–22, 1992.