IV Simpósio Nacional da Formação do Professor de Matemática

PROBABILIDADE ALÉM DA COMBINATÓRIA

TÓPICOS E PROBLEMAS REAIS COM FOCO NO RACIOCÍNIO PROBABILÍSTICO

Vitor Amorim
Graziele Mozer

ANPMat
Associação Nacional dos Professores de Matemática na Educação Básica

Probabilidade além da combinatória

Tópicos e problemas reais com foco no raciocínio probabilístico

Probabilidade além da combinatória - Tópicos e problemas reais com foco no raciocínio probabilístico

Capa: Pablo Diego Regino
Projeto gráfico: Cinthya Maria Schneider Meneghetti

Distribuição
Associação Nacional dos Professores de Matemática na Educação Básica
https://www.anpmat.org.br / email: secretaria@anpmat.org.br

ISBN 978-65-88013-10-6

IV Simpósio Nacional
da Formação do Professor de Matemática

PROBABILIDADE ALÉM DA COMBINATÓRIA

TÓPICOS E PROBLEMAS REAIS COM FOCO NO RACIOCÍNIO PROBABILÍSTICO

Vitor Amorim
Graziele Mozer

1ª edição
2020
Rio de Janeiro

ANPMat

Agradecimentos

Agradecemos à Associoção Nacional dos Professores de Matemática na Educação Básica (ANPMat) pela oportunidade e por todo o apoio necessário para o desenvolvimento deste livro e a apresentação do minicurso no IV Simpósio Nacional da Formação do Professor de Matemática.

Agradecemos também a atenção, apoio e carinho no acolhimento dos professores e alunos da UFES e do IFES que participaram da organização do evento.

Agradecemos ainda ao amigo Cezar Teixeira pelas valiosas sugestões de exemplos e situações-problema do mundo real envolvendo conceitos probabilísticos que nos inspiraram na elaboração de vários tópicos deste material.

Por fim, agradecemos especialmente ao nosso grupo dos *amigos de simpósios*, professores que sempre participam dos Simpósios da Formação do Professor de Matemática, dispostos a compartilhar seus conhecimentos enriquecendo a formação dos companheiros de profissão e contribuindo para o desenvolvimento do ensino da Matemática no Brasil.

Prefácio

Este livro foi desenvolvido como texto de base para minicurso *Probabilidade Além da Combinatória: Tópicos e Problemas Reais com Foco no Raciocínio Probabilístico*, ministrado no IV Simpósio Nacional da Formação do Professor de Matemática, realizado em Vitória-ES, em novembro de 2019.

O objetivo central deste texto é apresentar uma discussão sobre a abordagem tradicional do tema Probabilidade no ensino básico, apontando necessidades de renovação e, em seguida, apresentar e discutir tópicos, ferramentas e problemas de Probabilidade focados em aplicações reais, que sejam, em geral, independentes das técnicas de contagem e que permitam desenvolver no aluno o raciocínio probabilístico.

A prática tradicional do ensino de Probabilidade e sua abordagem em textos didáticos frequentemente têm como foco problemas probabilísticos cuja principal competência exigida do aluno é a aplicação de técnicas de contagem estudadas previamente em Análise Combinatória. A despeito da importância que essa competência tem, bem como sua aplicação em Probabilidade, o ensino desse tema focado em problemas combinatórios desenvolve nos alunos a falsa ideia de dependência e até mesmo de indissociabilidade entre Probabilidade e Análise Combinatória.

O raciocínio probabilístico e o raciocínio combinatório são independentes e devem ser relacionados, quando adequado, e distinguidos, quando necessário. Além disso, dadas as dificuldades intrínsecas aos problemas de contagem, o foco excessivo em problemas combinatórios pode criar maiores dificuldades e uma desnecessária aversão do aluno à unidade temática de Probabilidade e Estatística, que ganhou destaque recente na Base Nacional Comum Curricular (BNCC). Por fim, tal prática acaba deixando em segundo plano aspectos fundamentais do raciocínio probabilístico, que têm consequências diretas em situações cotidianas e aplicações em diversos contextos da Ciência.

Dessa forma, este livro tem como objetivo dar destaque a algumas competências probabilísticas fundamentais, através da discussão e propostas de inovação da abordagem tradicional e, principalmente, da apresentação de tópicos e problemas de probabilidade aplicados a situações reais, cuja modelagem, em geral, não depende das técnicas de contagem, mas sim de uma compreensão significativa dos conceitos elementares da Teoria das Probabilidades.

Entre esses conceitos destacaremos a abordagem frequentista e a lei dos grandes números, as possíveis definições axiomáticas, as propriedades da probabilidade e aplicações, probabilidade condicional e independência, teoremas do produto, da probabilidade total e de Bayes, e a distribuição binomial de probabilidades.

Apresentaremos também algumas possibilidades de inovação no tratamento de Probabilidades no ensino médio, abordando três tópicos da área que não costumam ser trabalhados nesse segmento, mas que têm grande potencial de contextualização, foco interdisciplinar e integração com outras áreas da Matemática: a distribuição geométrica de probabilidades, as cadeias de Markov e a probabilidade geométrica.

Em relação à distribuição geométrica, veremos que muitos problemas relevantes e de fácil compreensão podem ser modelados por uma distribuição de probabilidades dispostas em progressão geométrica, uma oportunidade para contextualização e aplicação desse conceito. Por outro lado, uma abordagem introdutória ao conceito de cadeia de Markov reforça habilidades probabilísticas fundamentais, como partição de um evento em um diagrama de árvore, e traz uma possibilidade de aplicações de sistemas lienares e de multiplicação e potências de matrizes, conceitos geralmente estudados sem aplicação no ensino médio. Por fim, problemas clássicos da área conhecida como probabilidade geométrica trazem uma visão diferente da probabilidade e integram conceitos de duas áreas aparentemente desconexas, a Geometria e a Probabilidade.

Finalmente, este livro não tem como objetivo servir de livro-texto para alunos de ensino médio, apesar de conter vários exemplos e exercícios que podem ser interessantes para esse público. O alvo principal deste texto é o público de professores de Matemática que atua no ensino básico e na formação inicial e continuada de professores. Não temos ainda a pretensão de que todo o conteúdo presente aqui possa ser abordado em uma turma regular de ensino básico. Acreditamos, contudo, que os tópicos e problemas aqui apresentados possam ser destacados para o uso em sala de aula, conforme a realidade de cada escola e de cada turma. E entendemos, por outro lado, que o contexto dos itinerários formativos, inseridos na recente reforma do ensino médio, nas áreas de Matemática e Ciências da Natureza pode apresentar oportunidades de aprofundamentos propostos neste livro.

Sumário

Capítulo 1

Introdução: Por Que Probabilidade Além da Combinatória?

A prática tradicional do ensino de probabilidade na escola básica, em geral circunscrita ao ensino médio, concentra-se frequentemente no estudo de problemas e propriedades da Probabilidade restritos ao escopo dos espaços amostrais finitos e equiprováveis. Nesse tipo de espaço, é comum adotar a definição clássica de probabilidade, onde a probabilidade da ocorrência de um evento é dada pela razão entre o número de "casos favoráveis" ao evento e o número de "casos possíveis do experimento", em um sentido que tornaremos explícito mais à frente.

Nesse contexto, a contagem dos elementos do espaço amostral e do evento para o qual se deseja inferir uma probabilidade de ocorrência torna-se a ferramenta principal para a resolução de exemplos e exercícios propostos. Assim, são abundantes nos textos didáticos os problemas de probabilidade cuja principal habilidade requerida ao aluno é a aplicação das técnicas de contagem para, no final do problema, determinar a probabilidade como quociente dos dois números calculados. Em muitos casos, é solicitado ao aluno apenas a indicação dos símbolos utilizados em combinatória (combinação, permutação etc.) na fração que determina a probabilidade. Ou seja, a compreensão exigida do aluno para o problema é focada nas questões combinatórias da contagem do número de elementos de dois conjuntos e, nesses casos, não é sequer avaliado o resultado numérico da probabilidade como indicador de expectativa sobre a ocorrência de um evento.

Ainda que reconheçamos a importância do desenvolvimento do raciocínio combinatório na etapa do ensino médio, bem como sua aplicação em problemas de probabilidade, o raciocínio probabilístico e o raciocínio combinatório são independentes e devem ser desenvolvidos de diversas formas, relacionando-os, quando adequado, e distinguindo-os, quando necessário. Ao tratar majoritariamente de problemas com foco combinatório, podem-se trazer pelo menos quatro consequências não desejadas:

1) Desenvolve-se nos alunos (e em parte dos professores) a falsa ideia de dependência e até mesmo de indissociabilidade das áreas de Probabilidade e Análise Combinatória;

2) Deixa-se de desenvolver nos alunos boa parte do raciocínio probabilístico elementar que, como veremos à frente, é fundamental não só para as aplicações modernas da Teoria das Probabilidades em diversas áreas do conhecimento, como para lidar com problemas e informações cotidianas relacionadas à área;

3) Muitos problemas da área têm soluções tanto pelo caminho combinatório quanto pelo caminho probabilístico. Soluções de problemas com uso dos raciocínios implícitos nas propriedades da probabilidade, nos importantes teoremas de Bayes e da probabilidade total, no raciocínio condicional, entre outros, podem ser uma oportunidade para o professor desenvolver esses tópicos de forma simultânea com a abordagem combinatória, comparar as soluções e permitir que o aluno possa escolher a estratégia que lhe parece mais eficiente. Focar apenas em soluções combinatórias elimina as oportunidades didáticas que a abordagem por dois olhares distintos pode trazer.

4) Dadas as dificuldades intrínsecas aos problemas combinatórios, que têm elevado grau de variação de níveis de dificuldades, o foco do ensino de Probabilidade em problemas combinatórios pode trazer uma desnecessária aversão ou medo dos alunos às áreas de Probabilidade e Estatística. No caso dessa última, que guarda ainda menos dependência do raciocínio combinatório, o problema da aversão já foi constatado por diversos educadores estatísticos e relatados em documentos da área.

Compreender e aplicar à modelagem e resolução de problemas diversas técnicas do pensamento probabilístico deve ser o foco do ensino de Probabilidade no ensino básico, e restringi-lo a problemas envolvendo espaços finitos e equiprováveis, definição clássica de probabilidade e cálculos combinatórios faz com que o aluno deixe de desenvolver competências, habilidades e pensamentos críticos que são fundamentais nessa área.

De início, a própria definição clássica de probabilidade, às vezes apresentada como única, traz essa restrição. Excetuando-se alguns poucos problemas baseados em espaços finitos não equiprováveis, o estudo de probabilidade no ensino médio raramente explora conceitos e definições de probabilidade diferentes da clássica. A definição frequentista de probabilidade, por exemplo, é uma das bases das aplicações em Estatística e, por outro lado, os problemas envolvendo espaços amostrais infinitos podem ser introduzidos nessa etapa do ensino sem a necessidade de utilização de ferramentas matemáticas avançadas. Nesse sentido, as diversas definições axiomáticas de probabilidade adquirem papel importante no desenvolvimento da teoria. Ainda no campo dos espaços amostrais infinitos, podem ser estudados problemas elementares da probabilidade geométrica, um tópico de compreensão geralmente intuitiva e que gera exemplos curiosos de aplicação.

Adicionalmente, as propriedades da probabilidade e os conceitos de probabilidade condicional e independência são parte central do pensamento probabilístico e estão presentes em situações cotidianas de forma mais frequente do que é percebido pela maioria das pessoas, como no resultado de um exame médico, por exemplo. Além disso, tais conceitos têm aplicações infindáveis em diversas áreas do conhecimento científico e tecnológico.

Sendo assim, o objetivo central deste livro é apresentar tópicos, ferramentas e problemas de Probabilidade que sejam em sua maioria independentes das técnicas de contagem e que permitam desenvolver no aluno o raciocínio probabilístico "puro", com atenção especial às aplicações dessas habilidades na resolução de problemas reais envolvendo aleatoriedade. Colocando em questão as práticas tradicionais e mais frequentes em textos didáticos, este livro pretende também rediscutir as formas de definição de probabilidade juntamente com seu significado prático, as propriedades da probabilidade e suas consequências na interpretação e resolução de problemas, a forma de abordagem dos conceitos de probabilidade condicional e independência, e uma forma diferente de abordagem para distribuição binomial de probabilidade.

Por fim, pretende-se apresentar algumas possibilidades de inovação no tratamento do tema. Alguns tópicos que não costumam ser apresentados no ensino médio, mas que são adequados ao público, como distribuição geométrica de probabilidade, cadeias de Markov e probabilidade geométrica, podem incentivar os professores leitores a buscarem novas possibilidades de desenvolvimento da competência probabilística pelos alunos, por meio de aplicações em problemas modernos de probabilidade e de possibilidades de contextualização de conceitos de outras áreas da Matemática, como progressões, matrizes, geometrias euclidiana e analítica, entre outras. Mostraremos ainda como a utilização de recursos tecnológicos, como GeoGebra e planilhas eletrônicas, podem ser amplamente aproveitados no desenvolvimento desses temas.

Capítulo 2

Definição Clássica: Potenciais de Inovação e Limitações

"Esta coisa estranha chamada "probabilidade" é o preço que você paga para não ter que lidar com toda a física do lançamento de moedas."
– Ralph Teixeira e Augusto Morgado

Neste primeiro capítulo, faremos uma breve análise e discussão sobre a prática tradicional do ensino de probabilidade, guiados por textos didáticos de relevância, especialmente livros didáticos do PNLD. Analisaremos a definição dada ao conceito de probabilidade e suas limitações, o distanciamento das aplicações e a forte relação com as técnicas da Análise Combinatória nos problemas abordados, criando uma dependência não desejada entre o raciocínio probabilístico e o raciocínio combinatório.

Em particular, na primeira seção nos debruçaremos sobre a abordagem inicial de Probabilidade no ensino básico, discutindo questões sobre a natureza de um experimento aleatório, o significado teórico e prático de atribuir um número como probabilidade de um evento e as limitações da definição clássica de probabilidades. Entendemos que, em geral, questões conceituais e de significado prático são deixadas de lado para priorizar exemplos e exercícios envolvendo cálculos de probabilidades e aplicação de técnicas de contagem, sem que os resultados obtidos e o processo para obtê-los tenham significado para os alunos.

O *deficit* na compreensão desses conceitos costumam trazer consequências reais que se observam no dia a dia: afirmações de que um cálculo probabilístico estava errado porque ocorreu um evento menos provável, interpretações de que alta probabilidade de chuva significa certeza de chuva, receitas mágicas para ganhar na loteria, jogos na Mega-Sena com números metade par metade ímpar "porque são mais prováveis", entre outras.

Por outro lado, na segunda seção, chamaremos a atenção para um problema que motivou a elaboração do minicurso que deu origem a este livro. O foco excessivo do ensino de probabilidades em problemas cuja principal habilidade requerida é a aplicação de técnicas de contagem. Como já dissemos na introdução, mostraremos como essa abordagem pode prejudicar o desenvolvimento da competência probabilística básica nos alunos, trazendo uma série de consequências involuntárias não desejadas e deixando de aproveitar o tempo dedicado ao ensino de Probabilidade para o trabalho com questões mais relevantes relacionadas a tópicos normalmente pouco desenvolvidos.

2.1 Definição Clássica de Probabilidade

Tradicionalmente, introduz-se o conceito de probabilidade com uma noção prévia de experimentos aleatórios e do seu respectivo espaço amostral Ω. Introduz-se também a ideia de eventos e suas operações, para então apresentar a definição clássica de probabilidades, na qual a probabilidade de ocorrer um evento $A \subset \Omega$ é dada pela razão entre o número de "casos favoráveis" à ocorrência de A, e o número de "casos possíveis" do experimento. A seguir, faremos um apanhado de alguns desses conceitos antes de iniciar as discussões que mencionamos acima.

Experimentos Aleatórios: *São experimentos que embora repetidos inúmeras vezes sob as mesmas condições, conduzem a resultados essencialmente distintos.*

Por exemplo, "lançar a bolinha num jogo de roleta e observar o número da canaleta que ela cai" ou "escolher uma laranja no pé e verificar seu peso registrado numa balança" são experimentos aleatórios.

Experimentos Determinísticos: *Ao contrário dos aleatórios, os experimentos determinísticos se repetidos sob as mesmas condições conduzem a resultados essencialmente iguais.*

Por exemplo, "soltar um objeto em queda livre de uma determinada altura e verificar o tempo de queda" ou "medir a temperatura que determinada quantidade de água atinge ao começar o processo de ebulição" são experimentos determinísticos. Quando nos referimos a resultados essencialmente iguais, queremos dizer que eventuais diferenças nos resultados de experimentos determinísticos devem-se à imprecisão dos instrumentos de medida, interferências humanas, influências externas ao experimento, entre outras. Essas diferenças em geral são classificadas como desprezíveis.

Antes de prosseguir com a exposição dos conceitos e definições básicas de Probabilidade, colocamos aqui uma questão importante (e complexa):

Qual é o real significado de um experimento ser classificado como aleatório?

É considerado aleatório um experimento com viés, ou seja, em que alguns resultados tenham tendência maior do que outros? Por exemplo, quando um atleta profissional de basquete arremessa um lance livre, o resultado do arremesso pode ser chamado de aleatório, ainda que a habilidade e o treinamento duro do atleta faça com que ele acerte com muito mais frequência que erre?

É considerado aleatório um número "randômico" gerado pelo computador? Mesmo que seja gerado após uma sequência de cálculos determinísticos?

Essas questões são bastante profundas e, em alguns aspectos, ainda são objeto de discussão acadêmica. Foram colocadas aqui para enfatizar a importância da discussão do tema com os alunos e do desenvolvimento da compreensão do conceito de aleatoriedade, incluindo suas subjetividades e limitações.

Na referência [18], encontramos uma discussão sobre as definições que o autor chama de probabilidade determinística e subjetiva. Na primeira, um experimento é considerado aleatório se, após um número grande de repetições, não houver evidência de viés nas frequências de ocorrências dos resultados possíveis. Na segunda, um experimento é considerado aleatório se não soubermos ou não pudermos prever seus resultados. Citamos a seguir dois trechos dessa discussão:

"Na probabilidade determinística, julgamos uma amostra pelo modo como se apresenta; já na probabilidade subjetiva, julgamos uma amostra pelo modo como é produzida."

"...num mundo perfeito, o lançamento de um dado seria sempre aleatório pela primeira definição, mas não pela segunda...no mundo real e imperfeito, porém, o lançamento do dado é aleatório de acordo com a segunda definição, mas não com a primeira."

Isso ocorre porque em um mundo perfeito existiria um dado perfeito, que produziria frequências iguais de resultados para cada uma de suas faces. Mas no mundo real e imperfeito, não temos essa garantia. De fato, isso dificilmente ocorre em experimentos reais.

O professor Humberto Bortolossi trata desse assunto no artigo [2], do qual destacamos os seguintes trechos:

"...é preciso definir antes o que se entende por um número ou evento aleatório. Esta é uma questão polêmica, sendo ainda debatida por psicólogos, filósofos, matemáticos, físicos e cientistas da computação."

"...o lançamento de uma moeda também pode ser pensado como um processo determinístico (complexo, mas ainda determinístico), regido pelas leis da Física."

Por fim, citamos um trecho bem humorado do excelente primeiro capítulo da referência [20], cujo nome é "O que é Probabilidade?":

"...esta coisa estranha chamada "probabilidade" é o preço que você paga para não ter que lidar com toda a física do lançamento de moedas"

Como dissemos, esta questão sobre a natureza da aleatoriedade não é trivial e é uma discussão longa. Mas é um assunto que deve ser abordado e discutido em um nível adequado à idade e à maturidade intelectual dos alunos. Afinal, o desenvolvimento do raciocínio probabilístico começa com a compreensão da aleatoriedade e das formas matemáticas de modelar problemas aleatórios. Mais ainda, a habilidade de identificar aleatoriedade é uma das propostas para ensino fundamental na BNCC (Base Nacional Comum Curricular).

Ao longo deste livro adotaremos a definição de aleatoriedade descrita antes das citações acima e que certamente tem algum grau de subjetividade.

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

Por exemplo, para o experimento "escolher ao acaso uma peça da linha de produção de uma fábrica e verificar se há ou não defeitos", temos

$$\Omega = \{\text{perfeita}, \text{defeituosa}\}.$$

Evento: *Qualquer subconjunto de* Ω*.*

Por exemplo, para o experimento "selecionar aleatoriamente pessoas na rua e verificar seu mês de nascimento", considere o evento E: "escolher uma pessoa nascida no primeiro trimestre". Nesse caso, temos

$$\Omega = \{\text{Janeiro}, \text{Fevereiro}, \cdots, \text{Dezembro}\} \quad \text{e}$$

$$E = \{\text{Janeiro}, \text{Fevereiro}, \text{Março}\}$$

Eventos Elementares: *Cada elemento do espaço amostral de um experimento é um evento elementar ou ponto amostral de* Ω*. No exemplo anterior, cada mês é um evento elementar.*

Além dos tópicos acima também é comum abordar operações entre eventos, introduzindo ou revisando as noções de complementar, união, interseção entre outras, que também podem ser vistas como tópicos de Teoria dos Conjuntos. Por esse motivo não abordaremos esses conceitos aqui.

Definição 2.1 (Definição Clássica)**.** *Dado um experimento aleatório, considere um evento* E *do espaço amostral* Ω*, onde* Ω *é finito e equiprovável. Ou seja, assumimos que os eventos elementares têm chances iguais de acontecer. Chamamos de probabilidade de* E *ocorrer, o número:*

$$P(E) = \frac{n(E)}{n(\Omega)} = \frac{\textit{número de casos favoráveis}}{\textit{número de casos possíveis}}$$

onde $n(E)$ *e* $n(\Omega)$ *são os números de elementos de* E *e de* Ω*, respectivamente.*

A Definição Clássica é um conceito fundamental da Probabilidade e muitos problemas relevantes são modelados por ela. Os livros didáticos do ensino básico costumam dar um tratamento bastante amplo às suas propriedades e diversos tipos de problemas modelados por ela, ainda que haja uma presença desnecessariamente grande de problemas artificiais.

Por um lado, reconhecemos que alguns problemas com jogos e situações fictícias, além de seu papel histórico na Teoria das Probabilidades, têm a vantagem da fácil compreensão para fins didáticos. Contudo, tanto em situações cotidianas quanto em questões científicas e tecnológicas, há uma abundância de problemas reais que podem tornar o ensino de probabilidade mais significativo para o aluno e ainda proporcionam uma oportunidade de contextualização e de trabalho interdisciplinar no ensino básico. Nesse sentido, apresentamos abaixo dois problemas de aplicação que são modelados por essa definição.

Exemplo 2.1. *Pessoas afetadas por uma certa doença hereditária que resulta em anemia podem ser homozigóticas* (AA) *ou heterozigóticas* (Aa)*. Esses homozigóticos apresentam a forma mais grave da doença, e os heterozigóticos, a mais branda. Os indivíduos homozigóticos* (aa) *são normais. Calcule a probabilidade de um casal, sendo a mulher normal e o homem com a doença na forma branda, terem um filho:*

a) normal (evento A).

b) com a doença na forma branda (evento B).

c) com a doença na forma grave (evento C).

Nesse tipo de situação, muito frenquente em livros de Biologia do ensino médio, cada componente A ou a vem de um dos pais. Assim, o indíviduo AA, por exemplo, herdou um A do pai e outro A da mãe. E é considerado que a chance de um indivíduo herdar um A é a mesma de um herdar um a quando for possível.

Como a mãe é normal, ela é homozigótica do tipo (aa). Já o pai é heterozigótico (Aa), pois tem a doença em sua forma branda. Daí, convém a construção de uma tabela com os possíveis resultados dessa combinação.

Tabela 2.1: Problema da Hereditariedade

mãe/pai	**A**	**a**
a	Aa	aa
a	Aa	aa

Agora contando os casos favoráveis aos eventos e os casos possíveis do experimento, podemos responder os itens. Observe que são 4 casos possíveis. Em 2 desses casos (1ª coluna) o filho seria heterozigoto e apresenta a forma branda da doença, nos outros 2 casos (2ª coluna), o filho seria normal. E não há casos em que o filho apresentaria a doença na forma grave, sendo impossível que esse casal tenha um filho nessa condição. Portanto:

$$P(A) = \frac{2}{4} = \frac{1}{2} \qquad P(B) = \frac{2}{4} = \frac{1}{2} \qquad P(C) = \frac{0}{4} = 0$$

Exemplo 2.2. *No Brasil as loterias oficiais possuem várias modalidades: Lotofácil, Quina, Mega-Sena, entre outras. Qual a probabilidade de um apostador específico conseguir o prêmio máximo de cada uma, fazendo um jogo simples? Será que os prêmios máximos são inversamente proporcionais às probabilidades de ganhá-los ou existe alguma modalidade mais vantajosa?*

O enunciado fala sobre a possibilidade de os prêmios e as probabilidades serem inversamente proporcionais porque espera-se que quanto maior for o prêmio, menor será a probabilidade de ganhá-lo. Vamos verificar se essa relação de proporcionalidade realmente existe. Para isso, restringiremos o exemplo apenas às três modalidades citadas no enunciado.

Na Lotofácil o apostador deve escolher de 15 a 18 dezenas entre as 25 disponíveis. Na Quina deve-se escolher de 5 a 15 entre 80 dezenas e na Mega-Sena, de 6 a 15 entre 60 dezenas.

Os prêmios variam a cada concurso de cada modalidade, dependem da quantidade de apostas, se está acumulado ou não, se haverá divisão entre outros eventuais ganhadores, entre outros fatores. Então consultamos, no *site* da Caixa Econômica Federal, os valores que cada ganhador de maior pontuação levou em uma certa semana para tomarmos como base.

Lotofácil - $R\$ \; 838.856,32$
Quina - $R\$ \; 1.232.735,16$
Mega-Sena - $R\$ \; 34.615.569,28$

Faremos os cálculos da probabilidade de ganhar em cada modalidade considerando apenas jogos simples (aqueles em que se escolhe a menor quantidade de dezenas possível).

Sejam os eventos L: "ganhar o prêmio máximo na Lotofácil fazendo um jogo simples", Q: "ganhar o prêmio máximo na Quina fazendo um jogo simples" e M:" ganhar o prêmio máximo na Mega-Sena fazendo um jogo simples".

Na Lotofácil, como são sorteadas 15 dezenas de forma não ordenada entre as 25 dezenas diponíveis, temos:

$$n(\Omega) = C_{25,15} = \frac{25!}{15!10!} = 3.268.760.$$

E, $n(L) = 1$, pois só há um grupo de 15 dezenas que possibilita o apostador de ganhar. Assim:

$$P(L) = \frac{1}{C_{25,15}} = \frac{1}{3.268.760}.$$

Na Quina, são sorteadas 5 dezenas (também não ordenadas) entre as 80 disponíveis, então:

$$\begin{aligned} n(\Omega) &= C_{80,5} = \frac{80!}{5!75!} = 24.040.016 \\ \Longrightarrow P(Q) &= \frac{1}{C_{80,5}} = \frac{1}{24.040.016}. \end{aligned}$$

E na Mega-Sena, sorteiam-se 6 dezenas entre as 60 disponíveis, então:

$$\begin{aligned} n(\Omega) &= C_{60,6} = \frac{60!}{54!6!} = 50.063.860 \\ \Longrightarrow P(M) &= \frac{1}{C_{60,6}} = \frac{1}{50.063.860}. \end{aligned}$$

Observe que a probabilidade de ganhar na Lotofácil é aproximadamente 7 vezes maior que a de ganhar na Quina, mas o prêmio (nesses concursos que consideramos) da segunda é menos que o dobro da primeira. Se prêmio e probabilidade nessas duas loterias fossem inversamente proporcionais, o prêmio da Quina deveria ser 7 vezes o prêmio da Lotofácil. Logo, a Lotofácil é mais vantajosa que a Quina, olhando por essa perspectiva.

A probabilidade de ganhar na Lotofácil é mais de 15 vezes maior que a de ganhar na Mega-Sena, e o prêmio da segunda é mais de 40 vezes o prêmio da primeira. O esperado era que o prêmio da Mega-Sena fosse 15 vezes o da Lotofácil. Então, a Mega-Sena é mais vantajosa que a Lotofácil.

E, a probabilidade de ganhar na Quina é praticamente o dobro de ganhar na Mega-Sena. O prêmio da segunda, porém, é 28 vezes o da primeira, a expectativa é a de que seria o dobro. Portanto, a Mega-Sena é mais vantajosa que a Quina.

Então, analisando simplesmente a probabilidade de ganhar em comparação com o valor do prêmio, temos uma ordem: Mega-Sena é a mais vantajosa das três e a Lotofácil é mais vantajosa que a Quina.

Observações: Usamos o valor pago de um único concurso em cada modalidade para simplificar a resolução. Mas como dissemos, esse valor varia muito de concurso para concurso. Mesmo um dia antes do sorteio não teríamos como saber esse valor, pois depende também do número de ganhadores que vão dividir o prêmio. Como não temos essa informação, se quisermos fazer a comparação acima, podemos fazer uma estimativa do prêmio a partir do valor acumulado e do número médio de ganhadores nos últimos 100 concursos ou mais.

Uma análise mais completa para esse problema seria utilizar o conceito de Valor Esperado abordado na Seção 6.2. Mas entendemos que a análise feita acima é bastante rica do ponto de vista didático para o desenvolvimento do raciocínio probabilístico. Em primeiro lugar, porque envolve uma situação cotidiana que todos conhecem e que provoca curiosidade nos alunos. Em segundo, porque fornece um modelo de raciocínio que pode ser aplicado a qualquer jogo de apostas, entre outras situações. Por fim, o problema acima prioriza a análise do resultado obtido juntamente com o significado de obter uma probabilidade no contexto em que ela se apresenta.

Ainda que a definição clássica de probabilidades proporcione um modelo capaz de resolver uma fatia importante de problemas relevantes, ela pressupõe duas condições fundamentais que Ω deve satisfazer: $1)$ ser finito; $2)$ seus eventos elementares devem ser considerados, por alguma razão, equiprováveis, ou seja, atribui-se a cada elemento de Ω chance igual de ocorrência.

Dessa forma, tal definição possui óbvias limitações, pois não pode ser aplicada em situações que envolvem espaços amostrais infinitos e/ou não equiprováveis. Vejamos alguns problemas probabilísticos que não satisfazem a pelo menos uma dessas condições.

(a) *Em um jogo de tabuleiro, dois dados são lançados e toma-se a soma das faces voltadas para cima para contar o número de casas que cada jogador deve "andar" na sua vez. Qual é a probabilidade de um jogador andar pelo menos 10 casas em uma rodada?*

O conjunto de resultados possíveis para a soma dos dados é

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

que é um espaço finito, porém, como o leitor deve ter notado, não é equiprovável. Veremos mais detalhes desse caso no Exemplo 3.5.

(b) *Qual é a probabilidade de um motorista recém-habilitado causar um acidente no seu primeiro ano com a CNH?*

Os resultados possíveis aqui são *"causa acidente"* e *"não causa acidente"*. Como são apenas dois resultados possíveis, devemos atribuir $1/2$ para cada evento elementar? Ou, em outras palavras, o espaço Ω do experimento é equiprovável?

(c) *Ao comprar uma lâmpada, qual é a probabilidade de que ela funcione por mais de cinco anos?*

Como o tempo de duração de uma lâmpada é uma grandeza real positiva, temos $\Omega = \mathbb{R}_+$, que é infinito (não enumerável, inclusive) e não é equiprovável. Sugerimos ao leitor que tente imaginar como seria aplicar a definição clássica a esse exemplo.

(d) *Ao lançar uma moeda sucessivamente até obter cara, qual a probabilidade de que essa ocorra após o quinto lançamento?*

Como o interesse do experimento é o número de lançamentos necessários até que se obtenha cara, então o espaço amostral deve conter os números inteiros positivos 1, 2, 3, ... Mas há algum limite para o número de lançamentos possíveis até o sucesso? Embora na prática seja bastante improvável lançar uma moeda honesta um número grande de vezes sem que ocorra uma cara, não temos como garantir que após um determinado número de lançamentos (ainda que bem grande) iremos consegui-la. Por isso, só nos resta considerar seu espaço amostral infinito, $\Omega = \mathbb{N}^*$.

(e) *Ao abordar pessoas na rua para participar de uma pesquisa, qual a probabilidade de que sejam necessárias pelo menos 10 abordagens até que se encontre alguém que aceite?*

Considerando que uma mesma pessoa possa ser abordada mais de uma vez, temos uma situação semelhante à do exemplo anterior de tentativas sucessivas até que se obtenha o sucesso, ou seja, $\Omega = \mathbb{N}^*$.

(f) *Ao atirar um dardo em um alvo circular, qual é a probabilidade de acertar a região central que vale mais pontos?*

Se o alvo circular tem raio r, então o problema pode ser modelado utilizando o espaço $\Omega = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 \leq r^2\}$, que é infinito.

(g) *Um experimento que consiste contar o número de partículas alfa emitidas por grama de um material radiotivo, num intervalo de tempo de um minuto. Calcule a probabilidade de haver mais de 2 emissões em um minuto.*

Novamente, não temos como garantir um número máximo de partículas emitidas nesse período. Logo, $\Omega = \mathbb{N}$.

No Capítulo 3, voltaremos a tratar desse assunto e apresentaremos outras possíveis definições de probabilidade que permitem lidar com as situações listadas acima. Entretanto, enfatizamos aqui a importância de desenvolver nos alunos a compreensão de que a definição clássica não é única e de que há condições que devem ser necessariamente satisfeitas para que ela seja aplicada. É importante que essa compreensão seja desenvolvida logo no início da abordagem do tema Probabilidade, pois está diretamente ligada ao significado de atribuir uma probabilidade a um evento.

Entre as reflexões que consideramos importantes de serem feitas em sala de aula no tabalho inicial com o tema Probabilidade, destacamos mais uma a seguir, cuja falta de compreensão gera confusões comuns entre a probabilidade teórica e o resultado observado.

Qual é o significado prático de atribuirmos um número para a probabilidade de um evento? De que forma isso se traduz na observação do fenômeno aleatório?

O que significa na prática dizer que a probabilidade de sair 3 no lançamento de um dado é $1/6$? Se lançarmos um dado equilibrado seis vezes em exatamente uma vez sairá o 3? Ou, aumentando a aposta, se lancarmos um dado 6000 vezes, teremos 1000 ocorrências do número 3? E se fizermos 6 milhões de lançamentos? E outra questão importante: o que ocorre com a frequência relativa de resultados iguais a 3? Ou seja, dos 6 milhões de lançamentos (ou dos 6000, ou dos 6), qual é a porcentagem de ocorrências do número 3?

Um famoso teorema conhecido como Lei dos Grandes Números (LGN) nos dá uma direção sobre isso. Informalmente, esse teorema afirma que a frequência relativa (porcentagem) das ocorrências de um evento A, quando o experimento é repetido muitas vezes de forma independente, converge para a probabilidade desse evento. Falaremos uma pouco mais dessa lei no Capítulo 3.

Mas mesmo sem entrar em detalhes, pela LGN, devemos ter as frequências relativas do evento citado acima convergindo para $1/6$. O que de forma alguma significa que em um número grande (ou pequeno) de lançamentos, estaremos infringindo as leis da probabilidade caso a frequência de ocorrências do número 3 não seja $1/6$. Em outras palavras, probabilidade igual a $1/6$ não significa ocorrência em $1/6$ das vezes em que o experimento foi repetido.

Isso não quer dizer que o número $1/6$ não tenha significado prático. Pelo contrário, ele nos dá uma medida teórica do que deve acontecer a longo prazo, do que devemos esperar daquele experimento. No final das contas, qualquer probabilidade calculada nos dá uma medida de expectativa sobre a ocorrência do evento.

Em [20], os autores apresentam uma discussão que pode parecer contraintuitiva para alguns, e que costuma ser objeto de confusão. O experimento proposto é imaginar uma sequência de lançamentos de moedas honestas na qual a quantidade de caras obtidas está 10 unidades à frente da quantidade de coroas.

A questão colocada é: se a frequência relativa deve convergir para $1/2$, então as coroas devem "recuperar o terreno"? E, se for assim, em algum momento, será mais provável acontecer coroa para que elas possam alcançar o número de caras? Estará correto esse raciocínio?

Certamente que não! Podemos estender essa ideia para uma vantagem de j unidades das caras em relação às coroas. Ainda que esse j seja grande, mas em algum sentido "comportado", se tivermos um padrão de n coroas e $n + j$ caras, o limite da frequência relativa será

$$\lim_{n\to\infty} \frac{n}{2n + j} = 1/2$$

Assim, se em algum momento o número de caras disparou e ficou 200 unidades à frente, não é necessário que as coroas recuperem a vantagem para que as leis da probabilidade continuem valendo. Nas palavras dos autores:

"a proporção aproxima-se de 0,5 não é o mesmo que o número de caras aproxima-se da metade do número de lançamentos"

Um experimento interessante que pode ser facilmente reproduzido em sala de aula é a simulação do lançamento de uma moeda um número grande de vezes, através de um *software* ou aplicativo para *smartphone* - computando-se em uma tabela e visualizando em um gráfico o número de caras a mais que o esperado e a frequência relativa de caras, conforme n cresce.

Seguindo a sugestão feita pelos autores em [20], fizemos nossa própria simulação de 10000 lançamentos de moedas no *Microsoft Excel*, com o objetivo de simular o experimento real realizado pelo matemático John Kerrich. Os gráficos das Figuras 2.1 e 2.2 apresentam os resultados dessa simulação. O primeiro deles traz, para cada valor de n, o número caras obtidas a mais do que o esperado, ou seja, a mais do que $n/2$. O segundo gráfico apresenta a evolução da frequência relativa do número de caras, confirmando as afirmações feitas acima.

Observe que o número de caras acima do esperado tem oscilação maior para valores maiores de n, enquanto a frequência relativa estabiliza-se em torno de $1/2$.

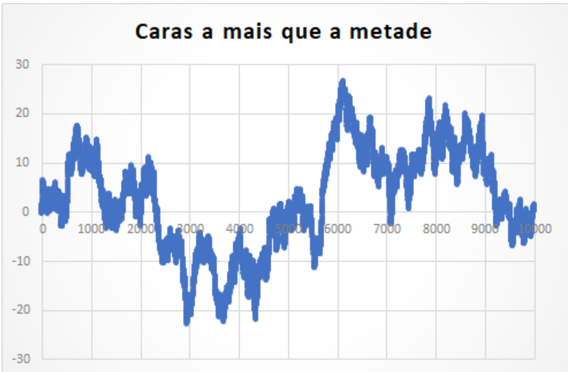

Figura 2.1: Simulação de 10000 lançamentos de uma moeda feita no *Microsoft Excel*. Número de caras a mais que a metade.

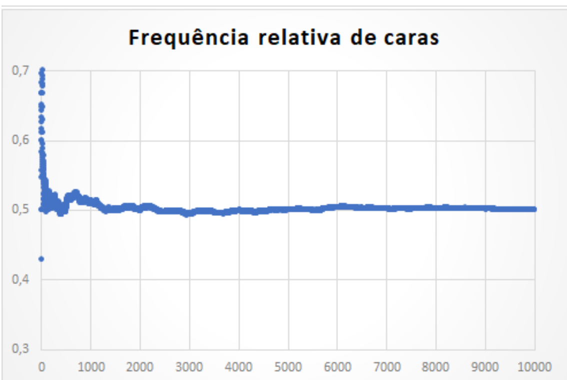

Figura 2.2: Simulação de 10000 lançamentos de uma moeda feita no *Microsoft Excel*. Frequência relativa do número de caras.

Para reproduzir um experimento como esse em planilhas eletrônicas, há um caminho relativamente comum. No caso do *Microsoft Excel*, pode-se seguir o seguinte procedimento:

1- Gere 10000 números aleatórios (ou qualquer outra quantidade) com o comando "=ALEATÓRIO()" na célula A1 e arraste para as 10000 primeiras células da coluna A;

2- Na célula B1, insira o comando "=SE(A1<0,5;1;0)". Em seguida arraste este comando até a célula B10000;

3- Tomando as corrências de 1 como caras, o número de caras pode ser obtido com o comando "=SOMA(B1:B10000)";

4- O número de caras a mais que a metade pode ser obtido criando uma colula C com todos os números inteiros de 1 a 10000. Em seguida, na célula D1, insira o comando "=SOMA($B$1:B1)-INT(C1/2)" e arraste até a célula D10000;

5- A frequência relativa de caras pode ser obtida na colula E com o comando "=SOMA($B$1:B1)/C1" e arrastando para as células seguintes;

6- Os gráficos podem ser obtidos selecionando as colunas C e D e depois C e E, utilizando a ferramenta de inserir gráficos.

Há diversas variações desse mesmo e de outros experimentos que podem ser obtidas em planilhas eletrônicas. Os próprios alunos podem ser desafiados a reproduzirem lançamentos de dados, jogo de roleta ou outros experimentos aleatórios.

Para experimentos de simulação em sala de aula, que consideramos bastante importantes para compreensão do significado da probabilidade, recomendamos ainda o uso de aplicativos para *smartphone*, pois os alunos podem fazer suas próprias simulações, gráficos e tabelas sobre os resultados obtidos. Na Figura 2.3, apresentamos três *prints* da tela do aplicativo *Random Number Generator Plus*. No primeiro deles, há uma simulação do lançamento de uma moeda 50 vezes; no segundo, o sorteio de 20 números de 1 a 1000; e, no terceiro, o lançamento de três dados.

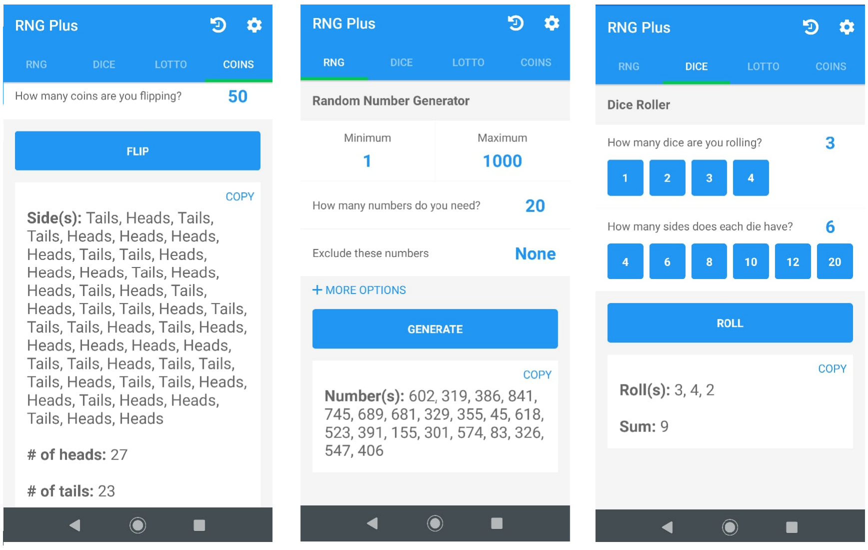

Figura 2.3: Simulações em aplicativo - Fonte: *Random Number Generator Plus*

Na Tabela 2.2, mostramos uma sugestão de simulações de lançamentos de uma moeda que podem ser feitas no aplicativo citado acima, com levantamento do número de caras, coroas, caras a mais que a metade, e a frequência relativa, para valores cada vez maiores que n.

Tabela 2.2: Simulações no aplicativo *Random Number Generator Plus*

Lançamentos (n)	**Nº de caras**	**Nº de coroas**	**Caras a mais que $n/2$**	**Frequência**
10	7	3	2	0,2
50	23	27	-2	0,46
100	46	54	-4	0,46
200	101	99	1	0,505
500	247	253	-3	0,494
1000	517	483	17	0,517

2.2 O Foco em Problemas Combinatórios

Nesta seção, vamos propor uma reflexão sobre a excessiva dependência das técnicas de contagem estudadas em Análise Combinatória contida nos problemas de probabilidade tratados no ensino médio. A despeito da importância inegável desses problemas, discutiremos, através dos exemplos, a possibilidade de que essa dependência possa ofuscar ou até dificultar o desenvolvimento do raciocínio probabilístico.

Vamos resolver e analisar alguns problemas comuns em textos didáticos de Probabilidade voltados para o ensino médio. Em seguida, vamos propor algumas questões sobre os aspectos didáticos desses problemas.

Exemplo 2.3. *No concurso da Mega-Sena, ganha o prêmio máximo o apostador que acertar as 6 dezenas sorteadas de um total de 60 dezenas. Qual é a probabilidade de:*
a) um apostador ganhar o prêmio máximo fazendo um jogo simples?
b) um apostador ganhar o prêmio máximo com um jogo de oito dezenas?
c) um jogador acertar a quadra (acertar 4 das 6 dezenas sorteadas) com um jogo simples?

$a)$ $\dfrac{1}{50.063.860}$, conforme resolvido no Exemplo 2.2.

$b)$ Sendo o evento B: "ganhar o prêmio máximo com um jogo de oito dezenas", temos

$$n(B) = C_{8,6} = \frac{8!}{6!2!} = 28$$

pois nessa situação o apostador escolheu 8 dezenas e qualquer agrupamento de 6 dessas 8 dezenas o permite ganhar. Como Ω é o mesmo para os três itens, temos:

$$P(B) = \frac{C_{8,6}}{C_{60,6}} = \frac{28}{50.063.860} = \frac{1}{1.787.995}$$

$c)$ Considere o evento C: "acertar a quadra com um jogo simples". Nesse caso, para o apostador ganhar a quadra devem ser sorteadas 4 das 6 dezenas que ele escolheu, e as outras 2 dezenas devem ser sorteadas entre as 54 que ele não escolheu. Caso contrário o jogador ganharia a quina (quando se acerta 5 das dezenas sorteadas) ou o prêmio máximo. Assim:

$$n(C) = C_{6,4} \cdot C_{54,2} = \frac{6!}{4!2!}\frac{54!}{52!2!} = 21.465$$

Portanto:

$$P(C) = \frac{C_{6,4} \cdot C_{54,2}}{C_{60,6}} = \frac{21.465}{50.063.860} = \frac{4.293}{10.012.772}$$

Considere agora as seguintes questões:

1) *Quais conceitos relacionados à Probabilidade o aluno deve dominar para resolver esse problema?*

2) *Em que aspectos estão as maiores dificuldades do problema para o aluno? Há mudança no nível de dificuldade entre os itens?*

1) Além de identificar o espaço amostral e os eventos de interesse do problema, o principal conceito de probabilidade exigido nos três itens para resolver o problema foi a definição clássica de probabilidade, ou seja, a ideia de que se deve dividir o número de "casos favoráveis" pelo de "casos possíveis". Todos os outros conhecimentos exigidos para resolver esse problema são relacionados às técnicas de contagem.

2) À medida que avançamos nos itens, o nível de dificuldade aumenta, mas essa dificuldade não avançou em raciocínio probabilístico e sim em raciocínio combinatório. As maiores dificuldades do problema estão em fazer as contagens dos números de "casos favoráveis" e de "casos possíveis". Mas nos itens $b)$ e $c)$ não foi exigida do aluno nenhuma habilidade probabilística que não tivesse sido exigida no item $a)$. Ou seja, a partir do item $b)$, esse problema poderia ser encarado como uma revisão ou reforço de Análise Combinatória.

E por isso deixamos a questão: dado que as técnicas de contagem foram previamente estudadas e exercitadas pelos alunos, o tempo utilizado em questões de contagem nesse momento não poderia ser mais bem aproveitado com avanços em tópicos de Probabilidade?

Entendemos que o exemplo acima é relevante e interessante para ser apresentado em sala de aula; os alunos costumam ser atraídos pelo assunto de loterias porque isso faz parte do dia a dia das pessoas. É importante, porém, identificar no ensino de probabilidades as oportunidades para o desenvolvimento do raciocínio probabilístico distinguindo-o do raciocínio combinatório.

Na resolução do Exemplo 2.2 que também trata de loterias, mais do que os cálculos combinatórios, o enunciado traz o questionamento sobre a relação custo-benefício de cada aposta, onde o aluno deve avaliar a probabilidade de ganho obtida em conjunto com o prêmio prometido, desenvolvendo a noção de comparação entre chance e ganho.

Exemplo 2.4. *Ao organizar de forma aleatória cinco livros de Matemática e três de Física em uma estante, qual é a probabilidade de que os livros de Física fiquem juntos?*

Neste exemplo temos $n(\Omega) = P_8 = 8!$, pois devemos organizar 8 livros de forma aleatória e enfileirada, e o número de organizações possíveis é dado pela permutação de 8 elementos.

Por outro lado, considere o evento E: "os livros de Física ficam juntos". Então, considerando o bloco de Física como um livro único que pode ser permutado entre os de Matemática (mas não podem se separar), temos para o evento E todas as permutações possíveis de 6 elementos, 5 livros mais o bloco. Além disso, para cada permutação dessa, os livros de Física podem permutar entre si. Dessa forma temos $n(E) = P_6 \cdot P_3$. Logo,

$$P(E) = \frac{6!3!}{8!} = \frac{3}{28}$$

Sobre esse exemplo, considere as questões:

1) *Em que aspectos esse problema é diferente do anterior?*

2) *A situação-problema proposta é natural?*

1) A diferença entre esse problema e o anterior encontra-se apenas na mudança das técnicas de contagem aplicadas, ou seja, na exigência do raciocínio combinatório. Não existe entre os dois problemas nenhuma mudança em relação aos conceitos probabilísticos envolvidos. Ou seja, assim como avaliamos os itens $b)$ e $c)$ do exemplo anterior, esse pode ser considerado mais um exercício de revisão de Análise Combinatória, caso o aluno já tenha compreendido as ideias básicas da definição clássica. Novamente, entendemos que o tempo dispendido aqui poderia ser mais bem aproveitado com o desenvolvimento de outros aspectos do raciocínio probabilístico.

2) Esse é outro aspecto do problema que queremos chamar a atenção. A situação apresentada no enunciado é artificial, não é um tipo de probabilidade que se buscaria saber em uma situação real. É claro que problemas fictícios podem ter um papel didático importante em certas situações. Mas não é esse o caso, pois, como vimos, não houve avanços nesse problema em termos de conceitos probabilísticos. Além disso, em relação às outras áreas da Matemática, a área da Probabilidade tem a vantagem da facilidade de se encontrar problemas reais e relevantes para o trabalho em sala de aula. O exemplo anterior sobre loterias e os Exemplos 2.1 e 2.2 são boas ilustrações desse fato.

De forma geral, problemas de probabilidade com foco no raciocínio combinatório são predominantes em materiais didáticos, provas, planejamentos didáticos e na abordagem tradicional em sala de aula no ensino básico. Isso é uma consequência natural da adoção da definição clássica de probabilidade como prioritária e, em alguns casos, como única possível.

No próximo capítulo, vamos apresentar outras possibilidades de definição de probabilidade e, com isso, outras formas de abordagem a problemas envolvendo aleatoriedade. Como vimos na seção anterior, há inúmeros problemas probabilísticos relevantes que não dependem de técnicas de contagem ou que não podem ser modelados pela definição clássica. De fato, mais do que apresentar outras definições de probabilidade, o objetivo principal deste livro todo é apresentar novas formas de abordagem a conceitos probabilísticos que não dependam apenas da definição clássica e das técnicas de contagem, priorizando situações-problemas reais.

Reforçando o que foi dito na Introdução, vamos propor a seguir reflexões sobre pelo menos quatro consequências não desejadas do ensino de probabilidade com foco no raciocínio combinatório e na definição clássica.

Reflexão 1. *Pode-se desenvolver nos alunos (e até em parte dos professores) a falsa ideia de dependência e até mesmo de indissociabilidade das áreas de Probabilidade e Análise Combinatória.*

Ao longo da preparação deste material e do minicurso que deu origem a ele, levantamos vários casos baseados em nossas experiências como docentes do ensino básico e também em relatos de colegas nos quais a ideia acima é bastante presente. É comum em conversas com alunos, ao se fazer uma revisão de Probabilidade e perguntar ao aluno o que ele lembra a respeito, ouvir respostas como: "são aqueles cálculos que usamos combinação, permutação e depois dividimos".

Naturalmente, mas não no mesmo sentido dos alunos, essa cultura está permeada inclusive entre colegas professores. O trabalho durante anos com textos didáticos tradicionais com tal abordagem faz com que o próprio professor do ensino básico por vezes tenha essa visão distorcida.

Um grupo de estatísticos preocupados com as discussões curriculares que o país vem atravessando, sugeriu certa vez, em uma reunião de trabalho sobre currículo promovida pela Sociedade Brasileira de Matemática (SBM), que a Análise Combinatória fosse ensinada depois da Probabilidade e não antes. Ou, pelo menos, que fosse ensinada em um momento distante, sem fazer a ligação entre as áreas.

Alguns professores de Matemática reagem a essa ideia como se fosse absurda, que isso não seria possível. Mas há alguns argumentos que apontam que a Análise Combinatória pode atrapalhar, em um sentido que veremos mais à frente, o desenvolvimento dos pensamentos estatístico e probabilístico.

No documento [14], produzido por uma comissão da Associação Brasileira de Estatística, alguns dos pontos acima são levantados, dos quais destacamos os trechos a seguir, que têm como objetivo fazer recomendações curriculares para o ensino básico:

"Enfatizar as noções de probabilidade introduzindo o conceito de incerteza em atividades cotidianas sem o uso de técnicas de análise combinatória e sem o cálculo de probabilidades de eventos finitos específicos em situações fictícias."

"A análise combinatória (uma área importante em nível superior que, a nosso ver, deve ser contemplada na escola básica em Números) não deve ser considerada nessa fase como pré-requisito para a área de Estatística. Para probabilidade, a construção dos diagramas de árvore (que embutem o raciocínio combinatório) é suficiente."

Há de se observar que, no formato tradicional do ensino de Probabilidade adotado em larga escala, a reação contrária dos professores de Matemática é mais do que justificada; não há como ensinar Probabilidade nesse formato sem ensinar Análise Combinatória antes. Assim, só uma mudança na forma de abordagem inicial de Probabilidade e Estatística poderia mudar tal realidade.

O fato é que essa falsa ideia de dependência entre as áreas não se sustenta nem em problemas práticos cotidianos, nem quando o aluno ingressa no ensino superior e começa a ter contato com aplicações da Probabilidade e da Estatística em suas áreas de atuação.

Reflexão 2. *Deixa-se de desenvolver nos alunos boa parte do raciocínio probabilístico elementar que, como veremos à frente, é fundamental não só para as aplicações modernas da Teoria das Probabilidades em diversas áreas do conhecimento científico e tecnológico, como para lidar com problemas e informações cotidianas relacionadas à área.*

Ao longo deste livro apresentaremos vários conceitos e problemas probabilísticos relevantes e aplicados que dependem do desenvolvimento do raciocínio probabilístico elementar. Saber utlizar as propriedades da probabilidade, utilizar diagramas de árvores, decompor um evento pela probabilidade total, identificar e fazer cálculos com situações de independência, o raciocínio condicional, o raciocínio frequentista, as relações com a estatística, as relações entre probabilidade teórica e resultado observado, entre outras, são habilidades probabilísticas fundamentais para resolver problemas envolvendo aleatoriedade.

Como observamos mais acima nesta seção, o tempo dispendido com problemas combinatórios permeados no ensino de probabilidade faz com que essas habilidades sejam cada vez menos ou nada desenvolvidas.

Apresentamos abaixo uma lista de enunciados de problemas relevantes cuja resolução não dependem de técnicas de contagem, mas dependem fortemente do raciocínio probabilístico elementar. Ao longo desse livro voltaremos a cada desses problemas para apresentar sua resolução.

Exemplo 2.5. *Em um exame médico, são conhecidas as taxas de falso positivo e falso negativo. Dada a incidência da doença na população, como calcular a probabilidade que uma pessoa que teve o exame positivo não esteja doente? E de quem teve o exame negativo para que esteja doente?*

Exemplo 2.6. *Como estimar a abrangência da imunização de uma vacina aplicada a uma certa população e que tem 70% de eficácia? Qual é o valor esperado para o número de imunizados em uma população?*

Exemplo 2.7. *Partindo um fio de espaguete em três partes, qual é a probabilidade de que elas formem um triângulo?*

Exemplo 2.8. *Um instituto de meteorologia divulgou sua previsão do tempo de amanhã para uma cidade litorânea, estimando uma probabilidade de 70% para ocorrência de chuvas e uma probabilidade de 60% para a ocorrência de ventos fortes. É possível descartar a ocorrência simultânea desses eventos climáticos? Se não, qual é a probabilidade máxima dessa ocorrência simultânea?*

Exemplo 2.9. *Durante a II Guerra Mundial, aviões americanos que voltavam do combate foram analisados para que houvesse reforço da estrutura nas regiões mais alvejadas por tiros. Essa escolha de reforço é a mais correta? De que forma o raciocínio condicional permite analisar o problema?*

Exemplo 2.10. *Uma britânica foi condenada à prisão por ter assassinado seus dois filhos recém-nascidos por sufocamento. A defesa alegava que a causa da morte fora síndrome da morte súbita infantil (SMSI). A acusação retrucou, afirmando que uma morte por SMSI tem probabilidade de ocorrência de* $1/8543$ *e, "portanto", duas mortes por essa causa teriam probabilidade de* $\left(\dfrac{1}{8543}\right)^2 = \dfrac{1}{72.982.849}$. *Há pelo menos dois erros na argumentação da acusação. Quais são eles?*

Reflexão 3. *Muitos problemas da área têm soluções tanto pelo caminho combinatório quanto pelo caminho probabilístico. Soluções de problemas com uso dos raciocínios implícitos nas propriedades da probabilidade, nos importantes teoremas de Bayes e da Probabilidade Total, no raciocínio condicional, entre outros, são uma oportunidade para o professor desenvolver simultaneamente com a abordagem combinatória, compará-las e permitir que o aluno possa escolher a estratégia que lhe parece mais eficiente.*

No Exemplo 5.10 do Capítulo 5, apresentamos um problema que pode ser resolvido usando tanto técnicas de contagem, quanto diagrama de árvore, que é baseado no Teorema da Probabilidade Total. Na nossa percepção e também na de uma aluna que nos procurou para perguntar sobre o problema, a abordagem probabilística pareceu mais simples do que a combinatória e, posteriormente, ajudou inclusive a encontrar o erro que estava sendo cometido nas contagens da solução combinatória.

A comparação de soluções por dois ou mais caminhos distintos já traz vantagens didáticas por si só, independentemente do conteúdo trabalhado. Mas no caso da Análise Combinatória e da Probabilidade, vemos ainda a vantagem de que determinados problemas de contagem dividem-se em muitos casos e, dependendo da complexidade, é comum que se conte algo a mais ou se esqueça de contar algo. A abordagem via probabilidades e diagrama de árvore ajuda a elucidar esse tipo de problema.

Além disso, há problemas de probabilidade em que tanto faz você contar com ordenação ou sem ordenação, mas se você contar os casos favoráveis com ordenação e os casos possíveis sem, a razão que define a probabilidade não será a correta. Dado que é comum ocorrer confusões entre os alunos sobre quando há ou não ordenações e como calcular suas possibilidades, eliminar essa dificuldade do problema com a abordagem probabilística parece ser bastante vantajoso.

Reflexão 4. *Problemas de contagem podem ter elevados graus de dificuldade, e o foco do ensino de Probabilidade em problemas combinatórios pode causar uma desnecessária aversão ou medo nos alunos ao trabalhar com a unidade temática de Probabilidade e Estatística da BNCC durante o ensino básico e também ao lidar com estudos futuros de conceitos dessa área.*

A dificuldade que alunos e também professores têm para resolver problemas combinatórios complexos é natural. Ainda que haja problemas padrões de contagem um pouco mais simples, muitas aplicações exigem cuidado e conhecimento de técnicas e macetes adquiridos após muita experiência com esse tipo de problema.

Devido a tais características, uma parcela dos alunos demonstra mais dificuldade em cálculo combinatório do que em outras áreas da Matemática, o que pode produzir até uma aversão a esse tópico.

Associada ao problema da falsa ideia de dependência e indissociabilidade entre probabilidade e combinatória, apontado na Reflexão 1, essa aversão pode trazer consequências desnecessárias ao aprendizado futuro de conceitos fundamentais de Probabilidade e Estatística no ensino superior. O aluno que cria essa aversão e relação de dependência pode transferir essa dificudade para a aprendizagem de tópicos importantes que não têm dependência nenhuma de técnicas de contagem.

As consequências não desejadas apontadas nas quatro reflexões acima são comumente observadas e relatadas por professores e alunos quando instigados a refletir sobre o assunto. Entretanto, sua ocorrência, em geral, se dá de forma involuntária, sem que os atores envolvidos no processo se deem conta. Isso ocorre, em parte, pelo caminho percorrido pelos materiais didáticos tradicionais, que quase sempre seguem um roteiro padrão com as características que apontamos acima. Dessa forma, esperamos que a abordagem da Probabilidade que apresentaremos nos próximos capítulos seja útil para ajudar a quebrar esse ciclo.

Capítulo 3

Definições de Probabilidade: os Grandes Números e os Axiomas

> *"A Teoria das Probabilidades como disciplina Matemática pode e deve ser desenvolvida a partir de axiomas, exatamente do mesmo modo que a Geometria e a Álgebra."*
>
> – Andrey Kolmogorov

Apresentaremos neste capítulo algumas definições de probabilidade que, em geral, recebem pouca atenção no contexto do ensino básico. Como vimos na Seção 2.1, existem incontáveis exemplos de problemas probabilísticos e situações reais para os quais a definição clássica de probabilidades não é um bom modelo matemático.

A definição frequentista de probabilidade, por exemplo, é utilizada em diversas aplicações práticas e possui nuances que devem ser objeto de reflexão em sala de aula para que seu uso não seja feito de forma equivocada, como veremos em exemplos mais à frente.

É importante salientar que as definições clássica, frequentista e geométrica (que veremos no Capítulo 8) têm apelo mais intuitivo e são suficientes para a modelagem de muitos problemas importantes em seus contextos. Uma abordagem cronológica a partir da história da Probabilidade, que consideramos recomendável em sala de aula, certamente trará essas definições antes de sua formalização.

Por outro lado, se olharmos de um ponto de vista teórico da matemática formal, a definição axiomática de probabilidade, introduzida na primeira metade do século XX por A. N. Kolmogorov, trouxe à Teoria das Probabilidades um caráter de formalidade, rigor e precisão típico de outras áreas da Matemática. Isso permitiu o desenvolvimento teórico e da pesquisa na área, durante um século marcado por avanços fundamentais e em grande quantidade na Probabilidade.

A definição axiomática, além de conferir rigor à probabilidade, tem a vantagem de incluir as outras definições como casos particulares. No contexto do ensino básico, a definição geral usualmente é adaptada para espaços amostrais finitos, e costuma estar presente em textos didáticos. Entretanto, nossa percepção é a de que ela é pouca explorada neste segmento, com raros exemplos de espaços finitos não equiprováveis que devem receber um cuidado maior ao se definir uma probabilidade. Entendemos ainda que a relação entre a definição clássica e a axiomática precisa receber uma atenção maior, chamando a atenção para o contexto histórico e para a necessidade de generalização.

Pensando um pouco mais à frente, apresentaremos também neste capítulo a definição axiomática mais geral que inclui espaços amostrais infinitos, dividindo-a em dois casos: espaços enumeráveis e não enumeráveis. A definição geral apresentada por Kolmogorov trata de espaços não enumeráveis, englobando as outras como casos particulares. Essas definições utilizam ferramentas matemáticas mais avançadas, como união e soma infinitas e, por isso, não são adequadas para o ensino básico.

Entretanto, essas definições são a base teórica para diversos problemas simples que apresentaremos neste livro, como problemas de probabilidade geométrica e sequências de lançamentos de moedas, por exemplo. Problemas desse tipo são intuitivos e podem ser resolvidos sem a necessidade de conceitos matemáticos de ensino superior. O embasamento teórico apresentado aqui tem como alvo o público de professores de matemática, mas suas aplicações são adequadas à sala de aula do ensino médio.

3.1 Definição Frequentista

Na Estatística, uma forma frequente de atribuir probabilidades a um evento A de um espaço amostral Ω é através da análise da frequência relativa desse evento, após n repetições do experimento associado a Ω. A ideia de utilizar frequência relativa como medida de probabilidade surge naturalmente a partir de aplicações práticas.

Se o evento A ocorre n_A vezes em n repetições independentes desse experimento, utiliza-se como aproximação para a probabilidade de ocorrência do evento A a frequência relativa de A dada por

$$f_n(A) = \frac{n_A}{n} \simeq P(A)$$

Como veremos, existem problemas que podem ser modelados por mais de uma definição de probabilidade, mas a modelagem frequentista aproxima-se mais da realidade nesses casos. Por outro lado, devemos tomar cuidado ao utilizar essa abordagem, que nem sempre é adequada. Por isso a compreensão dos conceitos que embasam a definição frequentista são fundamentais para evitar aplicações equivocadas.

Vejamos algumas situações nas quais a frequência relativa é mais adequada para a modelagem do que a definição clássica.

Exemplo 3.1. *Um empresa de crédito pessoal depende, entre outros fatores, da análise de riscos de inadimplência para que sua atividade seja sustentável. Ao* *analisar o perfil de um cliente, como estimar a probabilidade de que ele fique inadimplente?*

Esse tipo de situação assemelha-se à situação do motorista recém-habilitado na Seção 2.1. Se analisarmos sob o ponto de vista da Definição Clássica, o espaço amostral $\Omega = \{$"fica inadimplente", "não fica inadimplente"$\}$ teria seus eventos equiprováveis. Logo, a probabilidade desse cliente ficar inadimplente é 50%. Isso faz sentido em situações reais?

O mais plausível a se fazer é estudar o perfil do cliente e verificar com que frequência esse tipo de pessoa fica inadimplente. A partir daí, atribuir um número a essa probabilidade. Esse tipo de estudo, em geral, envolve uma série de técnicas estatísticas e de análise de perfil. Mas a frequência relativa dos dados observados está presente, em essência, para a determinação dessa probabilidade.

Exemplo 3.2. *Uma rede de locadora de automóveis possui 250 lojas na Grande São Paulo, sendo 180 na capital. Um cliente que retira um carro numa loja, pode devolver em qualquer outra nessa região.*

a) Qual é a probabilidade de que um carro alugado nessa rede seja devolvido em uma loja da capital?

b) Há mais de uma abordagem possível para esse problema?

a) Se todas as lojas tiverem a mesma probabilidade de receber o carro, teremos um problema modelado pela Definição Clássica. E então:

$$P(E) = \frac{180}{250} = 72\%$$

.

b) Será que as lojas nos bairros centrais ou próximas a aeroportos não recebem mais carros? Ou algum outro fator influencie na hora de escolher o local para devolver o carro? Na prática, o que poderia acontecer se um diretor da rede responsável pelo planejamento de espaços e frotas nas lojas partisse da hipótese de que todas as lojas recebem os carros de forma equiprovável?

Parece razoável afirmar que as lojas não terão devolução equiprovável, o que fará com que algumas delas necessitem de espaço e equipe de funcionários maior para dar conta da demanda de devoluções e reacomodação da frota. Nesse caso, a frequência relativa como medida de probabilidade para cada loja e, consequentemente, para as lojas da capital nos dá uma noção mais próxima da dinâmica real de devoluções.

Assim, se E é o evento em que um carro alugado nessa rede é devolvido em uma loja da capital, então, de posse dos dados de n devoluções passadas e do número n_E de devoluções na capital, um valor mais próximo da realidade seria:

$$f_n(E) = \frac{n_E}{n} \simeq P(E)$$

Neste ponto deparamos-nos com uma questão importante: a probabilidade como frequência relativa está bem definida? Com a aquisição de novos dados, o valor de $f_n(A)$ pode ser alterado, o que pode fazer a probabilidade de um evento ser um tanto quanto instável. Além disso, há alguma garantia de que a frequência observada de um evento seja uma boa medida para ocorrências futuras? Vamos tentar responder essas perguntas usando a famosa *Lei dos Grandes Números*, sobre a qual apresentaremos abaixo uma versão simplificada e informal.

Lei dos Grandes Números: *Considere um evento $A \subset \Omega$. A cada sequência de n de repetições independentes do experimento em questão, tal que $P(A)$ é constante, seja n_A o número de ocorrências do evento A. Então, com probabilidade igual a 1, tem-se:*

$$\lim_{n\to\infty} f_n(A) = \lim_{n\to\infty} \frac{n_A}{n} = P(A)$$

As versões precisas desse importante Teorema e suas respectivas demonstrações, que não fazem parte de nossos objetivos, podem ser encontradas em [9].

O leitor deve observar que o enunciado acima é bastante intuitivo. Por exemplo, imagine uma sequência de lançamentos independentes de um dado honesto, onde consideramos o evento $A = \{5, 6\}$. Nesse caso, para cada lançamento temos $P(A) = 1/3$. Sabemos que se lançarmos o dado um número grande de vezes, não necessariamente teremos $f_n(A) = 1/3$, mas o enunciado acima nos garante que quanto maior for o número de lançamentos mais próximo de $1/3$ estará $f_n(A)$. Isso deve ocorrer com probabilidade 1, que é uma afirmação cujo sentido não iremos aprofundar.

Mas de que forma essa lei pode nos ajudar na consistência da definição frequentista? Se fizermos as suposições de que existe uma probabilidade constante para um evento A e de que n_A foi obtido por meio de repetições independentes do experimento, então, pela Lei dos Grandes Números, a frequência relativa $f_n(A)$ tende a se estabilizar e convergir para $P(A)$, conforme n cresce.

Em outras palavras, nas condições dadas, $f_n(A)$ é uma excelente estimativa para a probabilidade "natural" de um evento A observado no mundo real. Obviamente, quanto maior o valor de n, mais precisa será a estimativa

Assim, nas condições acima, ficamos mais seguros quanto à existência do limite dado na seguinte definição.

Definição 3.1 (Definição Frequentista)**.** *Considere um evento $A \subset \Omega$ e, para todo $n \geq 1$, uma sequência de n repetições independentes do experimento em questão. Se n_A é o número de ocorrências de A, definimos:*

$$P(A) = \lim_{n\to\infty} \frac{n_A}{n} = \lim_{n\to\infty} f_n(A)$$

se o limite existir.

Da definição e das considerações anteriores temos, para n suficientemente grande:

$$f_n(A) \simeq P(A)$$

Na próxima seção mostraremos que, assim como no caso da definição clássica, a definição frequentista de probabilidade é um caso particular da definição axiomática. Por outro lado, ela também pode ser vista como consequência da construção axiomática por causa da Lei dos Grandes Números.

É importante ressaltar que a utilização da frequência relativa como probabilidade requer que sejam assumidas as condições de independência e existência de uma probabilidade constante para o evento A. Ou seja, tal definição também tem suas limitações. O exemplo a seguir mostra o sentido prático dessas limitações, comparando situações nas quais se discute a adequação da definição frequentista para modelar o problema.

Exemplo 3.3. *Considere as seguintes situações nas quais desejamos inferir a probabilidade de:*

1) O lançamento de uma moeda resultar em cara.

2) Um medicamento em testes fazer efeito em um paciente.

3) Um jogador de basquete acertar um lance livre durante um jogo.

4) Chover em uma certa região durante o dia de amanhã.

Em qual dos casos o uso da definição frequentista parece ser mais adequado? Por quê? Quais são os aspectos intuitivo e matemático envolvidos na diferença entre as quatro situações?

Cada situação acima terá apenas dois resultados possíveis: cara ou coroa, faz efeito ou não faz efeito, acerta ou não acerta, chove ou não chove. Com exceção da situação 1), não parece fazer sentido considerar esses espaços amostrais como equiprováveis. Então, a modelagem pela definição clássica não é uma boa opção para 2, 3 e 4. Vamos analisar a aplicação da definição frequentista em cada caso.

1) Nesse caso, é bastante razoável supor verdadeiras as duas hipóteses que embasam a convergência da frequência relativa, pois os lançamentos consecutivos de uma moeda ocorrem de forma independente e certamente a moeda não sofre alterações a cada lançamento, de modo que a probabilidade de sair cara é constante. Logo, apesar de a definição clássica ser mais simples e melhor para modelar o lançamento de uma moeda honesta, a definição frequentista também é uma boa opção. De fato, inúmeros experimentos já foram realizados nesse sentido, confirmando experimentalmente a convergência de $f_n(A)$ para $1/2$, alguns deles citados em [5].

2) Os testes com o medicamento entre os pacientes ocorrem de forma independente. Por outro lado, é razoável supor que a probabilidade de o medicamento fazer efeito é constante, pois sua composição permenece a mesma. Assim, para n suficientemente grande, a frequência relativa, com todos os cuidados e métodos que testes de medicamento exigem (como o duplo cego), é uma boa estimativa para a probabilidade em questão.

3) Nesse caso, apesar de as estatísticas de acerto por vezes serem usadas para descrever probabilidades na imprensa esportiva, as hipóteses de independência e probabilidade constante não são tão precisas. No primeiro caso, porque o fator confiança faz diferença no esporte e pode afetar o desempenho de um atleta. Uma sequência grande de erros (ou de acertos), em geral, trazem a carga psicológica para os próximos arremessos, podendo tornar os resultados dependentes dos resultados anteriores. Além disso, o desempenho dos atletas varia no decorrer da temporada, o que contraria o fato de a probabilidade de acerto ser constante. De qualquer forma, caso se observe ao longo da temporada uma estabilização das frequências relativas, poderíamos concluir que os fatores elencados acima têm pouca influência.

4) Esse caso é o mais intuitivo que os anteriores. Para determinar uma probabilidade de chuva para o dia de amanhã, faz sentido observar se choveu ou não nos últimos 365 dias e calcular sua frequência relativa? Certamente que não! Há épocas do ano com mais propensão a chuvas do que outras. Além disso, as ocorrências de chuva ou não entre os dias não são independentes. Condições para chuva em um dia podem favorecer a ocorrência ou não do dia seguinte. Por fim, é claro que a probabilidade de chover em um dia qualquer não é constante, depende de uma série de fatores climáticos a serem observados. Esse tipo de probabilidade meteorológica utiliza modelos mais sofisticados do que a simples frequência relativa.

Comentários. As situações elencadas acima e muitas outras no mesmo sentido trazem uma oportunidade de desenvolver uma atividade didática coletiva que abordaria variados conceitos probabilísticos e estatísicos envolvendo dados reais. Por meio da coleta de dados passados e sua análise através de tabelas, gráficos e medidas resumo, podemos responder vários dos questionamentos colocados, especialmente em relação à estabilização da frequência relativa e seu uso como medida de probabilidade. Por exemplo, no caso do basquete há fartos conjuntos de dados disponíveis na internet. Testar as hipóteses da estabilização da frequência é uma atividade que pode ser feita com a ajuda de vários grupos de alunos, utilizando jogadores distintos em temporadas distintas, entre outras possibilidades.

Exemplo 3.4 (Problema do Alvo)**.** *Pretende-se determinar uma probabilidade para cada pontuação que um atleta de tiro com arco-e-flecha pode obter em cada tiro. Há mais de um método para isso? Qual deles é o mais adequado?*

Um método muito comum em livros e gabaritos de concursos para se resolver questões desse tipo é usar a probabilidade geométrica (que será abordada no Ca-

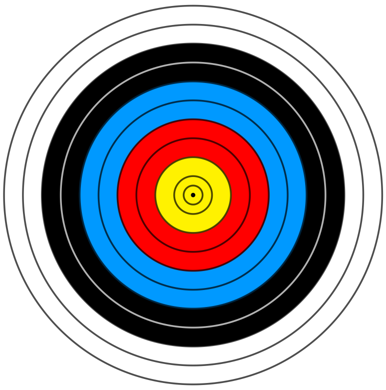

Figura 3.1: Problema do Alvo

pítulo 8), onde se calcula a área correspondente à superfície de certa pontuação na tábua e se divide pela área total do alvo. Note que uma consequência dessa definição de probabilidade é que regiões com áreas iguais terão probabilidades iguais. Esse cálculo reflete a situação real do arco-e-flecha?

Geralmente, um atleta treinará para acertar o mais próximo do centro do alvo, onde vale mais pontos. Para um atleta de alto nível, o que se observa é que é muito mais provável acertar as regiões centrais do que as periféricas. Por exemplo, a probabilidade de acertar a região azul (Figura 3.1) pode ser maior que a de acertar a região branca que está situada na borda, embora a branca tenha área maior que a azul.

Assim como no exemplo da rede de locadoras, o cálculo via probabilidade geométrica produz resultados que não refletem as probabilidades reais. Calcular as probabilidades desse modo faz o esporte parecer que é só questão de sorte.

Uma opção mais adequada para modelar o problema é atribuir probabilidades por meio de frequências relativas. Pode-se, por exemplo, usar um histórico recente de cada atleta para atribuir a distribuição de probabilidades para as pontuações possíveis (de 1 a 10, nas faixas de fora para dentro do alvo). Na Tabela 3.1, mostramos uma situação hipotética.

Pelos dados da tabela, há um levantamento dos últimos 500 tiros do atleta. As probabilidades para cada valor de pontuação i foram atribuídas através da frequência relativa $\dfrac{n_i}{500}$, obtida através da tabela. Para essa situação, a probabilidade de o atleta obter pelo menos 8 pontos com um tiro é dada por:

$$P(\{8, 9, 10\}) = 10,2\% + 10,8\% + 5\% = 26\%$$

Tabela 3.1: Problema do Alvo

Pontuação.	**Nº de Acertos** (n_i)	$f_i = \frac{n_i}{500} \simeq P(\{i\})$
0	3	0,6%
1	12	2,4%
2	23	4,6%
3	41	8,2%
4	44	8,8%
5	76	15,2%
6	80	16%
7	91	18,2%
8	51	10,2%
9	54	10,8%
10	25	5%

3.2 Definição Axiomática

Uma definição geral de probabilidades, que abarque diversos tipos de espaços amostrais, equiprováveis ou não, finitos ou infinitos, deve levar em conta algumas propriedades desejáveis para o modelo matemático que queremos desenvolver para experimentos aleatórios. Além disso, as propriedades da probabilidade já estabelecidas para o caso clássico devem continuar valendo neste modelo.

A definição apresentada abaixo é a definição geral de probabilidades – introduzida por Kolmogorov no século XX – adapatada para o caso de espaços amostrais finitos.

As três condições axiomáticas impostas para definir uma probabilidade P sobre um espaço amostral Ω, como em todo conjunto axiomático, são as mínimas necessárias para que os diversos modelos probabilísticos existentes sejam coerentes com a modelagem de problemas reais. As demais propriedades da probabilidade são deduzidas a partir desses axiomas.

Definição 3.2. *Seja Ω um espaço amostral finito. Considere uma função $P : \mathcal{P}(\Omega) \longrightarrow \mathbb{R}$ definida sobre o conjunto das partes de Ω. Ou seja, P associa a cada evento $A \subset \Omega$ um número real $P(A)$. Tal função é uma probabilidade se:*

(i) $0 \leq P(A) \leq 1 \quad \forall A \subset \Omega$;

(ii) $P(\Omega) = 1$;

(iii) $A, B \subset \Omega, A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$.

Proposição 3.1. *A funções $P : \mathcal{P}(\Omega) \longrightarrow \mathbb{R}$ definidas em 2.1 e 3.1 são probabilidades no sentido axiomático. Em outras palavras, as definições clássica e frequen-*

tista satisfazem os axiomas da Definição 3.2 e, portanto, são casos particulares da definição axiomática.

Demonstração. 1) Começando pela definição clássica, dado um evento $A \subset \Omega$, a razão:

$$P(A) = \frac{n(A)}{n(\Omega)}$$

define uma função $P : \mathcal{P}(\Omega) \longrightarrow \mathbb{R}$. Além disso, temos que:

$$
\begin{aligned}
&(i) \;\; A \subset \Omega \Rightarrow 0 \leq n(A) \leq n(\Omega) \Rightarrow 0 \leq P(A) = \frac{n(A)}{n(\Omega)} \leq 1\\
&(ii) \;\; P(\Omega) = \frac{n(\Omega)}{n(\Omega)} = 1\\
&(iii) \; A, B \subset \Omega, A \cap B = \emptyset \Rightarrow n(A \cup B) = n(A) + n(B)\\
&\Rightarrow \;\; P(A \cup B) = \frac{n(A \cup B)}{n(\Omega)} = \frac{n(A)}{n(\Omega)} + \frac{n(B)}{n(\Omega)} = P(A) + P(B)
\end{aligned}
$$

2) Para o caso da definição frequentista, suponha inicialmente que para todo $A \subset \Omega$, o limite

$$\lim_{n \to \infty} f_n(A) = P(A)$$

existe. Então, esse limite, define uma função $P : \mathcal{P}(\Omega) \longrightarrow \mathbb{R}$. Além disso, temos que:

$$
\begin{aligned}
&(i) \;\; 0 \leq n_A \leq n \Rightarrow 0 \leq f_n(A) = \frac{n_A}{n} \leq 1 \Rightarrow 0 \leq P(A) = \lim_{n \to \infty} f_n(A) \leq 1\\
&(ii) \; n_\Omega = n, \forall n \geq 1 \Rightarrow f_n(\Omega) = 1, \forall n \geq 1 \Rightarrow P(\Omega) = \lim_{n \to \infty} f_n(\Omega) = 1
\end{aligned}
$$

Para o terceiro axioma, note que se $A, B \subset \Omega$ com $A \cap B = \emptyset$, então o número de ocorrências simultâneas de A e B em n experimentos realizados deve ser 0 e, portanto, $n_{A \cup B} = n_A + n_B$. Logo,

$$P(A \cup B) = \lim_{n \to \infty} \frac{n_{A \cup B}}{n} = \lim_{n \to \infty} \frac{n_A}{n} + \lim_{n \to \infty} \frac{n_B}{n} = P(A) + P(B)$$

□

A partir da proposição acima, podemos concluir que todas as propriedades da Probabilidade que serão apresentadas no Capítulo 4 são válidas tanto para a definição clássica quanto para a definição frequentista de probabilidades, pois, como veremos, tais propriedades serão consequência dos axiomas da Definição 3.2.

A proposição que veremos a seguir dá-nos um critério simples para atribuir uma probabilidade P a qualquer espaço amostral finito, satisfazendo a definição axiomática. De forma resumida, qualquer espaço amostral finito pode ter uma probabilidade modelada de modo a corresponder com as características do experimento, bastando para isso atribuir números não negativos (probabilidades) aos eventos elementares tal que sua soma seja 1 e sem a necessidade de verificar a validade dos axiomas da Definição 3.2.

Proposição 3.2. *Considere um espaço amostral finito* $\Omega = \{w_1, w_2, \cdots, w_n\}$*. Sejam* $p_1, p_2, \cdots, p_n$ *números reais não negativos tais que* $p_1 + \cdots + p_n = 1$*. Então, a função* $P : \mathcal{P}(\Omega) \to \mathbb{R}$*, dada por:*

$$\begin{aligned} &P(\{w_i\}) = p_i, \quad \forall i = 1, \cdots, n \quad e \\ &\forall E \subset \Omega;\ P(E) = \sum_{w_i \in E} P(\{w_i\}) \end{aligned}$$

é uma probabilidade. Adicionalmente, se $p_1 = p_2 = \cdots = p_n$ *a função acima define uma probabilidade clássica.*

Demonstração. (i) Como $p_i \geq 0$, então $P(\{w_i\}) = p_i \geq 0$ para todo $1 \leq i \leq n$. Portanto, dado $E \subset \Omega$, temos

$$0 \leq P(E) = \sum_{w_i \in E} P(\{w_i\}) \leq p_1 + \cdots + p_n = 1$$

(ii)

$$\begin{aligned} P(\Omega) &= P(\{w_1, w_2, \cdots, w_n\}) = P(\{w_1\}) + P(\{w_2\}) + \cdots + P(\{w_n\}) \\ &= p_1 + p_2 + \cdots + p_n = 1 \end{aligned}$$

(iii) Sejam $E_1, E_2 \subset \Omega$, onde $E_1 \cap E_2 = \emptyset$. Então, $E_1 \cup E_2 \subset \Omega$ e:

$$\begin{aligned} P(E_1 \cup E_2) &= \sum_{w_i \in E_1 \cup E_2} P(\{w_i\}) = \sum_{w_j \in E_1} P(\{w_j\}) + \sum_{w_k \in E_2} P(\{w_k\}) \\ &= P(E_1) + P(E_2) \end{aligned}$$

A segunda igualdade acima se justifica pelo fato de E_1 e E_2 serem disjuntos.

Por fim, note que se $p_1 = \cdots = p_n = p$, então,

$$\begin{aligned} &1 = p_1 + \cdots + p_n = np \Rightarrow p = \frac{1}{n} \\ &\Rightarrow \forall E \subset \Omega;\ P(E) = \sum_{w_i \in E} P(\{w_i\}) = \sum_{w_i \in E} \frac{1}{n} = n(E) \cdot \frac{1}{n} = \frac{n(E)}{n(\Omega)} \end{aligned}$$

o que corresponde à definição clássica de probabilidade. □

Dessa forma, dado um espaço amostral finito Ω, uma probabilidade P pode ser associada a Ω atribuindo-se aos seus eventos elementares números não negativos cuja soma seja 1 e definindo a probabilidade de um evento E como a soma das probabilidades dos eventos elementares contidos em E.

Exemplo 3.5 (Jogos de Tabuleiro)**.** *Existem muitos jogos de tabuleiro nos quais o número de casas que cada jogador anda em sua vez na rodada é dado pela soma dos pontos obtidos no lançamento de dois dados equilibrados de 6 faces. Nesse* *tipo de jogo, o que importa para o jogador não são exatamente os números que saem nos dados, mas sim a sua soma. Determine a distribuição de probabilidades para o número de casas andadas nesse jogo. Em seguida, calcule a probabilidade de que um jogador ande mais que 9 casas em uma rodada.*

Nessa situação temos

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

pois esses são os resultados possíveis para a soma de dois dados.

Não faz sentido aqui considerar que esses eventos elementares são equiprováveis, pois, por exemplo, só há uma maneira de obter soma 2 (sair 1 em cada dado) mas há quatro modos de obter soma 5 (sair $(1, 4)$ ou $(2, 3)$ ou $(3, 2)$ ou $(4, 1)$).

Então, o mais sensato é fazer $P(\{2\}) = p$ e $P(\{5\}) = 4p$ e continuar usando o mesmo raciocínio para as outras probabilidades, onde obteremos, entre outros resultados, $P(\{11\}) = 2p$. Acrescentando a informação de que a soma dessas probabilidades deve ser 1, temos a equação:

$$\begin{aligned}&P(\{2\}) + P(\{3\}) + \cdots + P(\{12\}) = 1\\ &\Rightarrow p + 2p + 3p + 4p + 5p + 6p + 5p + 4p + 3p + 2p + p = 1 \Rightarrow p = \frac{1}{36}\end{aligned}$$

Assim, obtemos a seguinte distribuição de probabilidades para Ω.

$$\begin{array}{llll} P(\{2\}) = \frac{1}{36} & P(\{3\}) = \frac{2}{36} & P(\{4\}) = \frac{3}{36} & P(\{5\}) = \frac{4}{36} \\ P(\{6\}) = \frac{5}{36} & P(\{7\}) = \frac{6}{36} & P(\{8\}) = \frac{5}{36} & P(\{9\})\frac{4}{36} \\ P(\{10\}) = \frac{3}{36} & P(\{11\}) = \frac{2}{36} & P(\{12\}) = \frac{1}{36} & \end{array}$$

Observamos que a distribuição acima poderia ter sido obtida também analisando todos os 36 pares ordenados como equiprováveis e contando o número de pares que produz cada soma. Cada probabilidade acima seria dada pela razão da definição clássica.

Para responder a segunda pergunta do enunciado usamos o axioma da união (ou a Proposição 3.2). Queremos $P(\{10, 11, 12\})$:

$$\begin{aligned}P(\{10, 11, 12\}) &= P(\{10\}) + P(\{11\}) + P(\{12\})\\ &= \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{6}{36} = \frac{1}{6}\end{aligned}$$

Comentários. O ensino de probabilidade focado na definição clássica pode fazer com que um aluno acostumado a resolver problemas apenas com espaços amostrais equiprováveis responda $3/11$ para a segunda pergunta, pois identificaria 3 "casos favoráveis" para resultados maiores que 9 em um espaço amostral com 11 "casos possíveis". Por isso, reforçamos a importância do trabalho em sala de aula com espaços amostrais não equiprováveis.

3.3 Definição em Espaços Amostrais Infinitos

Na Seção 2.1, apresentamos uma sequência de problemas de probabilidade que não podem ser resolvidos pela definição clássica. Por isso tratamos de alguns deles nas seções anteriores através das definições frequentista e axiomática para espaço finitos. Entretanto, os problemas (c) a (g) daquela seção estão relacionados a espaços amostrais infinitos, para os quais a Definição 3.2 não pode ser aplicada devido à necessidade de lidar com problemas envolvendo a união infinita de eventos.

Nesta seção, veremos que para lidar com espaços infinitos, necessitamos de algumas ferramentas mais avançadas de matemática, com somas infinitas e operações infinitas entre conjuntos. Além disso, o tratamento dado a espaços amostrais infinitos não enumeráveis deve ser mais cuidadoso do que o dado aos enumeráveis e, portanto, a compreensão de conceitos básicos sobre cardinalidade é necessária.

Entretanto, não faz parte dos nossos objetivos aprofundarmo-nos nessas questões. O objetivo principal desta seção é apresentar um embasamento teórico para professores de matemática, fugindo um pouco do contexto do ensino médio, para que se possa trabalhar de forma segura com problemas simples envolvendo espaços infinitos, que podem ser bastante interessantes e acessíveis ao público do ensino médio.

Espaços Enumeráveis

Um conjunto infinito A é dito enumerável quando pode ser colocado em correspondência biunívoca com conjunto dos números naturais, ou seja, quando existe uma bijeção $f : \mathbb{N} \longrightarrow A$. Ainda que não seja nosso objetivo o aprofundamento dessa ideia, é importante, para o estudo de probabilidades, ter uma noção de que tipo de conjunto pode ser classificado como enumerável ou não enumerável. Até porque as ideias de variáveis discretas e contínuas, fortemente presentes na Probabilidade e na Estatística, estão diretamente relacionadas a esses conceitos.

No estudo de cardinalidade de conjuntos, é possível provar que $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ e qualquer um de seus subconjuntos infinitos são enumeráveis. Assim, nos exemplos da Seção 2.1, que reproduzimos abaixo, os espaços referentes aos itens (d), (e) e (g) são enumeráveis.

(c) Ao comprar uma lâmpada, qual é a probabilidade de que ela funcione por mais de cinco anos? $\Omega = \mathbb{R}_+$

(d) Ao lançar uma moeda sucessivamente até obter cara, qual a probabilidade de que essa ocorra após o quinto lançamento? $\Omega = \mathbb{N}^*$.

(e) Ao abordar pessoas na rua para participar de uma pesquisa, qual a probabilidade de que sejam necessárias pelo menos 10 abordagens até que se encontre alguém que aceite? $\Omega = \mathbb{N}^*$.

(f) Escolhendo-se aleatoriamente um ponto do disco com centro na origem do plano cartesiano e raio 1, qual é a probabilidade desse ponto estar no primeiro quadrante? $\Omega = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 \leq 1\}$.

(g) Em um experimento que consiste contar o número de partículas alfa emitidas por minuto por uma substância radioativa, calcule a probabilidade de haver mais de 2 emissões em um minuto. $\Omega = \mathbb{N}$.

Nesse tipo de espaço, a definição axiomática de probabilidade precisa ser estendida, em relação à Definição 3.2.

Definição 3.3. *Seja Ω um espaço amostral enumerável. Considere uma função $P : \mathcal{P}(\Omega) \longrightarrow \mathbb{R}$ definida sobre o conjunto das partes de Ω. Ou seja, P associa a cada evento $A \subset \Omega$ um número real $P(A)$. Tal função é uma probabilidade se:*

(i) $0 \leq P(A) \leq 1 \quad \forall A \subset \Omega$;

(ii) $P(\Omega) = 1$;

(iii) Se $\{A_1, A_2, ...\} = \{A_n\}_{n \geq 1} \subset \mathcal{P}(\Omega)$, *onde* $A_i \cap A_j = \emptyset,\ \forall i, j \geq 1;\ i \neq j$, *então:*

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

Note que a definição acima é praticamente igual à Definição 3.2, com exceção ao item (iii) que nos força a lidar com uniões e somas infinitas, devido à infinitude de Ω. De qualquer forma, a definição para espaços finitos é um caso particular dessa apresentada acima.

Pode-se mostrar ainda que as propriedades da Probabilidade decorrentes da definição axiomática para espaços finitos e as definições de probabilidade condicional e independência, que veremos nos próximos capítulos, permanecem válidas nesse novo modelo.

Apesar de não termos como objetivo o aprofundamento em espaços infinitos, mostraremos um exemplo abaixo cuja resolução pode ser feita no ensino médio, sem a necessidade de aprofundamentos.

Exemplo 3.6. *Ao lançar uma moeda sucessivamente até obter cara, qual a probabilidade de que essa ocorra após o quinto lançamento?*

Nesse exemplo, estamos interessados no número do lançamento em que se obteve o resultado cara, que deve ser um número inteiro positivo. Como não temos como fixar um número que nos dê a certeza que sairá cara até ali, devemos ter $\Omega = \{1, 2, 3, \cdots\} = \mathbb{N}^*$. Nosso evento de interesse é o conjunto $A = \{6, 7, 8, ...\}$, também infinito enumerável.

Agora observe que as probabilidades de se obter cara no j-ésimo lançamento, que representaremos por $P(\{j\})$, seguem um padrão. Denotando cara por K e coroa por C, a ocorrência no j-ésimo lançamento pode ser identificada com a sequência de ocorrências

$$\underbrace{CC\cdots C}_{j-1\ coroas}K$$

Ou seja:

$$P(\{1\}) = P(K) = \frac{1}{2} \quad P(\{2\}) = P(CK) = \frac{1}{4}$$
$$P(\{3\}) = P(CCK) = \frac{1}{8} \quad \cdots \quad P(\{j\}) = P\left(\underbrace{CC\cdots C}_{j-1\ coroas}K\right) = \frac{1}{2^j}$$

Note que a probabilidade assim definida satisfaz o axioma (ii) da definição dada acima, pois

$$\begin{aligned} P(\Omega) &= P(\{1,2,3,\cdots\}) = P(1) + P(2) + P(3) + \cdots \\ &= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = \frac{1/2}{1-1/2} = 1 \end{aligned}$$

onde usamos na última igualdade a fórmula da soma dos termos infinitos de uma PG. Os outros dois axiomas também podem ser verificados para a função P definida acima.

Usando novamente a fórmula da PG, obtemos a probabilidade pedida fazendo:

$$\begin{aligned} P(A) &= P(\{6,7,8,\cdots\}) = P(6) + P(7) + P(8) + \cdots \\ &= \frac{1}{64} + \frac{1}{128} + \frac{1}{256} + \cdots = \frac{1/64}{1-1/2} = \frac{1}{32} \end{aligned}$$

Comentários. O leitor deve notar que o exemplo acima pode ser trabalhado no ensino médio sem a necessidade da apresentação axiomática ou qualquer outro aprofundamento. As ferramentas utilizadas para a solução são todas conhecidas dos alunos desse segmento. Em geral, problemas de probabilidade envolvendo espaços infinitos enumeráveis envolvem somas infinitas (séries) e, por isso, não são adequados para o ensino básico. Mas esse caso é uma exceção, pois envolve soma infinita dos termos de uma PG, que é um tópico conhecido dos alunos.

Exemplos como esse em que as probabilidades formam uma progressão geométrica fazem parte de uma classe especial de distribuições de probabilidade, chamada de distribuição geométrica ou modelo geométrico (em alusão à PG). Na seção 6.3 do Capítulo 6 apresentaremos com mais detalhes esse modelo.

Espaços não Enumeráveis

Da definição dada anteriormente, segue que um conjunto não enumerável é aquele que não pode ser posto em bijeção com o conjunto dos números naturais.

No estudo de cardinalidade de conjuntos, demonstra-se que os conjuntos $\mathbb{R}^n$, para todo $n \geq 1$, não são enumeráveis. Além disso, os subconjuntos de $\mathbb{R}^n$ chamados de bolas – intervalos limitados em $\mathbb{R}$, discos em $\mathbb{R}^2$, esferas em $\mathbb{R}^3$ – e as uniões de bolas também não são enumeráveis, assim como outros tipos de subconjuntos não enumeráveis de $\mathbb{R}^n$, mas que não vamos expor aqui.

Dessa forma, os exemplos (c) e (f) citados mais acima apresentam espaços amostrais infinitos não enumeráveis. A definição de probabilidade nesses espaços não pode ser feita no conjunto $\mathcal{P}(\Omega)$ das partes de Ω. Ela deve ser feita em uma classe especial de subconjuntos de Ω conhecida como σ-álgebra de Ω. A explicação para esse fato é dada em estudos avançados de probabilidade e teoria de medida e não apresentaremos aqui.

Assim, nos limitaremos a dizer aqui que a Definição 3.3 de probabilidade pode ser usada em espaços amostrais infinitos e não enumeráveis, desde que o domínio da função P seja substituído por uma σ-álgebra de Ω. Assim, como o conjunto $\mathcal{P}(\Omega)$ é também uma σ-álgebra para conjuntos finitos e infinitos enumeráveis, a Definição 3.3, com a substituição citada acima, é a forma axiomática generalizada de definir probabilidade, que engloba todas as outras como casos particulares.

O cálculo de probabilidades em espaços infinitos não enumeráveis utiliza, em geral, conceitos matemáticos avançados e ultrapassa os limites propostos para este livro. Entretanto, no ensino básico, há um caso particular cuja exploração é possível no ensino básico e que pode trazer exemplos ricos e bastante interessantes: a chamada Probabilidade Geométrica, que abordaremos com detalhes no Capítulo 8.

3.4 Problemas Propostos

Problema 3.1. *Um levantamento foi feito com 400 alunos de uma escola sobre o número de pessoas residentes na sua moradia. A tabela abaixo mostra o resultado obtido. Suponha que essa amostra de alunos é representativa em relação a todo o corpo discente e considere o experimento de selecionar ao acaso um aluno dessa escola.*

Número de Residentes	2	3	4	5	6	7
Frequência Absoluta	25	105	155	65	32	18

a) Determine o espaço amostral desse experimento e atribua uma probabilidade a cada ponto amostral.
b) Calcule a probabilidade de o aluno selecionado morar em uma casa com pelo menos 5 pessoas.

Problema 3.2. *a) Cite pelo menos três problemas que podem ser modelados por mais de uma definição de probabilidade e analise a adequação de cada solução.*
b) Em particular, seguindo a ideia dos Exemplos 3.2 e 3.3, cite exemplos nos quais não faria sentido atribuir a definição clássica de probabilidades e, nesses casos,

explique por que podemos assumir as hipóteses que embasam a definição frequentista.

Problema 3.3. *Uma moeda equilibrada é lançada cinco vezes e anota-se o número de caras obtidas.*
a) Determine o espaço amostral desse experimento.
b) Defina uma probabilidade para esse espaço.
c) Calcule a probabilidade de que o resultado do experimento seja maior ou igual a 1.
d) Calcule a probabilidade de que o resultado do experimento seja par ou menor que três.

Problema 3.4. *Um dado é viciado de modo que um número par é três vezes mais provável de ocorrer do que um ímpar. Sendo assim, determine:*
a) a probabilidade associada a cada ponto amostral do espaço relativo ao lançamento desse dado;
b) a probabilidade de ocorrer um número par ou primo.

Problema 3.5. *O setor de controle de qualidade de uma indústria está fazendo um teste retirando peças da linha de produção até que se encontre uma peça defeituosa. Sabe-se que a proporção de peças defeituosas nessa indústria é de 1%. O objetivo do teste é anotar o número de tentativas até obter a primeira peça com defeito.*
a) Escreva uma representação para o espaço amostral desse experimento e defina uma probabilidade neste espaço.
b) Calcule a probabilidade de que se obtenha uma peça defeituosa após centésima tentativa.
c) Calcule a probabilidade de que se obtenha uma peça com defeito em uma tentativa de número par.

Capítulo 4

Propriedades da Probabilidade e Aplicações

"É notável que esta ciência, que surgiu da análise dos jogos de azar, tenha se tornado o mais importante objeto do conhecimento humano."

– Pierre Simon Laplace

Neste capítulo apresentaremos as principais propriedades da probabilidade e suas importantes aplicações na resolução de problemas e no desenvolvimento do raciocínio probabilístico.

Provaremos tais propriedades como uma consequência da Definição axiomática 3.2 para espaços amostrais finitos. Mas enfatizamos que as mesmas demonstrações podem ser estendidas para espaços amostrais infinitos, inclusive os não enumeráveis.

Uma vantagem de provar essas propriedades como consequência da definição axiomática é que qualquer função $P : \mathcal{P}(\Omega) \longrightarrow \mathbb{R}$ (ou com domínio numa σ-álgebra para o caso não enumerável) que satisfizer os axiomas de probabilidade terá essas importantes propriedades como consequência.

Das demonstrações que fizemos no Capítulo 3, seguirá como consequência que tanto a definição clássica quanto a frequentista de probabilidades têm as propriedades que provaremos a seguir. Mostraremos ainda no próximo capítulo que a probabilidade condicional é também uma probabilidade no sentido axiomático. Ou seja, mesmo quando estivermos trabalhando com condicionais, podemos aplicar essas propriedades.

No ensino básico, alguns textos didáticos provam essas propriedades apenas para o caso da definição clássica, que é um caso particular. Mas há também textos que as deduzem dos axiomas. No nosso entendimento, a segunda abordagem traz mais vantagens porque, em primeiro lugar, engloba uma situação maior de casos e não apenas um caso particular. Quando você demonstra apenas para a definição clássica, terá que demonstrar de novo toda vez que abordar um novo tipo de probabilidade.

Em segundo, porque evita a confusão que se faz quando os axiomas da Definição 3.2 são provados como propriedades da definição clássica. Por exemplo, quando se prova por essa definição que $P(\Omega) = 1$ como uma propriedade, fica mais difícil para o aluno entender essa condição como um axioma necessário para *qualquer probabilidade*, ou seja, uma verdade assumida que todas as probabilidades devem satisfazer.

Por fim, chamamos a atenção para alguns problemas que apresentaremos neste capítulo para os quais as propriedades da probabilidade têm um papel fundamental. Veremos que sem as propriedades, alguns problemas ficariam muito difíceis de serem resolvidos. Além disso, existem alguns raciocínios intrínsecos às propriedades que estão presentes na cerne de alguns problemas e até em usos cotidianos das probabilidades, como as ideias do complementar, da união e da inclusão de eventos. Dessa forma, a compreensão do significado de cada propriedade é um dos pontos centrais do desenvolvimento do raciocínio probabilístico.

4.1 Propriedades

Teorema 4.1 (Propriedades da Probabilidade)**.** *Se $P : \mathcal{P}(\Omega) \longrightarrow [0,1]$ é uma probabilidade no sentido da Definição 3.2, então:*

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

(ii) *Se $A_1, A_2, ..., A_n$ são eventos mutuamente excludentes, ou seja, $A_i \cap A_j = \emptyset$ para todo i, j tais que, $1 \leq i < j \leq n$, então*

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

(iii) *Para quaisquer $A, B \subset \Omega$, tem-se*

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

(iv) $P(A) = 1 - P\left(A^C\right), \forall A \subset \Omega$

(v) $A \subset B \subset \Omega \Rightarrow P(A) \leq P(B)$

Demonstração. (i) Dado um evento $A \subset \Omega$ qualquer, temos $A \cap \emptyset = \emptyset$ e $A \cup \emptyset = A$. Assim, seque do axioma da união que:

$$P(A) = P(A \cup \emptyset) = P(A) + P(\emptyset) \Longrightarrow P(\emptyset) = 0$$

(ii) Note que os eventos A_1 e $B = A_2 \cup ... \cup A_n$ são mutuamente excludentes. Então segue do axioma da união que

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P(A_1 \cup B) = P(A_1) + P(B) = P(A_1) + P\left(\bigcup_{i=2}^{n} A_i\right)$$

Utilizando o mesmo argumento acima, obtemos de forma recursiva

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P(A_1) + P\left(\bigcup_{i=2}^{n} A_i\right) = P(A_1) + P(A_2) + P\left(\bigcup_{i=3}^{n} A_i\right) = \ldots$$
$$= P(A_1) + P(A_2) + \ldots + P(A_n) = \sum_{i=1}^{n} P(A_i)$$

(iii) O evento $A \cup B$ pode ser reescrito através da união disjunta $A \cup (B - A)$. Por outro lado, temos também, de forma disjunta: $B = (B \cap A) \cup (B - A)$ (Figura 4.1). Dessa forma, aplicando novamente o axioma da união, temos

$$\begin{cases} P(A \cup B) = P(A \cup (B - A)) = P(A) + P(B - A) \\ P(B) = P(B \cap A) + P(B - A) \end{cases}$$
$$\Longrightarrow P(A \cup B) = P(A) + P(B) - P(B \cap A)$$

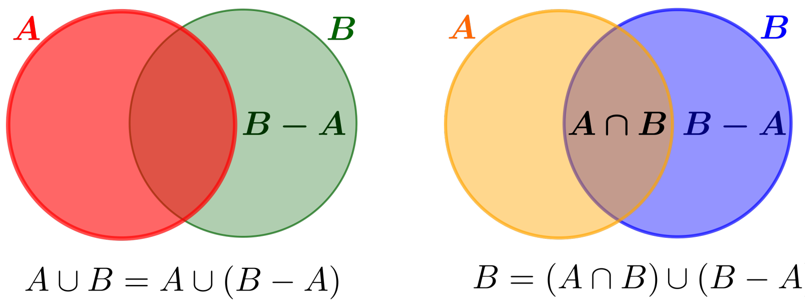

Figura 4.1: Propriedade - União de Eventos

(iv) Para todo $A \subset \Omega$, A e A^C são mutuamente excludentes. Logo,

$$A \cup A^C = \Omega \Longrightarrow P(A) + P\left(A^C\right) = P\left(A \cup A^C\right) = P(\Omega) = 1$$
$$\Longrightarrow P(A) = 1 - P\left(A^C\right)$$

(v)

$$A \subset B \Longrightarrow B = A \cup (B - A) \Longrightarrow P(B) = P(A) + P(B - A) \geq P(A)$$

□

4.2 Aplicações

Como dissemos, as propriedades acima são bastante úteis na resolução de problemas em espaços finitos. Além disso, elas contêm um significado probabilístico particularmente relacionado com a descrição de situações que, para alguns, podem parecer contraintuitivas, como veremos nos exemplos a seguir.

Exemplo 4.1 (Problema dos Aniversários)**.** *Em um grupo com* 30 *pessoas, qual é a probabilidade de que pelo menos duas delas façam aniversário na mesma data (dia e mês)? A partir de quantas pessoas no grupo essa probabilidade passa de* 90%*?*

Consideremos o evento A: "pelo menos duas pessoas do grupo fazem aniversário na mesma data".

Se assumirmos as datas de nascimento do ano como equiprováveis (e, portanto, aplicando a definição clássica), devemos contar quantas combinações de datas podem ter as 30 pessoas do grupo, para obter o número de casos possíveis; e quantas dessas combinações resultam em duas ou mais coincidências de datas, que são os casos favoráveis à ocorrência de A. Mas note que esta última contagem seria bastante trabalhosa, beirando o inviável.

Sendo assim, um modo mais fácil de resolver este problema é calcular a probabilidade de o evento A não ocorrer, e, a seguir, usamos a Propriedade (iv) para calcular $P(A)$.

Assim consideramos o evento A^C: "não há duas pessoas no grupo que façam aniversário na mesma data" ou, de forma equivalente, "todas as pessoas do grupo fazem aniversário em datas distintas".

Para obter $n(\Omega)$, podemos aplicar o Princípio Fundamental da Contagem. Considerando um ano com 365 dias, o total de combinações de datas de aniversário num grupo de 30 pessoas é dada por:

$$n(\Omega) = \underbrace{365 \cdot 365 \cdots 365}_{30\ fatores} = 365^{30}$$

Por outro lado, nosso evento A consiste em combinções de 30 datas de aniversário distintas. Assim, obtemos $n\left(A^C\right)$ fazendo:

$$n\left(A^C\right) = 365 \cdot 364 \cdot 363 \cdots (365 - (30 - 1)) = \frac{365!}{(365 - 30)!} = \frac{365!}{335!}$$

Logo,

$$P\left(A^C\right) = \frac{\frac{365!}{335!}}{365^{30}} \simeq 0,29$$

$$P(A) = 1 - P\left(A^C\right) \simeq 1 - 0,29 = 0,71 = 71\%$$

Usando o mesmo raciocínio acima podemos generalizar o resultado para um grupo de n pessoas. Desse modo obtemos:

$$P(A) = 1 - \frac{365!}{365^n \cdot (365 - n)!}$$

Atribuindo valores a n, verificamos que para $n > 40$ a probabilidade passa de 90%. Na Tabela 4.1, temos os cálculos das probabilidades de pelos menos duas pessoas fazerem aniversário no mesmo dia em um grupo de n pessoas, para vários valores de n.

Tabela 4.1: Problema do Aniversário

n	**Probabilidade** $\simeq$
10	$11,7\%$
20	$41,1\%$
30	$70,6\%$
40	$89,1\%$
50	$97,0\%$
60	$99,4\%$

Comentários. Este problema costuma ser popular quando apresentado em sala de aula, principalmente se a turma for atraída antes da resolução fazendo uma simulação com suas datas de aniversários. Se a turma tiver entre 30 e 40 alunos, as chances de encontrar dois aniversários iguais são grandes, como vimos. Os estudantes costumam se envolver e ficar bastante surpresos com o resultado que é bem contraintuitivo. E caso se tenha o azar de não encontrar dois alunos que atendam às condições do problema, o professor deve aproveitar o momento para discutir a questão de que um evento com probabilidade alta não ocorrerá necessariamente, o que remete à relação entre probabilidade teórica e resultado observado, abordada no Capítulo 2. Por outro lado, deve-se chamar a atenção para o fato de que ao repetir o experimento em muitas turmas com 30 alunos, devemos verificar sua ocorrência em aproximadamente 70% delas.

Devemos ressaltar também que neste problema os resultados numéricos tem um papel muito importante, e, como o leitor deve ter notado, não é possível obtê-los sem a ajuda de uma boa calculadora científica. Calculadoras, *softwares* e aplicativos para *smartphone*, conforme veremos ao longo de todo este livro, são grandes aliados ao ensino de Probabilidade. Além de nos proporcionar resultados numéricos importantes para a análise do problema (inviáveis de se obter manualmente), os recursos digitais nos poupam tempo com cálculos tediosos e nos permitem focar no que realmente importa para o problema.

Esse exemplo também fornece uma ótima oportunidade para utilizar a propriedade do evento complementar, uma vez que essa facilita bastante a resolução. Aliás, um bom exercício tanto para o leitor quanto para os alunos é tentar resolver o problema sem o uso dessa propriedade. Fica praticamente impossível! Mas a tentativa de resolver é importante para que se compreenda o quanto a propriedade do complementar é necessária.

É possível que o leitor julgue estranho o fato de termos resolvido esse exemplo usando um conceito de Análise Combinatória (Princípio Fundamental da Contagem), dado o título do livro. Aproveitamos para ressaltar que não é nosso objetivo condenar o uso dessa importante ferramenta. Nem mesmo deixar de usá-la em problemas de Probabilidade. Apenas queremos destacar que há muitos conceitos e propriedades a serem explorados e que são fundamentais para o desenvolvimento do raciocínio probabilístico. Caso o professor ainda não tenha ensinado Combinatória a sua turma, ainda há a possibilidade de resolvermos o problema aplicando uma ferramenta probabilística: o Teorema do Produto (Teorema 5.1), que veremos no Capítulo 5. Nesse caso, a probabilidade do complementar pode ser calculada pelo produto das probabilidades de não escolher uma data repetida, aluno por aluno. Ou seja:

$$P\left(A^C\right) = \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{365-(30-1)}{365} \simeq 0,29$$

$$P(A) = 1 - P\left(A^C\right) \simeq 1 - 0,29 = 0,71 = 71\%$$

Exemplo 4.2. *Dois dispositivos eletrônicos A e B são colocados em teste por um mês de forma ininterrupta. A probabilidade de que ocorra uma falha no dispositivo A, durante esse período, é de* $1/25$*, no dispositivo* B *é de* $1/120$ *e, em ambos,* $1/400$*. Qual é a probabilidade de que:*

a) pelo menos um dos dispositivos apresente uma falha no período mencionado?

b) nenhum deles apresente falha?

c) apenas o dispositivo A apresente falha?

$a)$ Consideremos os eventos A: "ocorre falha no dipositivo A" e B: "ocorre falha no dipositivo B". Para obter a probabilidade pedida, devemos perceber que "ocorrer falha em pelo menos um dos dispositivos" é o mesmo que "ocorrer falha em A ou ocorrer falha em B", o que é equivalente a ocorrer o evento união $A \cup B$.

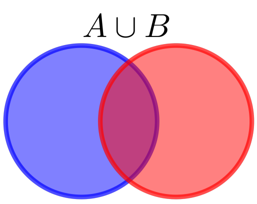

Figura 4.2: Evento "pelo menos um″

Ou seja, queremos calcular a probabilidade de $A \cup B$. A partir daí, basta utilizar a Propriedade (iii):

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{25} + \frac{1}{120} - \frac{1}{400} = \frac{11}{240} \simeq 0,046 = 4,6\%$$

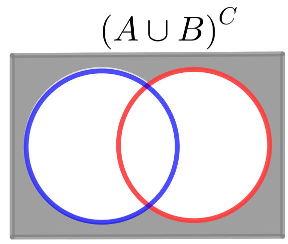

Figura 4.3: Evento “nem A nem B''

$b)$ Para este item, observe que “nenhum apresentar falha” é equivalente a “nem A nem B apresentam falha”.

Ou seja, devemos calcular a probabilidade do evento $(A \cup B)^C$: “não ocorre falha no dispositivo A e não ocorre falha no dispositivo B”. Uma vez que o item $a)$ já foi resolvido, basta usar a propriedade do evento complementar (Propriedade (iv)).

$$P\left((A \cup B)^C\right) = 1 - P\,(A \cup B) \simeq 1 - 0,046 = 0,954 = 95,4\%$$

$c)$ Este item é útil para destacar a diferença entre “ocorrer A” e “ocorrer apenas A”. No primeiro caso temos o evento A, que pode ocorrer ou não simultaneamente com B, já no segundo temos o evento $A - (A \cap B)$ (ou $A - B$).

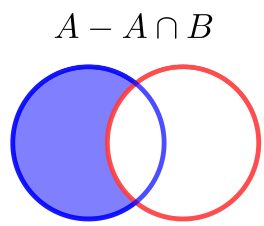

Figura 4.4: Evento “ocorrer apenas A''

Daí, basta fazer:

$$P\,(A - B) = P(A) - P\,(A \cap B) = \frac{1}{25} - \frac{1}{400} = \frac{3}{80} = 0,0375 = 3,75\%$$

A primeira igualdade acima, apesar de ser bastante intutivia (veja a Figura 4.4), pode ser provada de modo análogo à demonstração da Propriedade (iii). Repare que o conjunto A pode ser escrito como a união disjunta entre $(A - B)$ e $(A \cap B)$.

Assim:

$$A = (A - B) \cup (A \cap B) \Longrightarrow P(A) = P\left((A - B) \cup (A \cap B)\right)$$
$$\Longrightarrow P(A) = P(A - B) + P(A \cap B) \Longrightarrow P(A - B) = P(A) - P(A \cap B)$$

Comentários. Além de ser um bom exemplo para se trabalhar propriedades da probabilidade, é possível aproveitá-lo para revisitar o tópico de Operações entre Conjuntos. Aliás, usar diagramas de Venn facilita bastante a visualização e compreensão da resolução do problema e também das propriedades da Probabilidade.

Outro ponto importante é sobre a interpretação das frases dadas no problema como conceitos probabilísticos (ou de teoria dos conjuntos). A compreensão das ideias de "pelo menos um", "nenhum" e "apenas" e sua relação com os conceitos matemáticos pode ser um dos focos durante a discussão desse problema.

Exemplo 4.3. *Um instituto de meteorologia divulgou sua previsão do tempo de amanhã para uma cidade litorânea, estimando uma probabilidade de 70% para ocorrência de chuvas e uma probabilidade de 60% para a ocorrência de ventos fortes.*

a) É possível descartar a ocorrência simultânea das duas condições climáticas?

b) Qual são as probabilidades máxima e mínima para que ambas ocorram simultaneamente?

c) Se assumirmos a hipótese de que essas condições ocorrem de forma independente, qual é a probabilidade de que pelo menos uma das duas ocorra?

$a)$ Considere os eventos A: "ocorrer chuva" e B: "ocorrer ventos fortes". Se assumirmos a hipótese de que A e B são eventos que não ocorrem simultaneamente, ou seja $P(A \cap B) = 0$, então teremos

$$P(A \cup B) = P(A) + P(B) = 0,7 + 0,6 = 1,3$$

e pelo Axioma (i) da Definição 3.2, nenhuma probabilidade pode ser maior que 1. Portanto, podemos afirmar que há possibilidade de ocorrência simultânea das duas condições climáticas ($P(A \cap B) > 0$), não sendo possível descartá-la.

$b)$ Avaliar a probabilidade de ocorrência simultânea equivale a avaliar a probabilidade $P(A \cap B)$. O valor mínimo que ela pode assumir deve ser o necessário para que $P(A \cup B)$ seja menor que ou igual a 1. Por exemplo, não é possível termos $P(A \cap B) = 0,2$, pelo mesmo argumento utilizado no item $a)$. Analisando a equação dada pela propriedade da união (Propriedade (iii)), temos:

$$1 \geq P(A \cup B) = 0,7 + 0,6 - P(A \cap B) \Longrightarrow P(A \cap B) \geq 0,3$$

ou seja, o menor valor possível para a ocorrência simultânea é $0,3$.

Para avaliar o valor máximo de $P(A \cap B)$, observe que a propriedade (v) traz como consequência:

$$A \cap B \subset A \quad \text{e} \quad A \cap B \subset B \Longrightarrow P(A \cap B) \leq P(A) \quad \text{e} \quad P(A \cap B) \leq P(B)$$

Ou seja, a ocorrência simultânea das duas condições climáticas deve ter probabilidade menor ou igual às probabilidades de ocorrência de cada uma delas (o que parece ser bastante intuitivo). Assim concluímos que o valor máximo pedido é $0,6$, pois é o menor dos números entre $P(A)$ e $P(B)$. Dessa forma devemos ter:

$$0,3 \leq P(A \cap B) \leq 0,6$$

Ou seja, temos entre 30% e 60% de chances de ocorrência de chuvas com ventos fortes.

$c)$ A independência de eventos será abordada com mais detalhes no Capítulo 5. Entretanto, para fins de aplicação neste problema, quando dois eventos são independentes temos a relação $P(A \cap B) = P(A)P(B)$. Dessa forma, assumindo a independência, a probabilidade de que pelo menos uma das condições ocorra ($P(A \cup B)$) é dada por

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,7 + 0,6 - 0,7 \cdot 0,6 = 0,88$$

Comentários. À primeira vista este problema pode parecer semelhante ao anterior, mas ele exige um raciocínio um pouco mais sofisticado. São várias habilidades com o uso das propriedades e da definição que são exigidas e que, em geral, são subestimadas nos livros didáticos de ensino médio. Por exemplo, no item b), fizemos análises baseadas nas propriedades dos conjuntos, definição axiomática e usamos duas propriedades da probabilidade para poder estabelecer uma faixa de variação para a probabilidade de ocorrência simultânea dos eventos. Usamos também a Definição de Independência de Eventos (Definição 5.2), que veremos mais à frente, mas sua inserção neste exemplo ressalta formas distintas de lidar com as operações entre conjuntos, sua "traduação" para a linguagem probabilística (pelo menos um, ocorrência simultânea) e sua relação com as habilidades relacionadas ao uso das propriedades.

Exemplo 4.4 (Problema da Linda)**.** *Imagine uma mulher chamada Linda, de 31 anos de idade, solteira, sincera e muito inteligente. Cursou filosofia na universidade. Quando estudante, preocupava-se profundamente com discriminação e justiça social e participou de protestos contra armas nucleares. Considerando essas informações, qual das afirmações a seguir você julga ser a menos provável? Há alguma justificativa matemática para a sua resposta?*

A - Linda participa do movimento feminista;

B - Linda é bancária e participa do movimento feminista;

C - *Linda é bancária.*

Esse problema faz parte um experimento realizado por Daniel Kahneman e Alan Tversky, descrito em [18].

Observe que a afirmação B nada mais é que a ocorrência simultânea das afirmações A e C. Identificando cada afirmação com um evento, temos que $B = A \cap C$, ou seja,

$$B = A \cap C \subset A \quad \text{e} \quad B = A \cap C \subset C$$

Portanto, e pela propriedade dos subconjuntos (Propriedade (v)), temos que

$$P(B) = P(A \cap C) \leq P(A) \quad \text{e} \quad P(B) = P(A \cap C) \leq P(C)$$

Logo, a afirmação B é a menos provável.

Comentários. Quando Kahneman e Tversky fizeram o experimento com um grupo de pessoas, a maioria respondeu que a afirmação B era mais provável que a C. Isso aconteceu porque o enunciado teve a intenção de descrever uma mulher que aparentemente tem grande tendência a ser feminista, de acordo com esteriótipos criados pelos participantes da pesquisa. Mas note que, independentemente de conhecimento probabilístico, a opção B é mais restritiva que as opções A e C e, portanto, deveria ser avaliada como menos provável. Ainda assim, o experimento mostrou que a escolha mais racional pode ser superada por uma escolha baseada em concepções prévias.

Ao ler "Linda participa de movimento feminista" e "Linda é bancária e participa do movimento feminista" nas opções, em conjunto com a concepção prévia dos participantes, há a falsa confirmação (viés de confirmação) da expectativa de que ela seja feminista. Daí, algumas pessoas tomam isso como certo e esquecem que ao dizermos que "Linda é bancária" estamos englobando o fato de Linda ser bancária e feminista mais o fato de ela ser bancária e não ser feminista.

Diversos trabalhos citados em [18] relatam enganos desse tipo por profissionais da medicina, economia, engenharia e muitas outras. Por isso, acreditamos ser importante o desenvolvimento da compreensão conceitual de diversos aspectos da Probabilidade. Esse tipo de problema puramente conceitual pode colaborar para o desenvolvimento nesse sentido.

4.3 Problemas Propostos

Problema 4.1. *Na modalidade de pôquer conhecida como Texas Hold'em, cada jogador recebe no início da rodada duas cartas de uma baralho comum. Calcule a probabilidade de:*
a) um jogador receber duas cartas do mesmo naipe ou um par (par de dois, par de reis etc.).
b) um jogador não receber um par de figuras (valete, dama, rei).

Problema 4.2. *Dados os eventos $A, B, C \subset \Omega$:*
a) Mostre que $P(B) = 0 \Rightarrow P(A \cap B) = 0$ e $P(A \cup B) = P(A)$.
b) Mostre que $P(B) = 1 \Rightarrow P(A \cap B) = P(A)$ e $P(A \cup B) = 1$.
c) Se $P(A) = 0,4$, $P(B) = p$ e $P(A \cup B) = 0,7$, determinar os valores máximo e mínimo de p.
d) Se $P(A) = 0,4$, $P(B) = 0,8$ e $P(A \cap B) = 0,3$, determinar $P\left(A^C \cap B^C\right)$ e $P\left(A^C \cup B^C\right)$.
e) Mostre que

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

Problema 4.3. *Dez pessoas serão escolhidas aleatoriamente na rua. Calcule a probabilidade de que pelo menos duas tenham nascido no primeiro trimestre.*

Problema 4.4. *Considere o experimento em que dois dados equilibrados de seis faces são lançados e anota-se a diferença em módulo dos pontos obtidos.*
a) Determine o espaçco amostral deste experimento e defina sobre ele uma probabilidade.
b) Calcule as probabilidades dos seguintes eventos: (i) o resultado obtido é maior que 3 ou par; (ii) o resultado obtido é maior que 1.

Problema 4.5. *Com os dados do Exemplo 4.3, suponha que $P(A \cap B) = 0,5$. Calcule a probabilidade de:*
a) pelo menos uma das duas condições ocorrer;
b) nehuma delas ocorrer;
c) apenas a condição de ventos fortes ocorrer.
d) apenas uma das duas condições ocorrer.

Capítulo 5

A Condicional na Realidade e a Árvore das Probabilidades Totais

"A medicina é uma ciência da incerteza e uma arte da probabilidade."

– William Osler

Os conceitos de probabilidade condicional e independência são dois pilares centrais para o desenvolvimento do pensamento probabilístico elementar. Conforme comentado anteriormente, o excesso de exercícios com foco combinatório, tradicionalmente priorizados no ensino básico, acaba tornando a abordagem desses temas como secundária e perde-se uma oportunidade de desenvolver habilidades importantes que são exigidas na sua compreensão. Além disso, são deixadas de lado aplicações importantes e elementares da Teoria das Probabilidades que surgem tanto na Ciência, em diversas áreas, quanto em situações cotidianas.

Neste capítulo, apresentaremos a definição de probabilidade condicional, explicitanto o conceito intuitivo que a embasa e discutindo situações-problemas reais que são afetadas diariamente pela má compreensão desse tema. Mais que o cálculo e a definição matemática, mostraremos que a compreensão conceitual de probabilidade condicional (ou a falta dela) tem consequência direta na vida das pessoas.

Da mesma forma, discutiremos a ideia de independência de eventos, cujo conceito é também fundamental para o desenvolvimento do raciocínio probabilístico elementar. Mostraremos situações nas quais erros de avaliação que afetam a vida de pessoas foram cometidos pela comparação entre eventos como se fossem independentes, quando não eram.

Por fim, apresentaremos três teoremas fundamentais da Teoria das Probabilidades: o Teorema do Produto, o Teorema da Probabilidade Total e o Teorema de Bayes. Mostraremos várias aplicações das ideias contidas nesses teoremas em situações reais, com aplicação da importante técnica conhecida como diagrama de árvore, cuja base teórica dá-se no segundo teorema citado acima.

5.1 Conceitos, Definições e Aplicações

A conhecida ideia da probabilidade condicional, definida matematicamente abaixo, pode ser aplicada a qualquer tipo de espaço amostral: finitos, infinitos, equiprováveis ou não. A ideia intuitiva baseia-se no fato de que a probabilidade de um evento pode ser alterada (ou não) se recebermos a informação de que outro evento ocorreu. Por outro lado, a ideia de independência, diretamente relacionada à condicional, ilustra o fato de que a ocorrência de um evento não afeta nossas expectativas sobre a ocorrência de outro.

Definição 5.1 (Probabilidade Condicional)**.** *Dados dois eventos A e B de um espaço amostral e uma probabilidade P, onde $P(A) \neq 0$, a probabilidade condicional de B dado A é definida por:*

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

Observações: 1) Para captar a ideia intuitiva da definição acima, observe que quando recebemos a informação de que A ocorreu, devemos avaliar a probabilidade de B ocorrer apenas dentro do contexto A (e fora de A^C). Ou seja, nossa medida de probabilidade deve ser relativa à medida total do conjunto A e não de todo Ω. Por isso avaliamos a ocorrência simultânea de B e A ($P(A \cap B)$) em relação à ocorrência de A, dividindo por $P(A)$.

2) Em particular, para espaços amostrais finitos e equiprováveis temos:

$$P(B|A) = \frac{n(A \cap B)}{n(A)} \quad (*),$$

já que:

$$P(B|A) \overset{(**)}{=} \frac{P(A \cap B)}{P(A)} = \frac{\frac{n(A\cap B)}{n(\Omega)}}{\frac{n(A)}{n(\Omega)}} = \frac{n(A \cap B)}{n(A)}$$

Veja que esse caso particular $(*)$ é muito mais intuitivo que o caso geral (Definição 5.1), pois dada a ocorrência do evento A, devemos considerar apenas seus elementos como casos possíveis; e os casos favoráveis passam a ser os elementos de B que estão também em A, ou seja, $A \cap B$.

Por esse motivo, para trabalhar com alunos de ensino básico, uma opção que pode tornar mais fácil a compreensão da ideia envolvida neste conceito seria seguirmos a seguinte sequência didática:
(i) definir Probabilidade Condicional para conjuntos finitos e equiprováveis usando a igualdade $(*)$;
(ii) apresentar a igualdade $(**)$ como uma consequência da definição (fazendo o percurso contrário da demonstração) nesse tipo de espaço;
(iii) definir Probabilidade Condicional para conjuntos quaisquer (Definição 5.1)

através da igualdade $(**)$, como uma extensão da definição inicial, mas sem deixar de discutir seu o aspecto intuitivo apontado na Observação 1.

Exemplo 5.1. *No Exemplo 4.2, apresentamos uma situação na qual dois dispositivos eletrônicos A e B em teste têm probabilidade de falha de $1/25$ no dispositivo A, de $1/120$ no dispositivo B e, em ambos, de $1/400$. Sabendo que o dispositivo A falhou durante o período de teste, qual é a probabilidade de B falhar?*

Sabendo que A falhou, aplicamos a Definição 5.1 para obter:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/400}{1/25} = \frac{1}{16}$$

Note que a probabilidade de falha em B aumentou mais de 7 vezes com a informação da falha de A.

A proposição a seguir nos mostra que a probabilidade condicional é uma probabilidade no sentido axiomático como qualquer outra, e, portanto, possui todas as propriedades decorrentes. Isso pode facilitar o cálculo de probabilidades como $P\left(B^C|A\right)$ e $P(B \cup C|A)$ usando as propriedades que provamos no Capítulo 4.

Proposição 5.1. *Fixado um evento $A \subset \Omega$ tal que $P(A) \neq 0$, a função $P(X|A) : \mathcal{P}(\Omega) \to \mathbb{R}$ satisfaz os axiomas da Definição 3.2.*

Demonstração. Nosso objetivo é provar que:

$$\begin{aligned}
&(i) \;\; 0 \leq P(X|A) \leq 1 \quad \forall \; X \subset \Omega; \\
&(ii) \;\; P(\Omega|A) = 1; \\
&(iii) \;\; X, Y \subset \Omega; \;\; X \cap Y = \emptyset \Longrightarrow P(X \cup Y|A) = P(X|A) + P(Y|A)
\end{aligned}$$

(i) Dado $X \subset \Omega$, como $P(X \cap A) \geq 0$ e $P(A) > 0$, então:

$$P(X|A) = \frac{P(X \cap A)}{P(A)} \geq 0$$

Além disso,

$$X \cap A \subset A \Longrightarrow P(X \cap A) \leq P(A) \Longrightarrow P(X|A) = \frac{P(X \cap A)}{P(A)} \leq 1$$

Portanto, $0 \leq P(X|A) \leq 1$.

(ii) Nesse caso temos diretamente

$$P(\Omega|A) = \frac{P(\Omega \cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1$$

(iii) Dados $X, Y \subset \Omega$ tais que $X \cap Y = \emptyset$, temos que

$$(X \cap A) \cap (Y \cap A) = X \cap Y \cap A = \emptyset.$$

Ou seja, $(X \cap A)$ e $(Y \cap A)$ são disjuntos. Daí segue que:

$$\begin{aligned} P(X \cup Y|A) &= \frac{P((X \cup Y) \cap A)}{P(A)} = \frac{P((X \cap A) \cup (Y \cap A))}{P(A)} \\ &= \frac{P(X \cap A) + P(Y \cap A)}{P(A)} = \frac{P(X \cap A)}{P(A)} + \frac{P(Y \cap A)}{P(A)} \\ &= P(X|A) + P(Y|A) \end{aligned}$$

□

Exemplo 5.2. *No Problema dos Aniversários (Problema 4.1), qual passa a ser a probabilidade de que pelo menos duas pessoas, entre 30, façam aniversário na mesma data, se recebermos a informação de que não há ninguém no grupo nascido em janeiro?*

Considere os eventos A: "pelo menos duas pessoas do grupo fazem aniversário na mesma data" e B: "nenhuma pessoa do grupo nasceu em janeiro". O que se pede é a probabilidade condicional $P(A|B)$. De acordo com a proposição anterior e com o item (iv) do Teorema 4.1, podemos concluir que $P(A|B) = 1 - P\left(A^C|B\right)$. Sabendo que o problema é modelado pela definição clássica de probabilidade, temos:

$$P\left(A^C|B\right) = \frac{n\left(A^C \cap B\right)}{n(B)}$$

Da mesma forma como fizemos no Exemplo 4.1, concluímos que $n(B) = 334^{30}$, pois há 334 datas (sem o mês de janeiro) disponíveis para 30 pessoas. Além disso, temos

$$\begin{aligned} n\left(A^C \cap B\right) = 334 \cdot 333 \cdots 305 &\Longrightarrow P\left(A^C|B\right) = \frac{334 \cdot 333 \cdots 305}{334^{30}} \simeq 0,26 \\ &\Longrightarrow P(A|B) \simeq 0,74 \end{aligned}$$

Observe que, como esperado, a probabilidade de que duas pessoas façam aniversário na mesma data aumenta quando recebemos a informação de que não há ninguém nascido no mês de janeiro.

Esse exemplo (e muitos outros presentes em aplicações) torna explícito um ponto importante provado na proposição anterior: a probabilidade condicional é uma lei que se comporta da mesma forma, em termos de propriedades, que qualquer probabilidade não condicional, o que amplia nosso leque de ferramentas para lidar com problemas envolvendo condicionais.

Definição 5.2. *Dados dois eventos A e B de um espaço amostral, e $P(B) \neq 0$, dizemos que A é independente de B quando $P(A|B) = P(A)$*

Observe que a definição acima é bastante intuitiva. A expressão $P(A|B) = P(A)$ expressa matematicamente situações nas quais a ocorrência de um evento B não afeta a probabilidade de ocorrência de A. Por exemplo, no lançamento de uma moeda e de um dado equilibrados em sequência, a informação de que saiu cara na moeda não afeta as probabilidades de ocorrência no dado.

Proposição 5.2. *Dados os eventos $A, B \subset \Omega$ tais que $P(A), P(B) \neq 0$, temos:*

A é independente de $B \iff B$ é independente de $A \iff P(A \cap B) = P(A)P(B)$

Demonstração. Por um lado temos

$$A \text{ é independente de } B \iff \frac{P(A \cap B)}{P(B)} = P(A|B) = P(A)$$

$$\iff P(B) = \frac{P(A \cap B)}{P(A)} = P(B|A) \iff B \text{ é independente de } A$$

Por outro lado

$$B \text{ é independente de } A \iff \frac{P(A \cap B)}{P(A)} = P(B|A) = P(B)$$

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

□

Observações: 1) A primeira equivalência da Proposição acima nos diz que se $P(A), P(B) \neq 0$ e houver independência de um dos lados, então haverá do outro. Quando isso ocorrer, poderemos usar a expressão "A e B são independentes entre si".

2) Muitos livros definem independência de eventos por meio da igualdade

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

e apresentam $P(A|B) = P(A)$ como consequência. Essa definição tem a vantagem de ser mais geral que a anterior, pois inclui os casos em que $P(A) = 0$ ou $P(B) = 0$. Acreditamos inclusive que ela seja a mais adequada para cursos de nível superior e de formação continuada para professores de Matemática. Entretanto, para o ensino básico, entendemos que a definição pela qual optamos neste texto seja a mais adequada por ser mais intuitiva e também porque os eventos com probabilidade 0 que não sejam o conjunto vazio dificilmente aparecem nos problemas tratados neste segmento. Assim, fica mais fácil para que o aluno entenda e aceite as justificativas.

3) Outra vantagem da definição que adotamos é a possibilidade de usar frases como "A é independente de B" ou "B é independente de A", pois o sentido da independência pode ter apelo intuitivo na resolução de alguns problemas.

Para facilitar a escrita, é comum representar a interseção de vários eventos $A_1 \cap A_2 \cap \cdots \cap A_n$ da forma simplificada $A_1 A_2 \cdots A_n$. Utilizaremos esse artifício em vários exemplos deste livro, como no exemplo a seguir em que $FSSS$ representará a ocorrência simultânea de falha no primeiro componente e "não falha" nos três seguintes, ou seja, $F \cap S \cap S \cap S$.

Exemplo 5.3. *Um dispositivo eletrônico é formado por quatro componentes independentes. Cada um deles pode apresentar falha durante o funcionamento com probabilidade igual a* $0,1$. *Se dois ou mais componentes falharem, o dispositivo todo deixa de funcionar. Calcule a probabilidade disso acontecer.*

Considere o evento A no qual o dispositivo falha durante o funcionamento. Como esse evento envolve duas ou mais falhas de cada componente, vamos calcular a probabilidade do complementar A^C, que é equivalente a ocorrer uma ou nenhuma falha.

Representando os eventos "falha" de cada componente por F e "não falha" por S, estamos interessados na ocorrência de $SSSS$, $FSSS$ ou qualquer outra das 4 ordenações com um F e três Ss. Como os eventos são independentes, temos $P(FSSS) = P(F)P(S)^3 = P(SSFS)$, e isso vale para todas as 4 ordenações possíveis. Logo,

$$\begin{aligned} P\left(A^C\right) &= P(SSSS) + 4 \cdot P(FSSS) = P(S)^4 + 4 \cdot P(F)P(S)^3 \\ &= 0,9^4 + 4 \cdot 0,1 \cdot 0,9^3 = 0,9477 \\ &\Longrightarrow P(A) = 1 - 0,9477 = 0,0523 = 5,23\%. \end{aligned}$$

Comentários. A definição de independência para três ou mais conjuntos envolve a ideia de que a probabilidade da interseção é igual ao produto das probabilidades para todas as combinações dos conjuntos envolvidos tomados dois a dois, três a três etc. Não entendemos ser necessário esse nível de detalhes para o desenvolvimento no conceito no ensino básico.

Outra questão importante que veremos em exemplos a seguir é a necessidade da compreensão conceitual da ideia de independência, fazendo sua relação com a definição matemática e com problemas práticos. Consideramos importante que os alunos sejam expostos a diversas situações-problema nas quais devam identificar a independência ou não dos eventos envolvidos e fazer a relação com o significado matemático em cada caso.

Nos exemplos a seguir, veremos inúmeros casos de enganos cometidos (até mesmo por especialistas) com base no senso comum sobre questões contraintuitivas da aleatoriedade, que só a compreensão teórica dos conceitos de probabilidade condicional e independência podem evitar.

Exemplo 5.4 (Tubarão X Mosquito)**.** *Um famoso empresário com muitos seguidores no Twitter publicou em seu perfil a postagem da Figura 5.1 (*print screen*). Em tradução livre, a postagem diz: "Por que eu iria preferir encontrar um tubarão na natureza a um mosquito: Mosquitos matam mais pessoas em um dia que tubarões em um ano."*

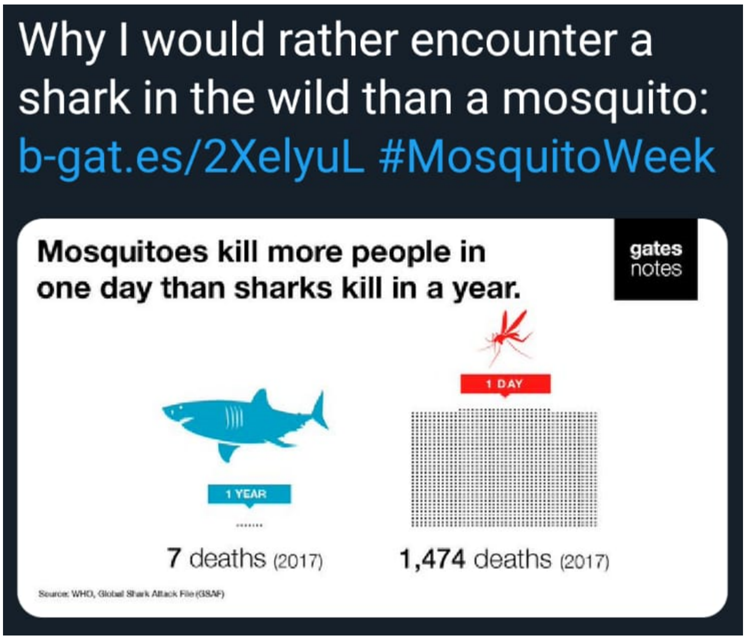

Figura 5.1: Problema do Tubarão X Mosquito

Antes de analisar a postagem, propomos as questões: de acordo com os dados apresentados, é justificável a comparação feita no *post*? Você passaria a ter mais medo de mosquito do que de tubarão após ver esses dados? Como avaliar o problema usando conceitos probabilísticos?

Quando o empresário afirma que prefere encontrar um tubarão a encontrar um mosquito fica implícito em sua fala que ele está avaliando haver um risco (probabilidade) maior de morte ao encontrar um mosquito. Mas observe que o encontro com o animal tem um papel fundamental na avaliação dessa probabilidade.

As palavras dele dão a entender que dado que uma pessoa encontra um mosquito, a probabilidade da mesma morrer é maior do que quando a pessoa encontra um tubarão. Mas a Figura 5.1 não nos dá essa informação. O que temos é que mosquitos matam mais que tubarões e provavelmente isso acontece porque os seres humanos têm muito mais contato com mosquitos do que com tubarões.

O engano na postagem ocorre por causa da diferença entre $P(A)$ e $P(A|B)$. A figura nos dá uma ideia da probabilidade de ser morto por um tubarão durante um ano e por um mosquito durante um dia ($P(A)$), mas não dá informação nenhuma sobre o risco de morte **dado que** uma pessoa encontrou um desses animais ($P(A|B)$). Para avaliar isso deveríamos ter os dados (praticamente inacessíveis) do número de encontros de pessoas com mosquitos e tubarões e o números de vezes que esses encontros provocaram mortes. Mas mesmo sem acesso a essa informação talvez tenhamos uma resposta intuitiva que deixamos para reflexão do leitor: qual dos encontros você acredita ser mais perigoso?

Muitos erros de interpretação de conceitos probabilísticos presentes em situações reais vêm da confusão que às vezes se faz entre as probabilidades $P(A)$, $P(A|B)$ e $P(B|A)$, que em geral são diferentes. Abaixo, são descritas algumas falácias advindas desse erro no raciocínio probabilístico que têm consequências diretas na vida das pessoas. Os próximos Exemplos 5.5 e 5.6 estão relatados em [18].

Exemplo 5.5 (Falso Positivo)**.** *Em estudos feitos na Alemanha e nos Estados Unidos, pesquisadores pediram a médicos que estimassem a probabilidade de uma mulher assintomática com idade entre 40 e 50 anos ter câncer de mama, dado que o exame foi positivo e sabendo que 7% das mamografias mostram câncer onde não existe. Além disso, eles sabiam que a incidência real da doença na população é de 0,8% e que a taxa de falsos negativos era de aproximadamente 10%.*

Entre os alemães, um terço dos médicos estimou a probabilidade pedida em cerca de 90% (a mediana foi 70%). E 95% dos médicos americanos concluíram que a probabilidade seria próxima de 75%. O valor correto é aproximadamente 9,4%.

Observe que a taxa de falsos negativos de 10% dada no problema significa que entre as mulheres que têm câncer e fazem o exame, 10% recebem resultado negativo. Assim, para o grupo de alemães citado acima, é possível que tenham usado a condicional invertida, uma vez que, essa taxa significa que

$$P(-|doente) = 10\% \Rightarrow P(+|doente) = 90\%$$

A probabilidade pedida foi, porém, $P(doente|+)$ que, como veremos na Seção 5.3, é aproximadamente igual a $9,4\%$, um valor muito distante da estimatva feita pela maioria dos médicos americanos e alemães.

Podemos fazer ainda outras especulações sobre como foram obtidas as probabilidades próximas a 70% ou 75%, mas dada a distância da probabilidade correta, um alerta importante fica desse exemplo: a falta de conhecimento de conceitos básicos de Probabilidade pode afetar inclusive profissionais de alta formação, com consequência direta na vida das pessoas.

Exemplo 5.6 (Falácia da Acusação)**.** *A britânica Sally Clarck foi condenada à prisão em 1999 por ter assassinado seus dois filhos recém-nascidos por sufocamento. O primeiro morreu com 11 semanas de vida, depois ela engravidou novamente e o segundo bebê morreu 8 semanas após nascer. A defesa alegava que a causa da morte fora síndrome da morte súbita infantil (SMSI), já que a autópsia não conseguira detectar nenhuma causa em nenhum dos dois casos. A acusação retrucou, afirmando que uma morte por SMSI tem probabilidade de ocorrência de* $1/8543$ *e, "portanto", duas mortes por essa causa teriam probabilidade de* $(1/8543)^2 = 1/72.982.849$*. Isso foi o suficiente para mandá-la para prisão. Nenhuma outra prova substancial foi apresentada. Você consegue detectar algum erro nessa argumentação da acusação já que foi um argumento puramente probabilístico?*

Primeiramente, multiplicar as probabilidades como foi feito presume que os eventos sejam independentes. Será que não há alguma disposição genética ou algum fator ambiental que tenha possibilitado essas mortes? O fato de o primeiro filho morrer por SMSI não interfere na chance do segundo seguir o mesmo caminho? Segundo relatado em [18], algumas semanas depois demonstrou-se que esses eventos de fato não são independentes. Ou seja, o argumento deveria ser negado como evidênica contra a ré.

Como se não bastasse o argumento acima, vamos analisar outro ponto. O enunciado apresenta um dado importante: as crianças morreram. Apesar de ser muito óbvia essa informação, ela não foi levada em consideração nos cálculos acima. A probabilidade anunciada de 1 para 8543 é a de que uma criança que está viva morra de SMSI. Mas na situação da britânica, as crianças já morreram, e desejamos avaliar a probabilidade de a morte ter sido por SMSI, dada essa informação, ou seja, a probabilidade condicional.

Na época, estimativas foram feitas por um matemático que, com base nas informações disponíveis, concluiu que a probabilidade de dois bebês terem morrido por SMSI era 9 vezes maior que a de terem sido assassinados. Para não deixar o leitor curioso sobre o desfecho do caso da Sally, ela foi solta da prisão após 3 anos e meio por conta de uma informação sobre a autópsia que foi omitida na época, mas que ao ser exposta anos depois ajudou a absolvê-la.

Não somos especialistas na área do Direito, mas usar probabilidade para direcionar investigações é bem diferente de condenar alguém **apenas** com base nesses números, mesmo que calculados de forma correta. Assim como um evento com alta probabilidade de acontecer pode não ocorrer, um evento com baixa probabilidade pode, sim, ocorrer.

Além disso, usar o produto de probabilidades para a ocorrência de dois eventos de forma simultânea, sem saber se são independentes, ou usar uma probabilidade no lugar de uma probabilidade condicional são erros comuns, mas nesse caso o erro custou a prisão de uma ré que anos depois foi absolvida.

Exemplo 5.7 (Viés de Sobrevivência)**.** *Nas situações descritas abaixo, estão algumas conclusões baseadas em dados que podem ser classificadas como falaciosas, pois estão contaminadas pelo chamado viés de sobrevivência. Esse termo é usado na Estatística quando, ao se analisar uma situação de seleção, são investigados apenas os casos que "sobreviveram" a esta seleção, e os demais não. Como justificar tal viés utilizando conceitos probabilísticos?*

a) (Finanças) *Em um livro que busca mostrar caminhos para a riqueza, o autor constata que pessoas que enriqueceram têm, comprovadamente, maior predisposição à tomada de risco. Então, o autor concluiu que a predisposição ao risco é recomendável para quem busca riqueza.*

Nessa mesma linha, estudos mostraram que muitos empreendedores de sucesso abandonaram a faculdade para investir nos seus projetos. Alguns concluíram com isso que a faculdade é perda de tempo para quem deseja empreender.

No primeiro caso não foram incluídos na análise aqueles que estão predispostos ao risco e nunca ficaram ricos (os "não sobreviventes"). Constata-se a alta probabilidade de a pessoa ser predisposta a riscos dado que ela enriqueceu, mas conclui-se que há uma alta probabilidade de a pessoa ficar rica dado que ela se predispõe ao risco (inversão de condicional).

No segundo caso, acontece algo semelhante. Os dados mostram probabilidade grande de a pessoa ter deixado a faculdade dado que ela teve sucesso, mas se concluiu que a probabilidade de uma pessoa ter sucesso dado que deixou a faculdade é boa.

As sentenças apresentadas nas conclusões podem até estar corretas, não estamos julgando isso. Mas não podem ser afirmadas baseadas apenas nos dados fornecidos acima.

b) (Guerras Mundiais) *Uma história ocorrida durante a I Guerra Mundial diz que capacetes de metal começaram a ser usados durante as batalhas europeias, mas logo a seguir o número de lesões na cabeça disparou.*

A princípio parece que algo muito controverso aconteceu, já que os capacetes deveriam reduzir esses números. Foi então que perceberam que os capacetes fizeram sim o efeito desejado, uma vez que esses novos ferimentos, anteriormente seriam mortes. A análise de eficácia deveria ter sido feita incluindo os não sobreviventes com ou sem o capacete de metal.

Outra situação muito interessante, mas dessa vez na II Guerra Mundial, é que pesquisadores estavam estudando os danos nos aviões que voltavam das batalhas. Analisaram quais eram as áreas mais afetadas e sugeriram que essas áreas fossem reforçadas, como mostra a ilustração fictícia da Figura 5.2. Foi quando o matemático Abraham Wald, da Universidade de Columbia, analisou e recomendou que fizessem o contrário. As áreas reforçadas deveriam ser exatamente aquelas não alvejadas nos aviões que retornaram.

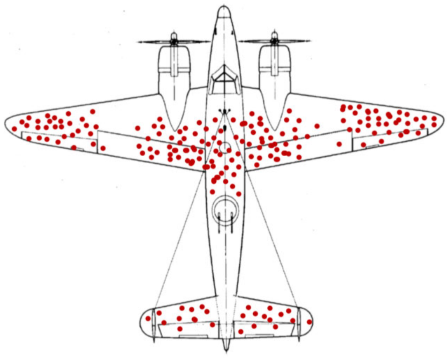

Figura 5.2: Viés de Sobrevivência - Fonte: [15]

O que Wald percebeu foi o viés de sobrevivência. Apenas as aeronaves que voltavam estavam sendo consideradas. As partes afetadas nas aeronaves sobreviventes não eram vitais, os aviões eram alvejados nelas e retornavam em segurança. Provavelmente os aviões alvejados nas outras partes (intactas no aviões que voltaram) estavam sendo abatidos.

Em termos matemáticos, foram avaliadas probabilidades de cada parte do avião ser alvejada, dado que o avião retornou da batalha. Deveria-se procurar, porém, a probabilidade de o avião retornar, dado que foi atingido em certas áreas, para então reforçar as mais sensíveis.

c) (Medicina) *Em um estudo sobre controle de pacientes com câncer no pâncreas, compararam pessoas com o câncer e pessoas sem o câncer para identificar fatores de risco. Mas este tipo de câncer é muito letal e avança rápido e a maior parcela dos doentes faleceram tão rápido que não entraram nos estudos, ficando apenas aqueles que tiveram uma sobrevida maior e com fatores de risco menos fatais do que os que morreram logo. Conclusão, fatores de risco importantes não foram detectados com o estudo.*

Comentários. Os exemplos acima nos dão uma ideia de quantos equívocos podem acontecer com assuntos relacionados a probabilidade condicional, dos mais ingênuos aos que trazem consequências mais graves. Seja desconsiderando a dependência entre eventos, seja invertendo a condicional ou mesmo desconsiderando-a. Essas concepções erradas podem gerar efeitos negativos graves para a sociedade. Por isso, é fundamental que o ensino de Probabilidade tenha esses conceitos como um de seus focos principais, destacando a diferença entre $P(A)$, $P(A|B)$ e $P(B|A)$ e as relações de independência, não apenas com exercícios focados em cálculos das probabilidades, mas também com exercícios que desenvolvam a leitura e a interpretação dos dados, com situações-problemas reais e de caráter qualitativo.

5.2 Teorema do Produto e Eventos Subsequentes

O Teorema do Produto é um dos mais importantes da Teoria das Probabilidades, com inúmeras aplicações em diversos tipos de problemas. Note que ele tem um aspecto intuitivo: para avaliar a probabilidade de ocorrência de vários eventos simultaneamente, você avalia a probabilidade do primeiro, depois do segundo dado que já avaliou o primeiro, depois do terceiro dados os dois primeiros, e assim sucessivamente.

Teorema 5.1 (Teorema do Produto)**.** *Dados os eventos $A_1, A_2, \cdots A_n$ de um espaço amostral, onde $P(A_1 \cap A_2 \cap \cdots \cap A_n) \neq 0$, temos*

$$\begin{aligned} P(A_1 \cap A_2 \cap \cdots \cap A_n) \;&=\; P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|(A_1 \cap A_2) \\ &\quad\cdots P(A_n|(A_1 \cap A_2 \cap \cdots \cap A_{n-1})) \end{aligned}$$

Demonstração. Observe inicialmente que, pela definição de probalidade condicional, temos $P(A \cap B) = P(A)P(B|A)$ para quaisquer $A, B \subset \Omega$, desde que $P(A) \neq 0$. Ou seja, se tomarmos $n = 2$, teremos:

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2|A_1)$$

Para $n = 3$:

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P((A_1 \cap A_2) \cap A_3) \\ &= P(A_1 \cap A_2) \cdot P(A_3|A_1 \cap A_2) \\ &= P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \end{aligned}$$

Estendendo o raciocínio para um n qualquer:

$$\begin{aligned} &P(A_1 \cap A_2 \cap \cdots \cap A_n) = P((A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \cap A_n) \\ &= P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \cdot P(A_n|A_1 \cap A_2 \cdots \cap A_{n-1}) \\ &\vdots \\ &= P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \end{aligned}$$

□

O Teorema do Produto permite-nos obter a probabilidade de ocorrência de vários eventos simultaneamente (ou, como veremos, em sequência) a partir das probabilidades de cada um condicionadas às ocorrências dos demais.

Exemplo 5.8 (Urnas e Sorteios em Sequência)**.** *Considere uma urna com três bolas brancas e sete bolas vermelhas. Duas bolas são retiradas uma após a outra e sem reposição. Determinar a probabilidade de que a primeira bola seja branca e a segunda, vermelha.*

Nesse tipo de problema, é comum materiais didáticos utilizarem interseção para representar eventos sequenciais. No nosso caso, sortear uma bola branca e uma bola vermelha, ***em sequência****, seria equivalente ao evento $B \cap V$. Qual é a explicação matemática para isso?*

O modo mais comum de resolver este problema é considerar a sequência de sorteios de bolas na urna como uma interseção de eventos e, então, aplicar o Teorema do Produto, como faremos a seguir.

Queremos que a primeira bola seja branca e a segunda vermelha. Assim, dados os eventos B: "sair bola branca" e V: "sair bola vermelha", temos

$$P(B \cap V) = P(B) \cdot P(V|B) = \frac{3}{10} \cdot \frac{7}{9} = \frac{21}{90} = \frac{7}{30}$$

A justificativa para $P(V|B)$ segue do fato de que ao sair a bola branca no primeiro sorteio, a urna fica com 9 bolas, das quais sete são vermelhas.

Mas por que podemos interpretar os eventos sequenciais "sair bola branca" e "sair bola vermelha" como se fosse a interseção dos dois eventos? Muitos livros apresentam a solução acima, que é intuitiva, sem uma justificativa; outros justificam a interseção com conectivo "e": sair bola branca **e** depois vermelha, seria equivalente a uma interseção. Mas seria a interseção de quais eventos? Vamos analisar essa situação.

Se considerarmos como espaço amostral as bolas da urna, teremos

$$\Omega = \{b_1, b_2, b_3, v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$$

sendo b_i e v_j, as bolas brancas e vermelhas, respectivamente, $i \in \{1, 2, 3\}$ e $j \in \{1, 2, \cdots, 7\}$.

Assim teremos $B = \{b_1, b_2, b_3\}$ e $V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$, ou seja, $B \cap V = \emptyset$. Então, $P(B \cap V)$ não deveria ser zero? A resolução acima está errada? Ou eventos em sequência podem ser interpretados como interseção de eventos?

Uma explicação para esse fato é a de que os resultados possíveis de dois eventos em sequência podem ser vistos como pares ordenados. Nesse caso, teremos um novo espaço amostral Ω' como o conjunto de todos os pares ordenados possíveis de serem obtidos nos dois sorteios. Mais precisamente

$$\Omega' = \{(x, y); x, y \in \Omega, x \neq y\}$$

já que o sorteio é sem reposição. Assim, $n(\Omega') = 10 \cdot 9 = 90$.

Consideremos agora o evento B': "sair bola branca na 1^a retirada" e o evento V': "sair bola vermelha na 2^a retirada", ou seja, $B' = \{(b_i, y); b_i, y \in \Omega, b_i \neq y\}$ e $V' = \{(x, v_j); x, v_j \in \Omega, x \neq v_j\}$. Com isso, obtemos

$$n(B') = 3 \cdot 9 \quad n(B' \cap V') = 3 \cdot 7$$

Desse modo, faz todo o sentido pensarmos em interseção de eventos, já que o evento $B' \cap V'$ seria o conjunto de todos os pares de Ω' do tipo (b_i, v_j), que representam ocorrência de bola branca no primeiro sorteio **e** bola vermelha no segundo.

Vamos calcular então $P(B')$ e $P(V'|B')$. Para esse segundo cálculo, usaremos a expressão obtida na Observação 2) da Definição 5.1:

$$P(V'|B') = \frac{n(B' \cap V')}{n(B')}$$

Assim,

$$P(B') = \frac{n(B')}{n(\Omega')} = \frac{3 \cdot \cancel{9}}{10 \cdot \cancel{9}} = \frac{3}{10}$$

$$P(V'|B') = \frac{\cancel{3} \cdot 7}{\cancel{3} \cdot 9} = \frac{7}{9}$$

$$\Longrightarrow P(B' \cap V') = P(B').P(V'|B') = \frac{3}{10} \cdot \frac{7}{9} = \frac{21}{90} = \frac{7}{30}$$

Observe que os fatores em comum eliminados no cálculo de $P(B')$ derivam do fato de que uma vez escolhida uma bola branca, a outra bola pode ser qualquer uma das não escolhidas, ou seja, a segunda escolha não afeta a probabilidade de B' ocorrer. Como consequência, $P(B') = P(B)$.

Com argumento similar, podemos analisar $P(V'|B')$ sem nos preocupar com pares ordenados, apenas considerando o fato de a urna ter uma bola branca a menos.

Assim, as probabilidades $P(B \cap V)$ e $P(V|B)$, dadas no início do problema, só têm sentido para o problema quando relacionadas à pares ordenados de ocorrências, ou seja, $P(B' \cap V')$ e $P(V'|B')$. Entretanto, para fins de cálculo, podemos pensar na ocorrência de eventos em sequência sem recorrer a pares ordenados, já que $P(B') = P(B)$ e $P(V|B)$ pode ser avaliada pensando em um único (e novo) sorteio, dado o resultado do primeiro, ainda que a justificativa de seu cálculo esteja atrelada à $P(V'|B')$.

Nesse mesmo sentido, eventos sequenciais como "ocorrer branca e, em seguida, vermelha" podem ser pensados como uma sequência de eventos únicos e representados por $B \cap V$, ainda que seu sentido matemático remeta a pares ordenados.

É claro que o raciocínio acima pode ser estendido para um número qualquer de bolas, de cores diferentes e de sorteios, e também para quaisquer eventos em sequência. Portanto, temos uma justificativa para resolver este tipo de problema pelo primeiro modo sem o receio de cometer erros. Ou seja, para efeito de cálculos, eventos subsequentes podem ser interpretados como interseção de eventos.

5.3 Diagrama de Árvore e o Teorema da Probabilidade Total

Teorema 5.2 (Teorema da Probabilidade Total)**.** *Se* $\Omega = A_1 \cup A_2 \cup \cdots \cup A_n$*, com* $A_1, A_2, \cdots A_n$ *disjuntos dois a dois e com probabilidade positiva, então para todo* $B \subset \Omega$*, temos*

$$\begin{aligned} P(B) &= P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_n) \\ &= P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + \cdots + P(A_n) \cdot P(B|A_n). \end{aligned}$$

Demonstração. Observe inicialmente que o conjunto B pode ser dado por

$$\begin{aligned} B &= B \cap \Omega = B \cap (A_1 \cup A_2 \cup \cdots \cup A_n) \\ &= (B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n) \end{aligned}$$

Em seguida, note que, pela hipótese sobre os conjuntos A_i, $i = 1, \cdots, n$, os eventos $(B \cap A_1), \cdots, (B \cap A_n)$ são disjuntos dois a dois. Logo, aplicando a propriedade (ii) do Teorema 4.1, obtemos:

$$\begin{aligned} P(B) &= P\left((B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n)\right) \\ &= P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_n) \\ &= P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + \cdots + P(A_n) \cdot P(B|A_n) \end{aligned}$$

onde a última igualdade foi obtida aplicando o Teorema do Produto 5.1 a cada parcela. □

Observações: 1) Ao dividir o espaço amostral Ω em uma união de subconjuntos disjuntos, dizemos que estamos fazendo uma partição de Ω. Na Figura 5.3, ilustramos uma partição de Ω em seis subconjuntos $A_1, \cdots, A_6$ disjuntos e um evento $B \subset \Omega$. O Teorema da Probabilidade Total tem um significado intuitivo: a probabilidade de qualquer evento pode ser decomposta na soma das probabilidades desses eventos avaliados sob a restrição de conjuntos menores. A vantagem nessa decomposição é que em várias situações nós temos como determinar essas probabilidades restritas a certas condições, para depois obter a probabilidade do evento pela soma.

2) Como será possível observar em vários problemas que veremos à frente, esse teorema traduz-se em uma ferramenta poderosa de cálculo de probabilidades. Em particular, o teorema dá a base teórica para utilização de uma técnica bastante importante na resolução de problemas de Probabilidade conhecida como diagrama de árvore.

3) Para uma primeira apresentação dos Teoremas do Produto, da Probabilidade Total e de Bayes (que será enunciado na próxima seção) aos alunos de ensino básico,

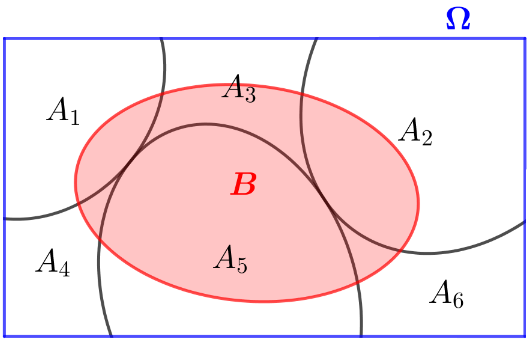

Figura 5.3: Partição de Ω

sugerimos considerar $n = 2$, para que as sentenças não pareçam muito complicadas. Além disso, a abordagem intuitiva através de um ou mais exemplos antes de apresentar os teoremas é sempre recomendável. Após uma compreensão efetiva pelos alunos, podemos estender para n qualquer, se avaliarmos necessário. Ressaltamos ainda que, havendo a compreensão intuitiva desses teoremas, os alunos podem aplicá-los com o auxílio do diagrama de árvore e sem a necessidade de memorizar fórmulas. A propósito, a memorização das fórmulas nunca deve ser um objetivo, e sim a compreensão do conceito envolvido.

Exemplo 5.9 (Problema de Monty Hall)**.** *Esse problema está relacionado com uma atração que fazia parte de um programa da TV americana apresentado por Monty Hall. Neste jogo, Monty Hall apresentava três portas fechadas aos participantes. Atrás de uma das portas havia um prêmio, e atrás de cada uma das outras havia um bode. Na primeira etapa o particpante escolhia uma porta que permaneceria fechada. Na segunda etapa, Hall abria uma das outras duas portas. Essa porta que o apresentador escolhia para abrir sempre tinha um bode, já que ele sabia onde estava o prêmio. Na terceira etapa, Hall pergunta ao participante se deseja manter a sua escolha inicial ou se pretende trocar e ficar com a outra porta que ainda está fechada. Qual a melhor estratégia para se ganhar o prêmio? Manter a escolha inicial ou trocar de porta? Ou essa escolha é irrelavante para o problema?*

Esse problema ficou mais famoso entre os matemáticos na década de 1990, quando a pergunta acima foi encaminhada a uma famosa colunista chamada Marilyn vos Savant, que costumava responder os leitores em um jornal. Ela respondeu que o participante teria maior probabilidade de ganhar se trocasse de porta. Sua resposta foi considerada contraintuitiva por muitos, já que havendo duas portas em jogo, a probabilidade de cada uma conter o prêmio deveria ser $1/2$. A conclusão da colunista virou objeto de polêmica na época e foi questionada até por professo-

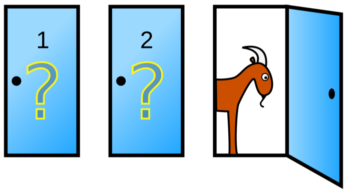

Figura 5.4: Problema de Monty Hall

res doutores em Matemática, que apelaram a ela através de correspondência para que corrigisse o "erro" perante o público. E essa visão obteve apoio da população; segundo uma pesquisa feita na época ([18]), 92% dos americanos achavam que ela estava errada.

Como justificar a conclusão da colunista utilizando o Teorema da Probabilidade Total?

Um modo de explicar essa conclusão é fazendo um diagrama de árvore. Esse diagrama é construído com base na relação obtida no Teorema 5.2 e faz com que o problema fique mais intuitivo. Para relacionar o diagrama da Figura 5.5 com o teorema, note que podemos particionar o espaço amostral do problema em dois eventos: A_1: "o participante escolheu a porta com o prêmio" e A_2: "o participante escolheu a porta com o bode". Dessa forma, temos $P(A_1) = 1/3$ e $P(A_2) = 2/3$ e podemos calcular a probabilidade dos eventos B: "vencer trocando de porta" e C: "vencer sem trocar a porta", aplicando o teorema da probabilidade total. Assim, temos

$$
\begin{aligned}
P(B) &= P(B \cap A_1) + P(B \cap A_2) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) \\
P(C) &= P(C \cap A_1) + P(C \cap A_2) = P(A_1)P(C|A_1) + P(A_2)P(C|A_2) \quad (5.1)
\end{aligned}
$$

Entretanto, vamos apresentar a seguir a resolução por um caminho intuitivo, diretamente pelo diagrama e sem usar a fórmula do teorema. No primeiro estágio, o participante escolhe uma porta e só há duas opções: a porta escolhida tem o prêmio ou tem um bode. A probabilidade de o participante escolher a porta com prêmio é $1/3$ e a de escolher uma porta com bode é $2/3$. No segundo estágio, temos apenas duas portas e já sabemos que em uma está o prêmio e na outra está um bode. Se o participante escolher a porta com prêmio inicialmente, ele ganha se não trocar, ou seja, a probabilidade de ganhar trocando de porta é 0 e a probabilidade de ganhar não trocando é 1; mas se a porta escolhida tinha bode, ele ganha se trocar, invertendo as probabilidades do primeiro caso. Veja a árvore ilustrada na Figura 5.5.

Assim, podemos dizer que a probabilidade de o participante ganhar o prêmio quando NÃO TROCA DE PORTA é dada pelos dois caminhos que levam ao evento

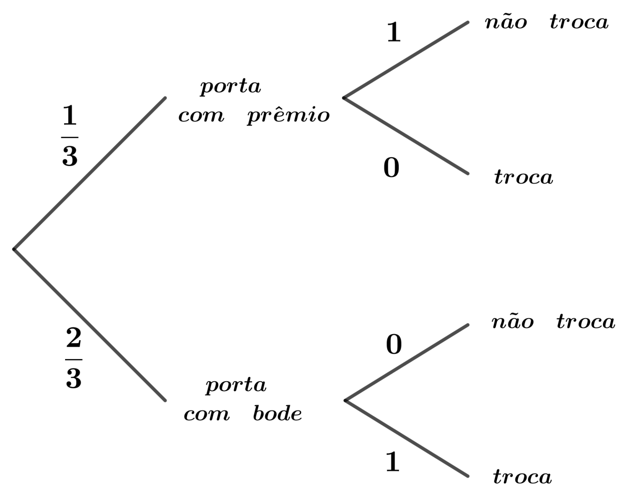

Figura 5.5: Problema de Monty Hall

"vencer sem trocar a porta" (na Figura, representado como "não troca"), ou seja:

$$P(C) = \frac{1}{3} \cdot 1 + \frac{2}{3} \cdot 0 = \frac{1}{3}$$

Já, a probabilidade de ganhar quando TROCA DE PORTA é:

$$P(B) = \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1 = \frac{2}{3}$$

Note que os dois cálculos acima são justificados por 5.1.

Portanto, trocar de porta dobra a probabilidade de ganhar se comparado a manter a escolha inicial, como afirmou Marilyn.

Comentários. Existem outras formas conhecidas de resolver esse problema, algumas delas bem criativas e que ajudam a eliminar o fator contraintuitivo. Por exemplo, imaginar um jogo com 100 portas, 99 bodes e que o apresentador abre 98 portas com bode na segunda etapa. Faria sentido dizer que a probabilidade de vencer trocando de porta é $1/2$ nesse caso só porque sobraram duas portas? Há muitas outras soluções disponíveis na internet. Pesquisar mais de uma solução em sala de aula, discuti-las com os alunos e compará-las pode ser uma atividade didática bastante rica.

Esta resolução está baseada no Teorema da Probabilidade Total (Teorema 5.2). Além de ser um exemplo interessante para trabalhar as habilidades vistas neste capítulo, é também um problema divertido, os alunos do ensino básico costumam gostar e se interessar por ele. Existem vários aplicativos para "brincar" com esse problema, simulando as escolhas e portas distintas para o prêmio; alguns permitem até mesmo alterar a quantidade de portas, o que traz uma nova variante para o problema. Pode ser uma boa oportunidade para usar os *smartphones* a favor das aulas.

Exemplo 5.10 (Solução Combinatória x Soluçao Probabilística)**.** *Duas peças de um jogo de dominó comum são sorteadas aleatoriamente, sem reposição. Calcule a probabilidade de que as duas peças tenham um número em comum.*

Aqui sugerimos ao leitor que, antes de ler nossa resolução, tente resolver o problema por dois modos, o primeiro usando Análise Combinatória e o segundo usando diagrama de árvore.

Esse exemplo apresenta uma ótima oportunidade para desenvolver e comparar uma solução com raciocínio inteiramente combinatório com uma com raciocínio probabilístico. Qual delas terá a preferência do leitor? E dos alunos?

Resolvendo usando métodos de Combinatória, temos:

- número de casos possíveis: $C_{28,2} = 378$.
- número de casos favoráveis escolhendo uma peça dupla - escolher a peça dupla e, em seguida, uma das 6 peças que repetem seu número: $7 \cdot 6 = 42$.
- número de casos favoráveis sem peça dupla - escolher o número que será repetido e, em seguida, um dos possíveis pares que repetem esse número: $7 \cdot C_{6,2} = 105$.

Resposta: $P(E) = \dfrac{147}{378} = \dfrac{7}{18}$.

Ou, então, considerando escolhas ordenadas:

- número de casos possíveis: $28 \cdot 27 = 756$.
- número de casos favoráveis escolhendo uma peça dupla - escolher uma peça dupla, uma das 6 peças que repete seu número e contar as 2 odenações possíveis para cada par: $7 \cdot 6 \cdot 2 = 84$.
- número de casos favoráveis sem peça dupla - escolher uma das 21 peças simples e, em seguida, escolher uma das 5 peças simples restantes que repete um dos números ou uma das 5 que repete o outro número da primeira peça: $21 \cdot 10 = 210$ (note que a contagem considera as ordenações de cada dupla de peças escolhidas).

Resposta: $P(E) = \dfrac{294}{756} = \dfrac{7}{18}$.

Agora faremos uma solução usando o artifício do diagrama de árvore (Figura 5.6).

No primeiro estágio, para a escolha da primeira peça, podemos tirar uma peça dupla (com números iguais) ou uma peça simples (com dois números diferentes).

A probabilidade de retirar uma peça dupla é $7/28$, pois são 7 peças duplas em um total de 28 peças em um jogo de dominó. Portanto, a probabilidade de retirar uma peça simples é $21/28$.

No segundo estágio (retirada da segunda peça), para o ramo de cima onde a primeira peça é dupla, podemos ter uma peça com número em comum com a primeira com probabilidade de $6/27$ ou peça sem número em comum com probabilidade $21/27$. Analogamente, para o ramo de baixo temos probabilidade $12/27$ para a segunda peça com número em comum, e $15/27$, caso contrário. Observe que as probabilidades referentes a não ter número em comum não nos interessam no problema, e poderíamos construir a árvore sem esses valores.

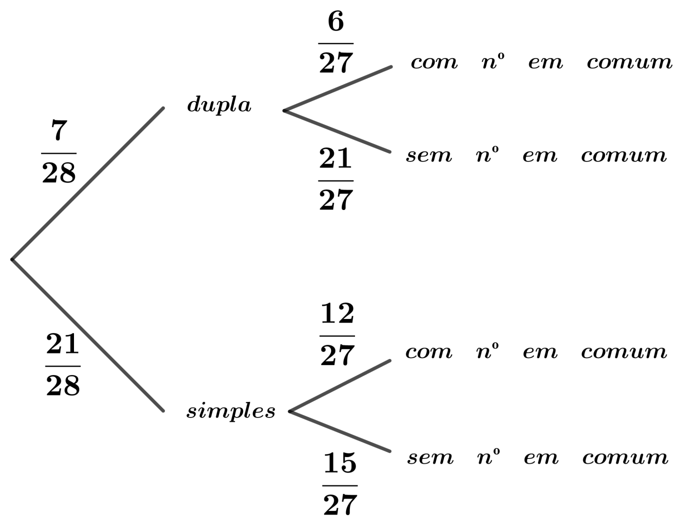

Figura 5.6: Problema do Dominó

Utilizando apenas os "ramos" que nos interessam nesse problema (pontas com número em comum) teremos:

$$P(E) = \frac{7}{28} \cdot \frac{6}{27} + \frac{21}{28} \cdot \frac{12}{27} = \frac{7}{18}$$

Comentários. É interessante observar e destacar com os alunos que, quando abrimos os ramos, a soma das probabilidades sempre deve ser igual a 1 (ou 100%), pois o Teorema da Probabilidade Total particiona Ω, que tem probabilidade 1. Na árvore acima temos $7/28 + 21/28 = 1$, $6/27 + 21/27 = 1$ e $12/27 + 15/27 = 1$. Caso isso não aconteça, pode ter ocorrido pelo menos um de dois problemas: não foram consideradas todas as possibilidades ou houve erro no cálculo das probabilidades.

Resolvemos o problema de três formas na intenção de fazer o leitor perceber que em determinadas situações usar Combinatória pode ser mais complicado e até mesmo confuso para o aluno, além de ser desnecessário nesse caso. Não são raras as situações em que problemas de contagem causam confusões até mesmo nos professores de Matemática. Entre os alunos, é comum a dúvida "a ordem de retirada das peças importa ou não?" ou "as peças são retiradas uma após a outra ou simultaneamente?". Questionamentos irrelevantes ao resolvermos pelo diagrama de árvore.

No caso desse problema, a solução combinatória pode trazer várias armadilhas. Observe que apresentamos duas soluções com técnicas de contagem: uma considerando ordenações, outra não considerando. As duas produzem a mesma probabilidade, mas é comum que alunos cometam o erro de contar o número de casos favoráveis com ordenação e o de casos possíveis sem, ou vice-versa. A propósito, no caso desse problema, erros em ordenações são até compreensíveis. Note que no caso ordenado multiplicamos por 2 um dos itens, mas não o fizemos nos outros, por causa da natureza dos pares em cada caso. Um sutileza difícil de ser percebida. Além disso, esse é um típico problema de contagem em que qualquer descuido resulta em contagem em excesso ou em falta. São várias armadilhas com potencial de erro na solução pelo caminho combinatório.

Acreditamos que a melhor opção didática para esses casos seja o desenvolvimento e comparação das duas soluções em sala de aula. Ao final, pode-se fazer a generalização de que tipos de problemas permitem ao aluno essa escolha, e deixar que cada um adote a estratégia que lhe for conveniente. Ainda que no problema acima tenhamos preferido a solução probabilística, podemos mudar diante de um problema novo e optar pela solução envolvendo contagem. O fundamental é que os alunos conheçam e possa usar as estratégias de resolução em cada caso.

Exemplo 5.11 (Jogo do Milhão)**.** *Em um programa de TV, um jogador deve responder perguntas de múltipla escolha com quatro alternativas para chegar à pergunta final do milhão, pergunta que vale 1 milhão de reais se for respondida corretamente. Ao aparecer uma pergunta que o jogador não tem a mínima ideia de qual é a resposta, é oferecida a oportunidade para que ele faça apenas uma das duas perguntas a seguir, com respostas corretas do apresentador do tipo "sim" ou "não".*
1) "A resposta é C?"; *2) "A resposta é B ou C?"*
Qual das estratégias é mais favorável: fazer a pergunta 1 ou a pergunta 2?

Observe como usar o Teorema da Probabilidade Total, através do diagrama de árvore (Figuras 5.7 e 5.8), permite a solução rápida desse dilema e facilita bastante a visualização da resolução.

Se o jogador escolher a pergunta 1) a probabilidade de a resposta ser "sim" é de $1/4$, e de ser "não" $3/4$, pois estamos supondo as quatro alternativas com igual probabilidade de ser a correta. Se a resposta for "sim" o jogador ganha, pois ele já saberá que C é a alternativa correta, logo, a probabilidade de ganhar é 1. Mas se a resposta for "não" o jogador fica com três possibilidades de alternativas, logo, ele terá probabilidade de ganhar igual a $1/3$. Então, a probabilidade de ganhar escolhendo a pergunta 1 é:

$$\frac{1}{4}\cdot 1+\frac{3}{4}\cdot\frac{1}{3}=\frac{1}{2}$$

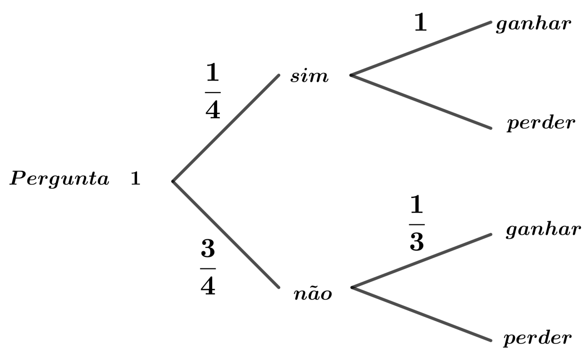

Figura 5.7: Problema do Jogo do Milhão

Mas se o jogador escolher a pergunta 2, a probabilidade de a resposta ser "sim" é $1/2$, e de ser "não" também é $1/2$. A resposta sendo "sim" ou sendo "não" o jogador fica com duas possibilidades de alternativas tendo probabilidade $1/2$ de ganhar. Logo, a probabilidade de ganhar escolhendo a pergunta 2 é:

$$\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{2}$$

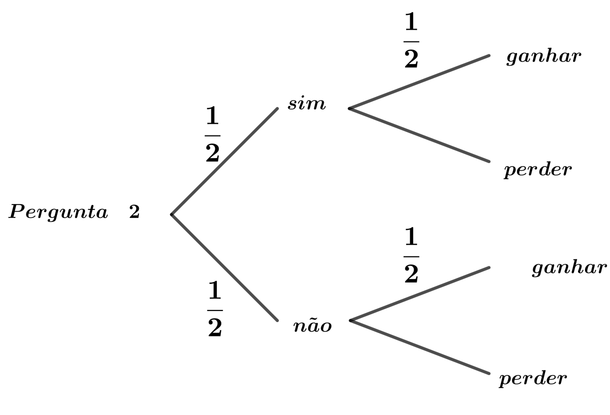

Figura 5.8: Problema do Jogo do Milhão

Portanto, em termos probabilísticos, qualquer que seja a pergunta escolhida, o jogador terá 50% de chance de escolher a alternativa correta como resposta.

Comentários. Outras variantes interessantes desse problema podem ser propostas em sala de aula pelos alunos e pelo professor, alteranto as possibilidades de perguntas. Pode-se ainda utilizar as regras de alguns programas reais desse tipo para analisar as melhores estratégias no caso de o jogador não saber a resposta: perguntar ao público? Aos universitários? Eliminar alternativas falsas?

5.4 Teorema de Bayes

O Teorema de Bayes é utilizado no contexto em que temos uma partição Ω e que, dada a ocorrência de um evento B, queremos saber a probabilidade de ter ocorrido cada evento da partição. Como veremos, esse raciocínio é muito útil para calcular a probabilidade de eventos passados.

Do ponto de vista prático e didático, a fórmula obtida pelo Teorema de Bayes não é necessária para resolver problemas nem para trabalhar com o conceito em sala de aula. Como veremos na demonstração abaixo, o cálculo da probabilidade condicional do teorema pode ser feito usando apenas a definição de condicional, o Teorema da Probabilidade Total (diagrama de árvore) e o Teorema do Produto. O Teorema de Bayes é apenas uma forma de generalizar a aplicação desses três conceitos e não é necessário, do ponto de vista do ensino básico, que os alunos memorizem essa fórmula.

Teorema 5.3 (Teorema de Bayes)**.** *Se $\Omega = A_1 \cup A_2 \cup \cdots \cup A_n$, com $A_1, A_2, \cdots A_n$ disjuntos dois a dois, tais que $P(A_i) \neq 0$ para todo $i = 1, \cdots, n$ e $P(B) \neq 0$, então temos:*

$$P(A_i|B) = \frac{P(A_i) \cdot P(B|A_i)}{P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + \cdots + P(A_n) \cdot P(B|A_n)}$$

Demonstração. A prova desse teorema deriva diretamente da definição de probabilidade condicional, do Teorema do Produto e do Teorema da Probabilidade Total, esse último sendo usado no denominador da expressão abaixo.

$$\begin{aligned} P(A_i|B) &= \frac{P(A_i \cap B)}{P(B)} \\ &= \frac{P(A_i) \cdot P(B|A_i)}{P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + \cdots + P(A_n) \cdot P(B|A_n)} \end{aligned}$$

□

Aplicaremos a ideia do Teorema de Bayes para resolver o Problema 5.5 que apresentamos na Seção 5.1. Vamos relembrar seu enunciado.

Exemplo 5.12 (Falso Positivo)**.** *Em estudos feitos na Alemanha e nos Estados Unidos, pesquisadores pediram a médicos que estimassem a probabilidade de uma mulher assintomática com idade entre 40 e 50 anos ter câncer de mama, dado que* *o exame foi positivo e sabendo que 7% das mamografias mostram câncer que não existe. Além disso, eles sabiam que a incidência real da doença na população é de 0,8% e que a taxa de falsos negativos era de aproximadamente 10%. Calcule a probabilidade que os médicos deveriam encontrar.*

A princípio, vamos elaborar o diagrama de árvore com as informações que temos e, a seguir, o completamos usando o fato de que a soma das probabilidades dos ramos deve ser 100%, como mencionamos nos comentários do Exemplo 5.10. Daí, teremos a árvore completa na Figura 5.9. Note que o segundo ramo foi obtido a partir das taxas de falso negativo e falso positivo. A taxa de falso negativo é a probabilidade de o exame dar negativo para a doença, dado que uma pessoa está com a doença, e a da falso positivo é a probabilidade de dar positivo, dado que a pessoa está sadia.

Uma dúvida comum que surge ao utilizar esse diagrama é sobre a ordem que devemos utilizar. Por exemplo, no diagrama abaixo poderíamos ter tentado colocar os primeiros ramos como sendo "positivo" ou "negativo" e deixar os ramos "doente" e "não doente" para depois. De fato, não há uma regra para saber quais ramos vêm primeiro. Mas observe que se tentássemos montar o diagrama ao contrário, faltariam informações, pois não temos, entre outras, a informação no enunciado sobre a probabilidade de estar doente, dado que o exame foi positivo; essa é justamente a probabilidade que estamos buscando. Assim, caso tentemos montar o diagrama na ordem errada, os dados do problema nos "avisarão". Do ponto de vista didático, é interessante que se faça a especulação entre as duas (ou mais) ordenações possíveis sempre que um problema desse tipo aparecer.

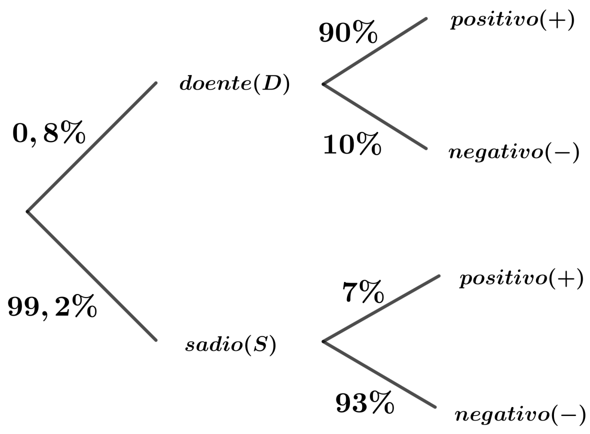

Figura 5.9: Problema do Falso Positivo

Observe que a informação que queremos é a probabilidade de que uma pessoa esteja doente (D), dado que o exame foi positivo $(+)$, $P(D|+)$. Mas o que temos até aqui com o enunciado é a probabilidade com a condicional invertida, ou seja, a probabilidade de que um exame seja positivo, dado que a pessoa está doente, $P(+|D) = 90\%$. Então, agora é o momento de usar o Teorema de Bayes (Teorema 5.3), mas, como dissemos, sem necessariamente apelar para sua fórmula; basta aplicar a definição de probabilidade condicional e buscar o resto das informações no diagrama de árvore:

$$P(D|+) = \frac{P(D \cap +)}{P(+)} = \frac{0,90 \cdot 0,008}{0,90 \cdot 0,008 + 0,07 \cdot 0,992} \simeq 0,094 = 9,4\%$$

Comentários. Observe que a probabilidade de um paciente ter exame positivo $P(+)$ foi calculada usando o Teorema da Probabilidade Total (Teorema 5.2) e que no numerador aplicamos o Teorema do Produto. Utilizando o diagrama de árvore, a definição de condicional e a ideia do produto, a fórmula do Teorema de Bayes não é necessária para a resolução do problema. Por isso, acreditamos que ele pode ser apresentado como uma generalização **após** os alunos terem desenvolvido a habilidade de resolver esse tipo de problema.

Observamos ainda que, dado o resultado do exame, o que calculamos foi a probabilidade de um evento já ocorrido (a pessoa ter adquirido a doença), ou seja, um evento do passado. Essa é a ideia principal que deve ser desenvolvida na compreensão de problemas como esse, além da técnica de resolução. Promover uma discussão sobre os significados teórico e prático de uma “probabilidade do passado” pode ser uma atividade didática bastante rica.

Tal exemplo tem a vantagem ainda de mostrar, por meio de uma situação real, o quão diferentes podem ser os valores de probabilidades com condicionais invertidas, e reforçar sobre a importância de um cidadão qualquer e profissionais das mais diversas áreas compreenderem conceitos básicos de Probabilidade, já que o problema envolve uma situação médica pela qual qualquer um pode passar.

5.5 Problemas Propostos

Problema 5.1 (Problema da Moeda de Bertrand)**.** *Há três caixas idênticas. A primeira contém duas moedas de ouro, a segunda duas moedas de prata e a terceira contém uma moeda de ouro e outra de prata. Escolhe-se uma caixa ao acaso, e sem olhar seu conteúdo, retira-se uma moeda. Sabendo que a moeda retirada é de ouro, qual a probabilidade de se retirar a segunda moeda e esta ser de ouro?*

Problema 5.2. *Quando um exame antidoping tem taxa de falso positivo de 1% e um atleta atesta positivo no exame, muitas pessoas concluem que a probabilidade de o atleta ser culpado é de 99%. Esse raciocínio está correto? Por quê?*

Problema 5.3. *Durante o julgamento, a acusação afirma que pesa sobre o réu o fato de que o criminoso foi visto saindo da cena do crime dirigindo um carro raro* *igual ao que o réu possui, cuja frequência na cidade onde ocorreu o crime é de* $1/10.000$*. Então, a acusação conclui que a probabilidade de ele ser inocente é de* $1/10.000$*. Esse raciocínio está correto? Por quê?*

Problema 5.4. *Intuitivamente, sabemos que sorteios em sequência com reposição são eventos independentes, e sorteios sem reposição são dependentes. Usando o exemplo da urna com três bolas brancas e sete vermelhas, mostre a partir da definição matemática de independência que a intuição acima se confirma.*

Problema 5.5. *Em uma caixa estão dez bolas que diferem apenas na cor: são seis brancas e quatro pretas. Ao retirar dessa caixa três bolas ao acaso, sucessivamente, qual é a probabilidade de sair, em qualquer ordem, duas bolas brancas e uma preta, se o sorteio for feito:*
a) com reposição? *b) sem reposição?*

Para o problema acima, é possível obter soluções por caminhos combinatórios ou aplicando unicamente ferramentas probabilísticas. Obtenha pelo menos uma de cada para compará-las.

Problema 5.6. *Três cartas são retiradas de um baralho comum. Se for dada a informação de que as três são vermelhas, qual é a probabilidade de que pelo menos uma seja de copas? Com a mesma informação, qual é a probabilidade de sair três figuras ou três cartas de copas?*

Problema 5.7. *a) Prove que se* A *e* B *são eventos independentes, então também são independentes os pares:* A^C *e* B*,* A *e* B^C*,* A^C *e* B^C*.*
b) Mostre que se $P(B) = 1$*, então* $P(A|B) = P(A)$*.*

Problema 5.8. *Duas máquinas A e B produzem 5000 peças em um dia, sendo um quinto produzido pela máquina A. Das peças produzidas pela máquina A, 4% são defeituosas e, pela máquina B, 1,5%. Uma peça da produção total de um dia é escolhida ao acaso, e constata-se que ela é defeituosa. Qual é a probabilidade de que ela tenha sido produzida pela máquina A? E qual a probabilidade de uma peça qualquer, escolhida aleatoriamente na produção diária dessas duas máquinas, ser defeituosa?*

Problema 5.9. *Os testes de DNA feitos na investigação de um crime apontam que um suspeito tem 80% de probabilidade de ser o culpado. Além disso, descobriu-se que a pessoa que cometeu o crime tem mais que 1,80m de altura, característica comum a 7% da população local. Sabendo que o suspeito citado acima possui essa característica, qual passa a ser a probabilidade de ele ser o culpado?*

Capítulo 6

Um Outro Olhar para a Binomial e para as PGs

"Probabilidades não fazem parte das moedas; probabilidades fazem parte das pessoas."

– Persi Diaconis

Uma variável aleatória discreta, de forma simplificada, é uma variável X de caráter aleatório cujos valores possíveis pertencem a um conjunto finito ou infinito enumerável. Por exemplo, se lançarmos uma moeda 10 vezes e denotarmos por X o número de caras obtidas, então X é uma variável aleatória discreta que pode assumir valores no conjunto $\{0, 1, 2, ..., 9, 10\}$.

Em geral, sequências de experimentos independentes repetidos e com dois resultados possíveis, como é o caso dos lançamentos de uma moeda descrito acima, têm suas probabilidades seguindo um certo padrão, o chamado modelo binomial que veremos mais à frante. Nesse sentido, os modelos probabilísticos discretos são generalizações de cálculos probabilísticos que aparecem com frequência em problemas relacionados a variáveis aleatórias discretas.

Em geral, no ensino médio, os alunos têm contato com um único modelo discreto: a distribuição binomial de probabilidades. Esse modelo está entre os mais importantes em termos de aplicações em Probabilidade e Estatística.

Entendemos, porém, que vários de seus aspectos e importantes aplicações no mundo real não são adequadamente explorados nesse segmento. Assim, as duas primeiras seções deste capítulo fazem um apanhado geral do modelo binomial e apresentam exemplos e propostas didáticas diferentes da abordagem tradicional nos livros didáticos, com sugestão do uso de *softwares* e aplicativos para a resolução de problemas, a relação com a média e o desvio padrão e a aproximação com a distribuição normal de probabilidades.

Por outro lado, outros modelos discretos de probabilidade têm papel importante na modelagem de fenônemos aleatórios, e sua utilização é acessível ao público do ensino médio. Em particular, destacamos a distribuição geométrica de probabilidades, que além de ser de compreensão intuitiva e modelar uma série de problemas relevantes, pode ser vista como uma possibilidade de aplicação e de contextualização de um conceito conhecido dos alunos do ensino médio: as progressões geométricas.

Assim, a distribuição geométrica de probabilidades será o tema da terceira seção deste capítulo, onde abordaremos exemplos de aplicação, propriedades e sua relação com as progressões.

6.1 Distribuição Binomial

Antes de apresentar a definição e as principais propriedades da distribuição binomial, vamos explorar o conceito através de um exemplo.

Exemplo 6.1 (Exemplo Introdutório)**.** *Considere uma prova com dez questões de múltipla escolha com cinco alternativas em cada e apenas uma delas correta. Suponha que um estudante vai responder todas as dez questões sorteando de forma independente uma das alternativas para marcar a resposta. Qual é a probabilidade de que o estudante acerte pelo menos três questões?*

Denotamos por X a variável aleatória que conta o número de questões em que o estudante marca a resposta correta. Note que X varia no conjunto $\{0, 1, 2, ...9, 10\}$. Queremos calcular $P(X \geq 3)$, mas o cálculo direto seria muito extenso, pois deveríamos calcular $P(X = 3)$, $P(X = 4)$,..., $P(X = 10)$ e somar todos esses valores. Então, vamos obter a probabilidade do complementar:

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$$

Denotaremos por A o evento em que o aluno acerta uma determinada questão e por E quando ele erra. O evento $\{X = 0\}$ corresponte à sequência de eventos $EEEEEEEEEE$. Como os sorteios são independentes e a probabilidade de cada resposta errada é $4/5$, temos

$$P(X = 0) = P(EEEEEEEEEE) = P(E)^{10} = \left(\frac{4}{5}\right)^{10} \simeq 0,1074$$

Já o evento $\{X = 1\}$ corresponde a todas as sequências com um A e nove Es. Podemos ter várias ordenações diferentes, como $EEAEEEEEEE$. Mudando a posição do A, temos 10 ordenações distintas, e a probabilidade de cada ordenação é dada por um produto cuja posição de A não altera o resultado. Ou seja, todas as ordenações possíveis têm a mesma probabilidade. Portanto:

$$P(X = 1) = 10 \cdot P(AEEEEEEEEE) = 10 \cdot \frac{1}{5}\left(\frac{4}{5}\right)^{9} \simeq 0,2684$$

Por fim, o evento $\{X = 2\}$ corresponde a todas as ordenações com oito Es e dois As, como $EEEAEAEEEE$. O total de ordenações para essas sequências é obtido utilizando a técnica de permutações com repetição, o que para o caso de duas letras coincide com o valor do coeficiente binomial:

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

Portanto:

$$P(X = 2) = 45 \cdot P(AAEEEEEEEE) = 45 \cdot \left(\frac{1}{5}\right)^2 \cdot \left(\frac{4}{5}\right)^8 \simeq 0,302$$

Logo,

$$P(X \geq 3) = 1 - P(X < 3) \simeq 1 - (0,1074 + 0,2684 + 0,302) = 0,3222$$

No exemplo acima, temos uma situação que aparece com frequência em diversos problemas probabilísticos. Temos uma sequência de experimentos aleatórios (sorteios das respostas) independentes entre si, com apenas dois resultados possíveis (A ou E), cuja probabilidade mantém-se a mesma para cada repetição do experimento.

Nessas condições, estamos interessados em calcular probabilidades relacionadas à variável X que conta o número de ocorrências de um dos dois resultados possíveis (A ou E), entre o total de repetições realizadas. O evento que estamos interessados em contar o número de ocorrências é chamado de *sucesso* (no nosso caso, ACERTO) e o outro de *fracasso*.

A distribuição de probabilidades de uma variável aleatória discreta X é uma função $P(X = k)$ que fornece a probabilidade de a variável aleatórioa X assumir o valor k. Para o caso acima, a distribuição de qualquer variável aleatória X nas condições dadas segue um padrão que explicitaremos à frente. Essa distribuição é chamada de distribuição binomial de probabilidades.

Definição 6.1. *Considere uma sequência de n experimentos aleatórios independentes com dois resultados possíveis: sucesso, com probabilidade p para cada experimento, e fracasso, com probabilidade $1 - p$ também para cada experimento. Se X é a variável aleatória que descreve o número de sucessos nessa sequência, então, a distribuição de probabilidades de X é chamada de Distribuição Binomial com parâmetros n e p.*

Teorema 6.1 (Teorema Binomial)**.** *Se X tem distribuição binomial com parâmetros n e p, então,*

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad \forall k; 0 \leq k \leq n$$

Demonstração. O evento $\{X = k\}$ corresponde à ocorrência de k sucessos e $n-k$ fracassos em qualquer ordem. Representando sucesso por S e fracasso por F, uma das sequências possíveis para o evento em questão é

$$\underbrace{SSS\cdots S}_{k\ sucessos}\ \underbrace{FFF\cdots F}_{n-k\ fracassos}$$

Dada a independência dos eventos, a probabilidade para a sequência de sucessos e fracassos acima é

$$P(SSS\cdots SFFF\cdots FFF) = P(S)^k \cdot P(F)^{n-k} = p^k(1-p)^{n-k}$$

Por outro lado, note que qualquer outra ordenação de k sucessos e $n-k$ fracassos tem a mesma probabilidade acima, pois só seria alterada a ordem dos fatores. Ou seja, devemos verificar o número de ordenações possíveis para essa sequência e multiplicar pela probabilidade de cada ordenação. Esse número de ordenações é dado pelas permutações de n elementos com k repetições de S e $n-k$ repetições de F. Portanto é dado por:

$$P_n^{k,n-k} = \frac{n!}{(n-k)!k!} = \binom{n}{k}$$

Logo,

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

□

Exemplo 6.2 (Tábua de Galton)**.** *A Tábua de Galton foi um dispositivo criado pelo matemático Francis Galton no século XIX com o objetivo de realizar simulações probabilísticas. Conforme ilustra a Figura 6.1, nesse dispositivo há uma abertura para a inserção de esferas que, ao serem despejadas, encontram obstáculos equidistantes dispostos em níveis, tomando dois caminhos possíveis em cada nível: esquerda ou direita. Ao final do percurso, cada esfera é coletada por canaletas na parte inferior.*

a) Considere uma tábua de Galton com 8 níveis de queda e nove canaletas numeradas de 0 a 8, da esquerda para a direita. Ao lançar uma esfera no dispositivo, qual é a probabilidade de ela ser coletada pela canaleta central? E de cair em alguma das canaletas dos extremos?

b) Em diversos vídeos na internet de reprodução desse experimento, pode-se observar um certo padrão na disposição das bolinhas depositadas nas canaletas do dispositivo. Como justificar a formação desses padrões?

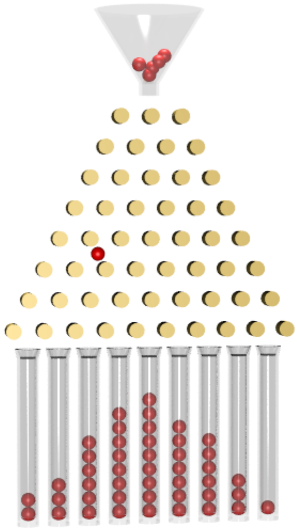

Figura 6.1: Tábua de Galton - Fonte: Wikimedia Commons

$a)$ Seja X a variável aleatória que descreve o número da canaleta em que uma esfera lançada na tábua foi coletada. Observe que, como as canaletas foram numeradas da esquerda para a direita, uma esfera cairá na canaleta 0 se, e somente se, em todos os choques com os obstáculos ela for direcionada para a esquerda (E). Ou seja, se ela seguir a sequência de quedas $EEEEEEEE$.

Analogamente, note que uma esfera será coletada na canaleta 2 se, e somente se, ela seguir **qualquer** sequência de quedas que tenha duas idas para a direita (D) e seis para a esquerda E, ou seja, uma sequência do tipo $EEDEDEEE$. E o mesmo raciocínio se aplica às demais canaletas.

Assim, a variável X (canaleta de coleta) coincide com o número de vezes que a esfera foi direcionada para a direita ou, na linguagem que usamos acima, o número de sucessos. É razoável assumir que os choques com os obstáculos são independentes entre si, e que a probabilidade de queda para esquerda ou direira em cada choque é $1/2$. Sob essa hipóteses, a variável X tem distribuição binomial com parâmetros $n = 8$ e $p = 1/2$. Logo, segue do Teorema 6.1 que

$$P(X = 4) = \binom{8}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^4 = \frac{35}{128} \simeq 0,2734$$

que é a probabilidade de uma esfera ser coletada pela canaleta central.

Por sua vez, a probabilidade de cair em alguma das canaletas extremas é menor que 1%:

$$P(X = 0) + P(X = 8) = \binom{8}{0}\left(\frac{1}{2}\right)^8 + \binom{8}{0}\left(\frac{1}{2}\right)^8 = \frac{1}{128} \simeq 0,0078$$

$b)$ Tanto na Figura 6.1 quanto em diversos vídeos e fotos encontrados na internet com a reprodução do experimento da tábua de Galton, podemos observar que, ao despejar um número grande de esferas, sua disposição nas canaletas forma aproximadamente um desenho padrão. Por mais que se repita o experimento um número grande de vezes, esse padrão aproximado é quase sempre observado.

Um primeira explicação para isso está no fato de que as probabilidades de cada esfera ser depositada nas canaletas seguem uma distribuição binomial com parâmetros n (número de níveis) e $1/2$ (parâmetro de sucesso fixo). Pela Lei dos Grandes Números, para um número grande de repetições do experimento, a frequência relativa de esferas em cada canaleta deve ser aproximadamente igual à probabilidade correspondente.

Assim, como a altura do empilhamento das esferas em cada canaleta é proporcional à sua frequência relativa, é de se esperar que o desenho formado no dispositivo aproxime-se do gráfico de distribuição de probabilidades. Na Figura 6.2, temos os gráficos de duas distribuições binomiais de probabilidade, para $n = 8$ (a tábua do exemplo) e $n = 50$. No eixo horizontal temos o números de sucessos (idas para a direita), e no vertical as probabilidades.

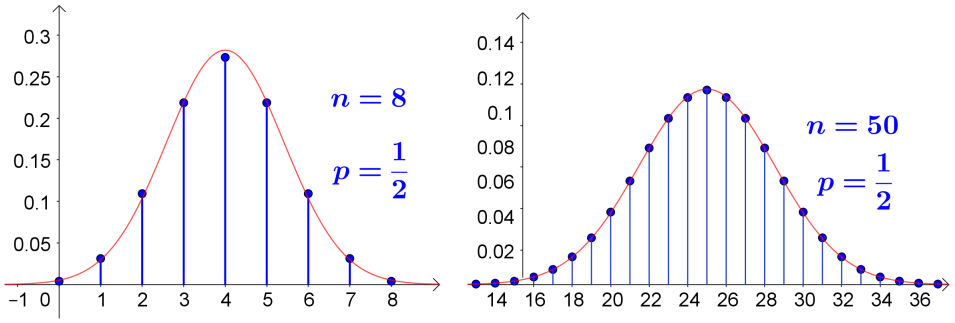

Figura 6.2: Distribuição Binomial

Há ainda outra forma de olhar para o problema. Algumas pessoas olham para o padrão formado em uma tábua de Galton e exergam o formato da famosa curva gaussiana, aquela em forma de sino que é traçada pela distribuição normal de probabilidades. A presença aproximada dessa curva tanto nos experimentos com a tábua de Galton quanto nos gráficos de distribuição binomial é sustentada em bases matemáticas.

De fato, um importante teorema estudado em cursos de Probabilidade e Estatística, conhecido como Teorema Central do Limite, afirma que, para valores cada vez maiores de n, a distribuição binomial converge em um certo sentido probabilístico para a distribuição normal.

Em termos informais, os pontos discretos da distribuição binomial aproximam-se cada vez mais do gráfico da curva normal. Isso pode ser observado na Figura 6.2, onde as curvas vermelhas são traçadas pela distribuição normal, e os pontos azuis formam o gráfico discreto da binomial. Note que para $n = 50$ a aproximação é melhor do que para $n = 8$.

Comentários. O experimento da tábua de Galton traz várias oportunidades de exploração didática da distribuição binomial. Inicialmente, professor e alunos podem buscar vários vídeos na internet com formas distintas de reproduzir o experimento. Há desde tábuas de Galton produzidas com material de baixo custo até algumas mais sofisticadas utilizando dispositivos eletrônicos em seu funcionamento. A busca pelos termos em inglês *Galton board* ou *Galton machine* pode abrir ainda mais possibilidades.

Uma atividade interessante pode ser a reprodução do experimento pelos alunos. Grupos diferentes podem construir tábuas diferentes. E a dedução de que a queda nas canaletas segue uma distribuição bonomial pode ser feita em conjunto pelos alunos, a partir de suas próprias tábuas. Bem como pode ser aproveitada a atividade para fazer estatísticas de frequências relativas, gráficos, médias, medidas de dispersão entre outras explorações de conceitos estatísticos e probabilísticos.

Exemplo 6.3. *Ao assumirmos que uma moeda é honesta, isso não significa necessariamente que em um número grande de lançamentos metade será cara, metade será coroa. Sabemos apenas, pela Lei dos Grandes Números, que a frequência relativa de caras e de coroas, converge para* $1/2$*, cada uma. Entretanto, imagine uma situação hipotética em que uma moeda foi lançada* 1000 *vezes, obtendo-se* 420 *caras. É consistente afirmar que a probabilidade de sair cara nessa moeda é igual a* $1/2$*, mesmo após observar esse resultado? Ou podemos concluir que essa moeda é viciada? Com qual segurança poderíamos concluir isso?*

Como comentado no início do capítulo, ao lançarmos várias vezes uma moeda honesta, a contagem do número X de caras (ou coroas) segue uma distribuição binomial com parâmetro de sucesso $p = 1/2$. No caso desse exemplo, temos também $n = 1000$.

Neste ponto, uma consideração importante deve ser feita. Para avaliarmos se 420 é um número de caras "aceitável" para 1000 lançamentos, uma primeira proposta seria calcular $P(X = 420)$. Se o resultado for muito próximo a zero, concluiríamos que é bastante improvável a moeda ser honesta e sair 420 caras. Mas esse raciocínio tem um problema.

Em uma distribuição binomial, quando o parâmetro n é grande, qualquer probabilidade $P(X = k)$ tende a resultar em um valor relativamente próximo a zero. Isso ocorre porque em uma distribuição com 1000 valores possíveis, é natural que cada um deles receba uma probabilidade pequena, pois a soma de todos deve ser igual a 1.

Assim, uma forma mais justa de avaliar se 420 é um número aceitável é calcular a probabilidade de ocorrência em um intervalo em torno desse número. Como veremos mais à frente, o intervalo em torno de 500 (o número esperado de caras) tem probabilidade alta, ainda que $P(X = 500)$ seja pequena.

E qual tamanho de intervalo podemos usar? Falaremos dessa questão de forma mais detalhada em um exemplo adiante. Por enquanto, diremos apenas que um bom intervalo é $\{390 \leq X \leq 450\}$. Vamos ao cálculo, utilizando o Teorema 6.1

$$P(390 \leq X \leq 450) = P(X = 390) + \cdots + P(X = 450) = \sum_{k=390}^{450} P(X = k)$$

$$= \sum_{k=390}^{450} \binom{1000}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{1000-k} = \sum_{k=390}^{450} \binom{1000}{k} \left(\frac{1}{2}\right)^{1000}$$

O leitor deve ter percebido que mesmo com uma calculadora em mãos o cálculo acima é bastante extenso. Mas aqui entra um ponto importante sobre as distribuições binomiais: a maioria dos problemas não artificiais e relevantes relacionados a essa distribuição tem n relativamente grande e/ou demanda um cálculo de probabilidade em intervalos, ou seja, um cálculo extenso como acima.

Para resolver esse problema, existem duas soluções: a aproximação pela distribuição normal, que discutiremos mais à frente, e o uso de *softwares* e/ou aplicativos. Utilizaremos a segunda para o cálculo acima.

No GeoGebra, há uma janela chamada calculadora de probabilidades. Em um de seus *menus*, é possível escolher a distribuição binomial, inserir seus parâmetros e calcular a probabilidade para qualquer intervalo. Como vantagem adicional, ainda vemos a representação gráfica da distribuição e uma tabela com todos os seus valores, no limite de aproximação de casas decimais escolhido. Conforme podemos verificar na Figura 6.3, obtivemos a probabilidade buscada:

$$P(390 \leq X \leq 450) \simeq 0,000865$$

O que nos permite concluir com segurança de $1 - 0,000865 = 0,999135 = 99,9135\%$ que uma moeda honesta não produziria nenhum resultado no intervalo $\{390 \leq X \leq 450\}$. Isso não nos permite concluir que a moeda é viciada, pois existe uma probabilidade positiva de ocorrer 420 caras ou mesmo algum número no intervalo acima, mas nos permite desconfiar com bastante segurança do equilíbrio da moeda.

Para finalizar, exibimos abaixo os resultados obtidos no GeoGebra para $P(X = 500)$ e $P(470 \leq X \leq 530)$, para fins de comparação com o resultado acima e para reforçar a questão da importância do intervalo para valores grandes de n.

$$P(X = 500) \simeq 0,0252 \quad \text{e} \quad P(470 \leq X \leq 530) \simeq 0,9463$$

Comentários. O uso de *softwares* e aplicativos no estudo de Probabilide e Estatística traz várias vantagens. O tempo gasto com cálculos extensos é mais bem aproveitado explorando conceitos e raciocínios mais profundos relacionados aos temas estudados. Além disso, permite abordar problemas mais relevantes e próximos da realidade. Em geral, problemas reais e relevantes possuem cálculos extensos, e a tecnologia é uma excelente aliada para a exploração desses problemas.

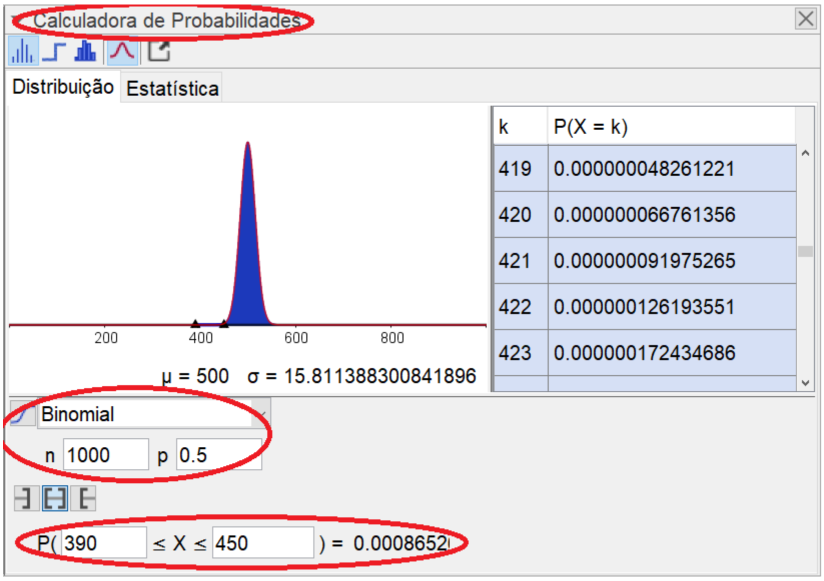

Figura 6.3: Calculadora de Probabilidades do GeoGebra

Em particular, no caso da distribuição binomial, é comum encontrar nos livros didáticos problemas cujo objetivo é calcular $P(X = k)$ para algum k ou calcular a probabilidade de X estar em algum intervalo pequeno. Isso acontece pela limitação de fazer cálculos extensos na mão ou mesmo na calculadora. Há diversos aplicativos para *smartphones* que podem resolver esse problema do cálculo, permitindo o foco em questões mais importantes. Nesse sentido, além do *software* GeoGebra, vamos sugerir mais dois aplicativos para *smartphone* em um dos exemplos da próxima seção.

Por fim, observe que o exemplo acima traz mais uma oportunidade para discussão sobre "probabilidade teórica x resultado observado", já levantada no Capítulo 2, no sentido de que raramente vamos obter 500 caras no lançamento de uma moeda 1000 vezes, mas, por outro lado, mais raramente ainda teremos resultados muito distantes de 500.

6.2 Valor Esperado

Existem dois conceitos importantes relacionados a variáveis aleatórias que em geral não são explorados no ensino básico: valor esperado e desvio padrão. Em nível superior, a abordagem aprofundada desses conceitos exige ferramentas matemáticas mais avançadas. Entretanto, entendemos que é possível fazer uma abordagem introdutória e intuitiva desses conceitos e aproveitá-los para analisar de forma mais profunda problemas reais.

O **valor esperado** de uma variável aleatória discreta é a média dos valores que ela assume ponderada pelas suas probabilidades. Denotamos por $E(X)$ o valor esperado da variável aleatória X. Assim, se X assume valores no conjunto $\{k_1, \ldots, k_\ell\}$, temos:

$$E(X) = k_1 P(X = k_1) + \ldots + k_\ell P(X = k_\ell)$$

Para ficar mais claro o aspecto intuitivo de $E(X)$, considere a variável aleatória X que descreve os resultados obtidos no lançamento de um dado honesto de 6 faces. Então temos:

$$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3,5$$

O significado teórico de $E(X) = 3,5$ reside no fato de que, **em média**, esperamos obter $3,5$ no lançamento de um dado. Do ponto de vista experimental, esse valor tem significados probabilístico e estatístico importantes, pois nas condições das hipóteses da Lei dos Grandes Números, as frequências relativas de ocorrência dos valores obtidos no dado aproximam-se das probabilidades desses valores. Ou seja, a média dos valores observados converge para $E(X)$.

Em outras palavras, se lançarmos um dado n vezes e obtivermos n_1 ocorrências da face 1, n_2 ocorrências da face 2 e assim por diante, temos que a média dos pontos obtidos é dada por:

$$\overline{X} = 1 \cdot \frac{n_1}{n} + 2 \cdot \frac{n_2}{n} + 3 \cdot \frac{n_2}{n} + 4 \cdot \frac{n_4}{n} + 5 \cdot \frac{n_5}{n} + 6 \cdot \frac{n_6}{n}$$

Pela Lei dos Grandes Números,

$$\lim_{n \to \infty} \frac{n_j}{n} = \frac{1}{6}, \ \forall j = 1, \ldots, 6 \Longrightarrow \lim_{n \to \infty} \overline{X} = E(X)$$

Assim, justificamos o fato de que $E(X)$ é também chamado de média da variável aleatória X. Em termos informais, o valor esperado é uma média teórica da variável, que se espera obter através da média atimética de resultados observados ao realizar o experimento muitas vezes.

Exemplo 6.4. *Ao lançar uma moeda honesta quatro vezes, qual é o valor esperado para o número de caras obtidas?*

Seja X o número de caras obtidas no lançamento da moeda. A distribuição de probabilidades de X é dada na tabela abaixo:

k	0	1	2	3	4
$P(X = k)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

Assim, temos:

$$\begin{aligned} E(X) &= 0 \cdot P(X=0) + 1 \cdot P(X=1) + 2 \cdot P(X=2) \\ &+ 3 \cdot P(X=3) + 4 \cdot P(X=4) \\ &= \frac{4}{16} + \frac{12}{16} + \frac{12}{16} + \frac{4}{16} = \frac{32}{16} = 2 \end{aligned}$$

O que confere com a nossa intuição. De fato, ao realizar o experimento descrito, esperamos que sejam obtidas, em média, 2 caras.

Exemplo 6.5. *No Exemplo 2.2 comentamos sobre uma outra forma de analisar qual seria a loteria mais vantajosa usando o conceito de Valor Esperado. Nesse contexto, vamos determinar o Ganho Esperado para a Mega-Sena, e deixaremos o cálculo das outras loterias como proposta de exercício para o leitor.*

Os prêmios variam muito de concurso para concurso, então tomamos as informações no *site* da Caixa Econômica Federal sobre os valores pagos a cada ganhador do concurso 2198 (16/10/2019), para utilizarmos nesse exemplo.

Cada ganhador levou:

sena: $R\$ \ 34.615.569,28$
quina: $R\$ \ 56.334,80$
quadra: $R\$ \ 859,26$

Seja X o ganho de um jogador específico que fez uma aposta simples (6 dezenas). Note que X é uma variável aleatória, e para obtermos seus os possíveis valores, devemos considerar o investimento feito na aposta. Como vamos calcular o valor esperado relativo a uma pessoa que fez apenas uma aposta simples, que custa $R\$ \ 4,50$, esse será o valor investido e deverá ser descontado do valor do prêmio.

Assim, devemos considerar que os ganhos reais ao apostar nesse concurso podem ser:

acertar a sena: $R\$ \ 34.615.569,28 - R\$ \ 4,50 = R\$ \ 34.615.564,78$
acertar a quina: $R\$ \ 56.334,80 - R\$ \ 4,50 = R\$ \ 56.330,30$
acertar a quadra: $R\$ \ 859,26 - R\$ \ 4,50 = R\$ \ 854,76$
não ganhar prêmio algum: $-R\$ \ 4,50$

As probabilidades de acertar a sena e a quadra já foram calculadas no Exemplo 2.3 e a probabilidade de acertar a quina, seguindo o mesmo raciocínio, é $\frac{81}{12.515.965}$. A probabilidade de o apostador não ganhar qualquer prêmio pode ser calculada como a probabilidade do complementar de ganhar algum prêmio. Como já temos as probabilidades de ganhar os outros prêmios, basta subtrairmos esses valores de 1 (observe que o resultado é muito próximo de 1).

$$1 - \left(\frac{1}{50.063.860} + \frac{81}{12.515.965} + \frac{4.293}{10.012.772}\right) = 1 - \frac{2.179}{5.006.386} = \frac{5.004.207}{5.006.386}$$

Assim, temos a tabela abaixo:

k	$R\$ \ 34.615.564,78$	$R\$ \ 56.330,30$	$R\$ \ 854,76$	$-R\$ \ 4,50$
$P(X=k)$	$\frac{1}{50.063.860}$	$\frac{81}{12.515.965}$	$\frac{4.293}{10.012.772}$	$\frac{5.004.207}{5.006.386}$

Portanto,

$$\begin{aligned}
E(X) &= 34.615.564,78 \cdot P(X = 34.615.564,78) \\
&+ 56.330,30 \cdot P(X = 56.330,30) + 854,76 \cdot P(X = 854,76) \\
&- 4,50 \cdot P(X = 4,50) \\
&\simeq 0,6914 + 0,3646 + 0,3665 - 4,4980 \\
&\simeq -R\$ \ 3,08
\end{aligned}$$

Logo, o ganho esperado para esse concurso da Mega-Sena é de aproximadamente $-R\$ \ 3,08$, ou seja, em média deve-se esperar um prejuízo deste valor.

Observações: 1) Como dissemos anteriormente, os valores dos prêmios variam muito em cada concurso. Então, um cálculo mais refinado do ganho esperado poderia usar média dos prêmios anteriores. Ou fazer uma estimativa do prêmio a partir do valor acumulado no concurso de interesse e de uma média envolvendo arrecadação e número de ganhadores nos últimos concursos.

2) Todo jogo de apostas tem por essência $E(X)$ negativo. Isso é uma condição necessária para que o jogo exista pois, caso contrário, quem o promove teria prejuízo. Isso significa que, a longo prazo, o apostador perde e a banca ganha.

O **desvio padrão** de uma variável aleatória discreta X, denotado por σ_X, é uma medida de dispersão dessa variável. Ele nos fornece uma informação teórica sobre o quanto os valores de X estão afastados da média $E(X)$, de forma ponderada pelas probabilidades.

Sua definição é semelhante à do desvio padrão de um conjunto de dados, com a ponderação das probabilidades no lugar da média aritmética simples. Não faz parte dos nossos objetivos desenvolver sua definição e exemplos neste material, mas acreditamos que esse conceito pode ser explorado de forma conjunta com o ensino de estatística descritiva, onde ele é abordado para um conjunto de dados.

Nosso interesse está no uso desse conceito no contexto da distribuição binomial, conforme veremos no teorema e no exemplo a seguir. Tanto a definição precisa de desvio padrão de uma variável aleatória X quanto a demonstração do teorema abaixo, que não faremos aqui, podem ser encontrados em [5].

Teorema 6.2. *Se uma variável aleatória X tem distribuição binomial com parâmetros n e p, então seu valor esperado e desvio padrão são dados por*

$$E(X) = np \quad e \quad \sigma_X = \sqrt{np(1-p)}$$

Note que a primeira fórmula do teorema acima poderia ter sido usada para obter $E(X) = 2$ no Exemplo 6.4, já que a variável X em questão tem distribuição binomial com parâmetros $n = 4$ e $p = 1/2$.

Exemplo 6.6. *Uma vacina com 70% de eficácia (como a da gripe) foi aplicada a uma população de 14500 pessoas.*

a) Qual é o valor esperado para o número de pessoas imunizadas e seu desvio padrão?

b) Qual é a probabilidade de que pelo menos 10.000 pessoas sejam imunizadas?

c) E qual é a probabilidade de que o número de pessoas imunizadas fique a no máximo dois desvios padrões de distância da média?

$a)$ Seja X o número de pessoas imunizadas na população em questão. Note que X conta o número de sucessos (imunizações) em uma sequência de 14500 experimentos independentes com probabilidade de sucesso constante igual a $0,7$. Ou seja, X tem distribuição binomial com parâmetros $n = 14500$ e $p = 0,7$. Portanto, segue do Teorema 6.2 que

$$E(X) = 14500 \cdot 0,7 = 10150 \quad e \quad \sigma_X = \sqrt{14500 \cdot 0,7 \cdot 0,3} \simeq 55,18$$

$b)$ De acordo com o Teorema 6.1, a probabilidade pedida é dada por

$$P(X \geq 10000) = P(X = 10000) + \cdots + P(X = 14500)$$
$$= \sum_{k=10000}^{14500} P(X = k) = \sum_{k=10000}^{14500} \binom{14500}{k} (0,7)^k (0,3)^{14500-k}.$$

Assim como ocorreu no Exemplo 6.3, nos deparamos com um cálculo trabalhoso demais para resolver manualmente ou apenas com calculadora. Dessa vez, vamos propor duas abordagens distintas, geralmente usadas para calcular probabilidades nesses casos.

A primeira delas é fazendo o uso de ferramentas tecnológicas. Entretanto, chamamos a atenção para o fato de que quanto maior o valor de n, mais demorado fica o processamento da calculadora de probabilidades do GeoGebra. De fato, toda ferramenta computacional tem suas limitações de processamento, mas existem aplicativos para *smartphone* que têm melhor desempenho para o cálculo de binomiais com n grande. É o caso do aplicativo *Probability Distributions*.

Na Figura 6.4, os dois primeiros *prints* mostram a tela desse aplicativo com o cálculo da probabilidade pedida e com os valores de $E(X)$ e σ_X, que também podem ser obtidos diretamente no aplicativo. Assim, verificamos que

$$P(X \geq 10000) \simeq 0,99673 = 99,673\%$$

Ou seja, temos bastante segurança de que pelo menos 10000 pessoas serão imunizadas.

A outra forma de abordar esse problema, comumente adotada em situações práticas, é utilizar uma conhecida propriedade da distribuição binomial que é consequência do Teorema Central do Limite. Como já comentamos, esse teorema é estudado em cursos de nível superior de Probabilidade e Estatística e tem importância central na área. Mas o que precisamos saber, no contexto da nossa proposta, é que ele garante que para valores suficientemente grandes de n, a distribuição binomial pode ser bem aproximada pela distribuição normal.

Esse fato pode ser observado graficamente. Na Figura 6.4, por exemplo, no primeiro *print* temos um gráfico discreto da distribuição binomial, e no terceiro, um gráfico contínuo da distribuição normal. Na escala adotada, eles são praticamente indistinguíveis.

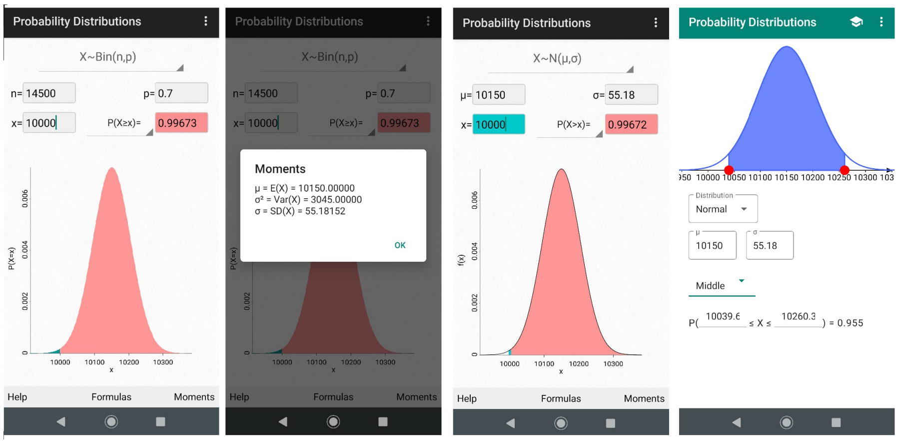

Figura 6.4: Cálculo de Probabilidades com Aplicativos - Fontes - Três primeiras telas: *Probability Distributions*; Tela mais à direita: *Probability Distributions Visualized*

Uma das vantagens de utilizar a distribuição normal como aproximação é que ela é mais facilmente encontrada em tabelas, *softwares* e em diversos *sites*. Ela resolve inclusive o problema da demora do GeoGebra, pois o cálculo da distribuição normal tem processamento mais rápido. Além disso, o aplicativo que citamos não faz cálculos para intervalos do tipo $P(a \leq X \leq b)$ e precisaremos disso para o próximo item. Então utilizaremos outro aplicativo que trabalha com esses intervalos, mas não tem a distribuição binomial, só a normal.

Voltando ao cálculo, para utilizar a distribuição normal como aproximação para a binomial, precisamos da curva normal correspondente aos nossos dados. A distribuição normal é determinada por dois parâmetros: a média, que determina a localização do ponto de máximo da curva, e o desvio padrão, que tem papel no achatamento dela. Aplicando esse dois parâmetros $E(X) = 10150$ e $\sigma_X \simeq 55,18$ obtidos para a binomial no item anterior, obtemos a curva normal que aproxima a distribuição binomial do problema. Inserindo os dados no aplicativo *Probability Distributions* (Figura 6.4) ou no GeoGebra (Figura 6.5), onde $\mu = E(X)$, obtemos:

$$P(X \geq 10000) \simeq 0,99672$$

Uma diferença notada apenas a partir da quinta casa decimal.

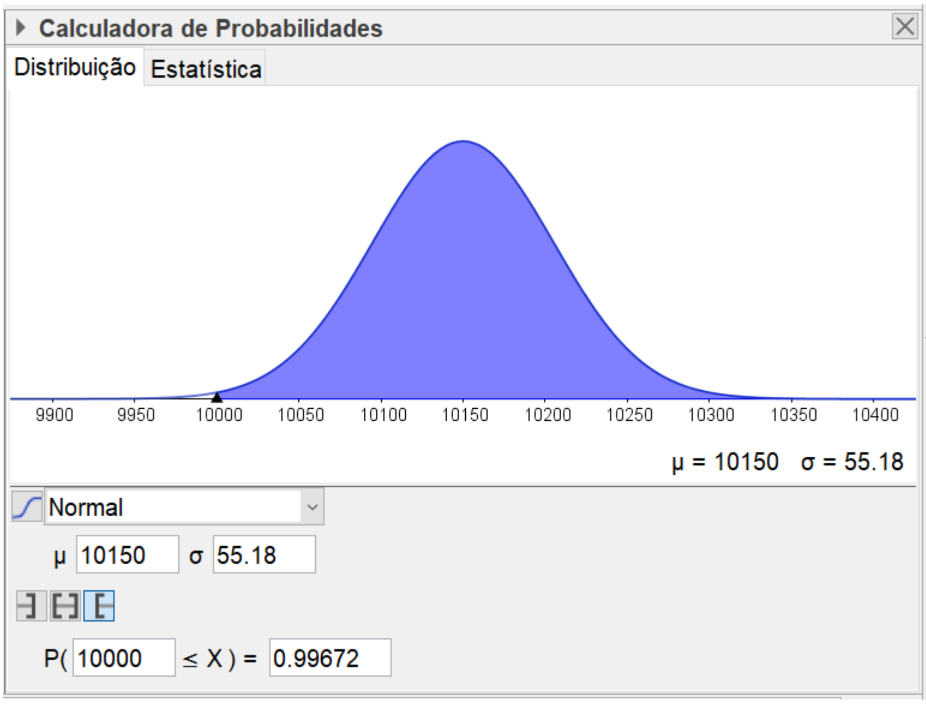

Figura 6.5: Distribuição Normal no GeoGebra

$c)$ Usando aproximação $\sigma_X \simeq 55,18$, temos

$$E(X) - 2\sigma_X \simeq 10039,64 \quad e \quad E(X) + 2\sigma_X \simeq 10260,36$$

Assim, a probabilidade pedida no enunciado é $P(10039,64 \leq X \leq 10260,36)$. Como dissemos, o primeiro aplicativo que usamos acima para distribuição binomial não dispõe desse tipo de intervalo. Mas podemos usá-lo mesmo assim. Lembrando que a variável binomial só assume valores inteiros, basta usar a seguinte estratégia:

$$\begin{aligned} P(10039,64 \leq X \leq 10260,36) &= P(X \leq 10260) - P(X \leq 10039) \\ &\overset{app}{\simeq} 0,97758 - 0,02281 = 0,95477 \end{aligned}$$

Onde o símbolo *app* significa que as aproximações foram obtidas no *Probability Distributions*.

A outra opção é utilizar a aproximação pela distribuição normal, como fizemos no item anterior. Tanto pelo GeoGebra quanto pelo aplicativo *Probability Distributions Visualized* (Figura 6.4 - à direita), obtemos:

$$P(10039,64 \leq X \leq 10260,36) \simeq 0,9545$$

o que confere de forma aproximada com o valor anterior.

Finalizamos com uma observação importante. No estudo de distribuição normal, é possível mostrar que, qualquer que seja a média μ e o desvio padrão σ, temos $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \simeq 0,9545$. Por isso esse intervalo é utilizado frequentemente em Estatística para intervalos de confiança para o nível de 95%. E por isso também que utilizamos um intervalo de tamanho 30 (um pouco menos que 2σ) no Exemplo 6.3.

Conhecendo essa propriedade, teremos sempre, sem fazer contas, a probabilidade do intervalo de dois desvios padrões de distância da média para qualquer problema envolvendo distribuição binomial, desde que n seja suficientemente grande.

Comentários. Sabemos que a distribuição normal somente é abordada em cursos de nível superior. No ensino de distribuição binomial, porém, ao fazer gráficos para valores grandes de n, a curva gaussiana aparece naturalmente. Entendemos que essa pode ser uma oportunidade para um primeiro contato intuitivo com a distribuição normal, sem a necessidade de ferramentas avançadas e sem a necessidade de definir formalmente tal distribuição e a função que determina seu gráfico, apenas utilizando ferramentas tecnológicas e mencionando algumas propriedades intuitivas, como a aproximação com a binomial, a localização da média, as características de dispersão entre outras.

A distribuição normal é, sem dúvida, a distribuição mais importante da Estatística e está presente em praticamente todas as suas aplicações. Em muitos países, o seu estudo introdutório faz parte do ensino básico. Ainda que não esteja presente na nossa BNCC, entendemos que as distribuições binomial e normal, no contexto acima, podem e devem ser um tópico central no ensino de Probabilidade e Estatística.

Sobre esse exemplo, chamamos a atenção ainda para a abordagem das medidas resumo média e desvio padrão, que têm papel central em qualquer análise estatística. Ao estudar estatística descritiva no ensino médio, o aluno já tem o primeiro contato formal e intuitivo com esses conceitos. Estendê-los para a distribuição binomial traz muitos ganhos, ainda que não seja adequado fazer a demonstração das fórmulas obtidas no Teorema 6.2. Observe que a análise de um problema real de vacinação ficou mais rica com essas medidas, que podem ser obtidas diretamente nos aplicativos ou pelas fórmulas.

6.3 Distribuição Geométrica

Entre as distribuições discretas de probabilidade, as duas abordadas neste capítulo são as mais importantes no contexto da modelagem de problemas envolvendo variáveis aleatórias discretas. Entretanto, só a distribuição binomial tem lugar em textos didáticos do ensino médio. A distribuição geométrica, além de ser de compreensão intuitiva e modelar uma série de problemas relevantes, traz uma oportunidade para a aplicação e contextualização das progressões geométricas, com as quais os alunos costumam ter contato no início do ensino médio. Vamos explorar um exemplo para introduzir seu conceito.

Exemplo 6.7 (Exemplo Introdutório)**.** *Um dado honesto de seis faces é lançado sucessivas vezes até que se obtenha o número 6. O interesse do experimento é no número de lançamentos necessários até que se obtenha sucesso no objetivo buscado.*

a) Qual é o espaço amostral desse experimento?

b) Qual é a probabilidade de que o sucesso seja obtido na terceira tentativa? E na quinta? Há uma expressão geral para a probabilidade de sucesso na k-ésima tentativa?

c) Como podemos justificar que o axioma $P(\Omega) = 1$ é satisfeito nesse exemplo?

d) Qual é a probabilidade de que o sucesso venha em um lançamento de número par? E em um lançamento de número ímpar? Os resultados obtidos são intuitivos?

$a)$ Como o interesse do experimento é o número de lançamentos necessários até que se obtenha o número 6, então o espaço amostral deve conter os números inteiros positivos 1, 2, 3, ...

Mas há algum limite para o número de lançamentos possíveis até o sucesso? Sabemos que é bastante improvável sair uma sequência grande de números no dado sem que o 6 apareça, mas não temos como colocar um limite que nos dê a certeza de que o 6 sairá até ali. Então nossa única alternativa é colocar $\Omega = \mathbb{N}^* = \{1, 2, 3, ...\}$.

$b)$ Vamos denotar por F (fracasso) o evento em que lançamos o dado e não obtemos o 6, e por S (sucesso) o evento no qual obtemos o 6. Claramente temos $P(F) = 5/6$ e $P(S) = 1/6$.

Seja X a variável aleatória que descreve o número de lançamentos feitos até a obtenção da face 6. O sucesso será obtido na terceira tentativa se, e somente se, tivermos a sequência de resultados FFS. Como os lançamentos são independentes entre si, segue que:

$$P(X=3)=P(FFS)=P(F)^2P(S)=\left(\frac{5}{6}\right)^2\cdot\frac{1}{6}=\frac{25}{216}$$

Analogamente:

$$P(X=5)=P(FFFFS)=\left(\frac{5}{6}\right)^4\frac{1}{6}=\frac{625}{7776}$$

De forma geral, para o sucesso na k-ésima tentativa, devemos ter $k-1$ fracassos nos primeiros lançamentos e sucesso no último. Portanto:

$$P(X=k)=P\left(\underbrace{FF\cdots F}_{k-1\ fracassos}\ S\right)=\left(\frac{5}{6}\right)^{k-1}\cdot\frac{1}{6}$$

c) Observando a expressão geral $P(X=k)$ obtida acima ou mesmo a sequência de probabilidades

$$\begin{aligned}&(P(X=1),P(X=2),P(X=3),\cdots,P(X=k),\cdots)\\ =&\left(\frac{1}{6},\frac{5}{36},\frac{25}{216},\cdots,\left(\frac{5}{6}\right)^{k-1}\cdot\frac{1}{6},\cdots\right)\end{aligned}$$

podemos perceber que essa sequência é uma progressão geométrica (PG) cujo primeiro termo é $a_1=1/6$ e a razão é $q=5/6$. Daí a justificativa para o nome dessa distribuição de probabilidades.

Assim, para obter $P(\Omega)$, podemos usar a conhecida fórmula para soma S dos infinitos termos de uma PG com razão menor que 1:

$$S=\frac{a_1}{1-q}$$

Logo,

$$\begin{aligned}P(\Omega)=P(X\in\{1,2,3,...\})&=\sum_{k=1}^{\infty}P(X=k)=\sum_{k=1}^{\infty}\left(\frac{5}{6}\right)^{k-1}\cdot\frac{1}{6}\\ &=\frac{1}{6}+\frac{5}{36}+\frac{25}{216}+\cdots=\frac{1/6}{1-5/6}=1\end{aligned}$$

d) Seja $A\subset\Omega$ o evento no qual o sucesso é obtido em um lançamento de número par. Pelo fórmula obtida no item b), temos:

$$\begin{aligned}P(A)=P(X\in\{2,4,6,...\})&=\sum_{k=1}^{\infty}P(X=2k)=\sum_{k=1}^{\infty}\left(\frac{5}{6}\right)^{2k-1}\cdot\frac{1}{6}\\ &=\frac{5}{36}+\frac{125}{1296}+...\end{aligned}$$

onde soma acima é a soma infinita dos termos de uma PG cujo primeiro termo é $a_1 = 5/36$ e a razão é $q = 25/36$. Portanto:

$$P(A) = \frac{5}{36} + \frac{125}{1296} + ... = \frac{5/36}{1 - 25/36} = \frac{5/36}{11/36} = \frac{5}{11}$$

Por outro lado, o sucesso em um lançamento de número ímpar é dado pelo evento A^C. Logo,

$$P\left(A^C\right) = 1 - P(A) = 1 - \frac{5}{11} = \frac{6}{11}$$

Uma análise apressada poderia afirmar equivocadamente que as probabilidades de sucesso em um lançamento par ou ímpar devem ser iguais, ou seja, $1/2$ para cada. Essa seria uma afirmação razoável, baseada na intuição. Mas note que os números ímpares saem na frente, pois a probabilidade de sair o número 6 logo de cara, no lançamento de número 1, é ligeiramente maior que a de obter o sucesso no lançamento 2. Essa comparação continua valendo para os lançamentos 3 e 4, 5 e 6 etc. Como sabemos, é sempre bom desconfiar da intuição quando se trata de probabilidades.

A situação modelada pelo exemplo acima ocorre com frequência em problemas probabilísticos. Uma sequência de experimentos independentes, com dois resultados possíveis (sucesso ou fracasso) com probabilidade constante, em que se deseja contar o número X de repetições necessárias até a obtenção do primeiro sucesso. A distribuição de probabilidades de qualquer variável aleatória X nessas condições é generalizada na definição a seguir.

Definição 6.2. *Considere uma sequência de experimentos independentes com dois resultados de interesse: sucesso, com probabilidade p, e fracasso com probabilidade $1-p$. Seja X a variável aleatória que descreve o número de realizações até a obtenção do primeiro sucesso. A distribuição de probabilidades de X é chamada de Distribuição Geométrica com parâmetro p.*

Teorema 6.3. *Se X é uma variável aleatória com distribuição geométrica de parâmetro p, então*

1) $P(X = k) = (1-p)^{k-1}p \quad \forall k \in \{1, 2, 3, ...\}$

2) O valor esperado e o desvio padrão de X são dados por

$$E(X) = \frac{1}{p} \quad e \quad \sigma_X = \sqrt{\frac{1-p}{p^2}}$$

Observações: 1) A demonstração do primeiro item do teorema acima é direta e segue a mesma lógica do item b) Exemplo 6.7. O evento $\{X = k\}$ corresponde a uma sequência de k eventos independentes em que os $k-1$ iniciais são fracassos e o k-ésimo é um sucesso. Por isso o produto $(1-p)^{k-1}p$.

$2)$ Ainda seguindo o exemplo anterior, podemos verificar que a distribuição geométrica de probabilidades está bem definida, no sentido de que a soma de todas as probabilidades possíveis resulta em 1. Basta somar os termos infinitos da PG formada pela sequência $P(X=k)=(1-p)^{k-1}p$ para $k=1,2,3,...$:

$$p+(1-p)p+(1-p)^2p+\cdots=\frac{p}{1-(1-p)}=1$$

$3)$ Não demonstraremos o segundo item do Teorema, por não fazer parte dos nossos objetivos. Sua demonstração pode ser encontrada em [5].

$4)$ O modelo geométrico tem uma variante na qual X conta o número de fracassos antes do primeiro sucesso. Nesse caso, as fórmulas da distribuição e do valor esperado mudam, respectivamente, para

$$P(X=k)=(1-p)^kp \quad \forall k\in\{0,1,2,...\} \quad e \quad E(X)=\frac{1-p}{p}$$

A fórmula para o desvio padrão permanece a mesma.

Exemplo 6.8. *Considere um experimento que seleciona pessoas na rua de forma independente e aleatória dispostas a participar de uma pesquisa de opinião que dura trinta minutos. Em média, sabe-se que uma a cada dez pessoas topa participar da pesquisa. Seja X o número de pessoas abordadas até que a primeira aceite participar da pesquisa.*

a) Determinar a probabilidade de encontrar um participante até a sexta tentativa.

b) Determinar a probabilidade de que sejam necessárias pelo menos vinte abordagens até que se encontre alguém que tope participar da pesquisa.

c) O que é mais provável: encontrar alguém que tope participar antes do valor esperado para X ou a partir dele?

d) Obter um gráfico da distribuição de probabilidades em questão. Qual gráfico de função contínua elementar contém os pontos desse gráfico?

$a)$ Pela descrição do experimento, a variável X tem distribuição geométrica com parâmetro $p=1/10=0,1$, que é a probabilidade de encontrar alguém que tope participar da pesquisa. Para encontrar a probabilidade pedida, ou seja, do evento $\{X\leq 6\}$, podemos usar a fórmula da soma dos primeiros termos de uma PG, como se segue:

$$\begin{aligned} P(X \leq 6) &= P(X = 1) + P(X = 2) + \cdots + P(X = 6) \\ &= 0,1 + 0,9 \cdot 0,1 + 0,9^2 \cdot 0,1 + \cdots 0,9^5 \cdot 0,1 \\ &= 0,1 \frac{0,9^6 - 1}{0,9 - 1} \simeq 0,4686 = 46,86\% \end{aligned}$$

$b)$ Realizar pelo menos 20 tentativas corresponde ao evento $\{X \geq 20\}$. Utilizando a fórmula da soma dos infinitos termos de uma PG, obtemos:

$$\begin{aligned} P(X \geq 20) &= \sum_{k=20}^{\infty} P(X = k) = 0,9^{19} \cdot 0,1 + 0,9^{20} \cdot 0,1 + \cdots \\ &= \frac{0,9^{19} \cdot 0,1}{1 - 0,9} = 0,9^{19} \simeq 0,1351 = 13,51\% \end{aligned}$$

Poderíamos ainda ter optado pelo seguinte raciocínio: evento “pelo menos 20 tentativas” é equivalente ao evento “ocorrer 19 fracassos”. Ou seja,

$$P(X \geq 20) = P\left(\underbrace{FFF \cdots F}_{19\ fracassos}\right) = 0,9^{19} \simeq 0,1351$$

$c)$ De acordo com o Teorema 6.3, temos $E(X) = 10$. Assim, encontrar alguém que tope antes do valor esperado corresponde ao evento $\{X < 10\}$. Dessa forma, aplicando novamente a fórmula da soma da PG, temos:

$$\begin{aligned} P(X < 10) &= \sum_{k=1}^{9} 0,9^{k-1} \cdot 0,1 = 0,1 \frac{0,9^9 - 1}{0,9 - 1} \simeq 0,6126 \\ \iff P(X \geq 10) &\simeq 1 - 0,6126 = 0,3874 \end{aligned}$$

Assim, por um lado, $E(X) = 10$ significa que, em média, devemos esperar 10 tentativas até obter o sucesso. Por outro lado, é mais provável conseguirmos alguém que tope antes das 10 tentativas do que a partir delas.

De modo informal, esse resultado contraintutitivo ocorre porque, no cálculo de $E(X)$, existem infinitos valores após o 10 que, ainda que tenham probabilidade pequena, “puxam” o valor esperado para cima.

Do ponto de vista experimental, realizações repetidas desse experimento produzirão majoritariamente valores menores que 10, mas para um número grande de repetições, a ocorrência de valores altos será suficiente para fazer a média convergir para 10.

Em geral, pode-se mostrar que a distribuição das probabilidades antes e a partir do valor esperado de uma distribuição geométrica depende do parâmetro p; mas que sempre teremos ao menos 50% da distribuição antes do valor esperado, sendo igual a 50% somente quando $p = 1/2$.

Note que nas distribuições normal e binomial (para n grande) vistas nas seções anteriores essa situação não ocorre, já que as probabilidades são distribuídas de forma simétrica em torno da média (aproximadamente, no caso da binomial com n grande), conforme ilustra a figura 6.6.

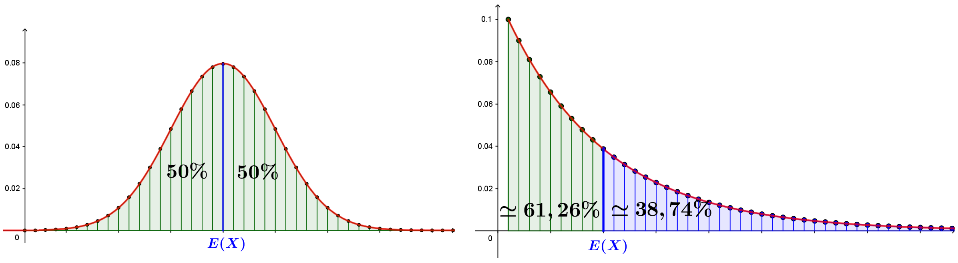

Figura 6.6: Valor Esperado e Distribuições

$d)$ Na Figura 6.7, temos em azul o gráfico discreto de pontos da distribuição geométrica com parâmetro $p = 0,1$. Como a sequência de valores $P(X = k)$, para $k = 1, 2, 3, ...$, forma uma PG, sabemos que seu termo geral é uma função exponencial restrita a valores naturais. Assim, o gráfico da função contínua $f(x) = 0,9^{x-1} \cdot 0,1$ contém os pontos do gráfico da distribuição.

Essa representação gráfica ilustra o fato intuitivo de que a probabilidade de esperar muito tempo para obter o sucesso decresce com o tempo, tendendo a zero.

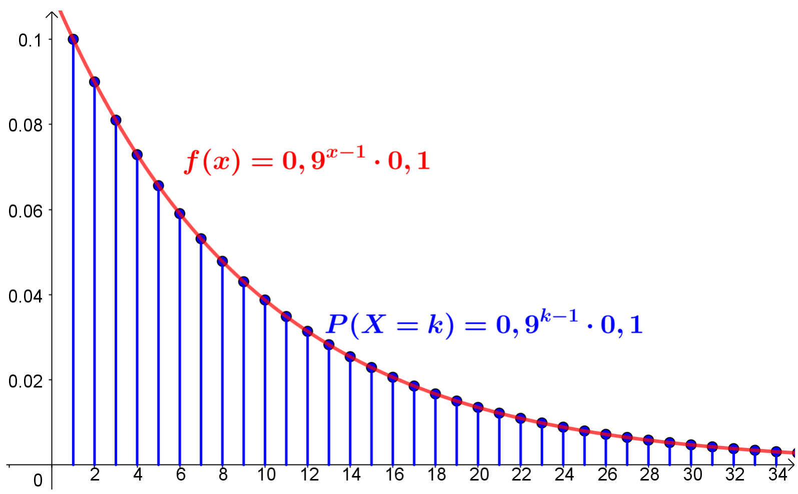

Figura 6.7: Distribuição Geométrica e Função Exponencial

Comentários. Da mesma forma que comentamos sobre a distribuição binomial, problemas reais envolvendo distribuição geométrica envolvem mais frequentemente eventos do tipo intervalo, como $\{X \leq k\}$ ou $\{X \geq k\}$, do que eventos pontuais como $\{X = k\}$. Por isso, no caso da distribuição geométrica, as fórmulas de soma de termos de uma PG finita ou infinita são usadas com bastante frequência. Contudo, para essa distribuição também temos a opção de usar aplicativos como o *Probability Distributions*, citado anteriormente.

Exemplo 6.9. *O setor de manutenção de uma indústria faz revisões diárias nas máquinas, paralisando a produção por um tempo determinado. Em média, 98,3% das inspeções não encontram problemas. Quando, porém, algum problema é encontrado, a retomada da produção pode atrasar por horas até que ele seja resolvido.*

a) A diretoria da empresa deve se planejar para quantas paralisações anuais, em média?

b) Sabendo que já se passaram 20 dias da última manutenção, qual é a probabilidade de que se passem mais três dias sem interrupção da produção fora do intervalo de inspeção?

$a)$ Estamos interessados na quantidade de dias decorridos até que seja necessária a próxima paralisação para manutenção. Assim, utilizando a terminologia do modelo geométrico, cada dia que passa sem que seja encontrado o problema é um "fracasso" com probabilidade $0,983$. E, portanto, o "sucesso" tem parâmetro $p = 1 - 0,983 = 0,017$. Logo, segue do Teorema 6.3 que o valor esperado para o número de dias entre manutenções é:

$$E(X) = \frac{1}{0,017} \simeq 58,8$$

Ou seja, a empresa deve se planejar para que, em média, haja de 6 a 7 paralisações anuais da produção.

$b)$ Note que a probabilidade pedida é de que se passem 23 dias sem manuntenção ($\{X > 23\}$), condicionada ao evento $\{X > 20\}$. Assim, usando a definição de probabilidade condicional obtemos

$$P(X > 23|X > 20) = \frac{P(\{X > 23\} \cap \{X > 20\})}{P(X > 20)} = \frac{P(X > 23)}{P(X > 20)}$$

A interseção acima resulta no evento $\{X > 23\}$, pois $\{X > 23\} \subset \{X > 20\}$. Isso ocorre porque $X > 23 \Rightarrow X > 20$.

O cálculo das probabilidades na fração acima pode ser resolvido por soma de PG ou usando o raciocínio empregado no item $b)$ do Exemplo 6.8, que é o que faremos:

$$P(X > 23) = P\left(\underbrace{FFF\cdots F}_{23\ fracassos}\right) = 0,983^{23}$$

$$P(X > 20) = P\left(\underbrace{FFF\cdots F}_{20\ fracassos}\right) = 0,983^{20}$$

Logo,

$$P(X > 23|X > 20) = \frac{0,983^{23}}{0,983^{20}} = 0,983^{3} = P(FFF) = P(X > 3) \simeq 0,95$$

O resultado acima ilustra uma propriedade da distribuição geométrica conhecida como "falta de memória". Isso significa que as ocorrências passadas nessa distribuição não afetam condicionalmente as probabilidades futuras. No caso acima, receber a informação de que já se passaram 20 dias sem manutenção, não faz a probabilidade de que "se passem mais 3 dias" seja afetada. É como se estivéssemos "partindo do zero" com uma contagem nova. Essa propriedade é uma caracterização da distribuição geométrica no conjunto das distribuições discretas e será generalizada no teorema a seguir.

Comentários. Um cuidado necessário ao trabalhar tanto com a distribuição binomial quanto com a geométrica é na utilização da terminologia "fracasso" e "sucesso", que nem sempre são usadas de acordo com o sentido real do termo. No exemplo acima, o que foi chamado de fracasso é, na realidade, considerado um sucesso, pois é um dia sem defeitos nas máquinas. Por isso é importante compreender, mais do que esses termos, a estrutura dos raciocínios que embasam as sequências de eventos que caracterizam as distribuições em questão.

Ainda sobre o exemplo acima, a propriedade da "falta de memória" é útil para mostrar situações reais contraintuitivas em que, mesmo que tenha passado muito tempo sem a ocorrência de um evento, ela não se torna mais provável por causa disso. Nesse sentido, é importante não confundir tempos longos sem ocorrência, que são improváveis na distribuição geométrica, com a contagem a partir de um certo momento, dada a informação passada.

Teorema 6.4 (Falta de Memória)**.** *Se X é uma variável aleatória com distribuição geométrica de parâmetro p, então, para todo $k, j \in \mathbb{N}$, tem-se:*

$$P(X > k + j|X > j) = P(X > k)$$

Demonstração. De forma análoga ao feito no Exemplo 6.9, temos

$$P(X > k + j|X > j) = \frac{P(\{X > k + j\} \cap \{X > j\})}{P(X > j)} = \frac{P(X > k + j)}{P(X > j)}$$

Usando soma de PG, obtemos:

$$P(X > k + j) = \sum_{i=k+j+1}^{\infty} (1 - p)^{i-1} p = \frac{(1 - p)^{k+j} p}{1 - (1 - p)} = (1 - p)^{k+j}$$

$$P(X > j) = \sum_{i=j+1}^{\infty} (1 - p)^{i-1} p = \frac{(1 - p)^j p}{1 - (1 - p)} = (1 - p)^j$$

Logo,

$$\begin{aligned} P(X > k + j|X > j) &= \frac{P(X > k + j)}{P(X > j)} = \frac{(1 - p)^{k+j}}{(1 - p)^j} \\ &= (1 - p)^k = P(X > k) \end{aligned}$$

□

6.4 Problemas Propostos

Problema 6.1. *Uma doença pode ser curada por meio de cirurgia em 80% dos casos. Foram selecionados 300 pacientes portadores dessa doença que serão submetidos a cirurgia.*
a) Qual a probabilidade de que pelo menos 250 pacientes sejam curados? E de que no máximo 280 sejam curados?
b) Determine o valor esperado e o desvio padrão do número de pacientes curados entre os 300 selecionados.
c) Determine a probabilidade de que o número de curados fique a uma distância máxima de um desvio padrão da média.
d) Obtenha uma representação gráfica da distribuição em questão e de sua aproximação normal utilizando o GeoGebra. Em seguida, calcule as probabilidades pedidas nos itens anteriores utilizando a aproximação pela normal.

Problema 6.2. *Um banco de sangue necessita de sangue do tipo O-Rh negativo. Suponha que* $p = 0,05$ *seja a proporção de indivíduos na população com esse tipo de sangue. Uma equipe do banco selecionará indivíduos na população até que se encontre algum portador do tipo procurado para solicitar a doação.*
a) Qual é a probabilidade de que seja necessário examinar ao menos 15 pessoas antes de encontrar o tipo desejado? E de encontrar a pessoa até a décima tentativa?
b) Em média, quantas pessoas devem ser selecionadas até que se encontre o tipo procurado?
c) Dado que já foram examinadas pessoas em um número igual à média, qual é a probabilidade de se examinar ao menos esse número outra vez?
d) Obtenha uma representação gráfica da distribuição em questão e de sua correspondente exponencial utilizando o GeoGebra.

Problema 6.3. *Para um exame com 25 questões do tipo verdadeiro ou falso, um estudante sabe a resposta correta de 17 questões e responde as demais "chutando".*
a) Calcule a probabilidade de que o aluno acerte pelo menos 90% das respostas.
b) Determine a média e o desvio padrão do número de acertos.
c) Suponha que outro estudante que vai fazer esse mesmo exame saiba a resposta de 15 questões e tenha probabilidade de acerto nas demais de 0,7. De qual dos estudantes espera-se o melhor resultado?
d) Qual dos estudantes terá desempenho mais homogêneo?

Problema 6.4. *Determine o ganho esperado para Lotofácil e Quina, utilizando os dados do Exemplo 2.2.*

Problema 6.5. *Qual é o número esperado de casas que um jogador deve andar no jogo de tabuleiro descrito no Exemplo 3.5?*

Problema 6.6. *Ao ministrar um remédio em testes a 10000 voluntários, observou-se efeito positivo em 6815 deles. Utilizando a estratégia aplicada ao Exemplo 6.3, avalie a seguinte hipótese: a taxa de sucesso do remédio é de 0,7.*

Problema 6.7. *a) Seja X uma variável aleatória com distribuição binomial de parâmetros n e 0.5. Mostre que se n é ímpar, então $P(X < E(X)) = P(X \leq E(X)) = 0,5$.*
b) O que acontece com essas probabilidades para n par?
c) Com a ajuda do GeoGebra ou do aplicativo Probability Distributions, analise o que ocorre com o caso do item anterior quando n cresce arbitrariamente.
d) E o que ocorre com os dois casos, n par ou ímpar, quando tomamos parâmetros quaisquer p e n, em cada uma das condições acima?

Problema 6.8. *a) Dada uma variável aleatória X com distribuição geométrica de parâmetro p, determine uma expressão em função de p para as probabilidades $P(X > E(X))$ e $P(X \geq E(X))$ e mostre que $P(X \geq E(X)) \leq 0,5$.*
b) Sob quais condições as distribuições das probabilidades antes e a partir do valor esperado se aproximam de 50%?
c) Mostre que quando o parâmetro p se aproxima de zero, as probabilidades $P(X > E(X))$ e $P(X \geq E(X))$ convergem para e^{-1}; e que quando p se aproxima de 1, essas probabilidades convergem para zero.

Capítulo 7

Probabilidade, Matrizes e Sistemas Lineares: as Cadeias de Markov

"A Ciência não pode prever o que vai acontecer. Só pode prever a probabilidade de algo acontecer."

– César Lattes

As cadeias de Markov constituem um tópico moderno na Teoria das Probabilidades que surgem para modelar problemas envolvendo sequências de variáveis aleatórias com uma característica específica que será detalhada neste capítulo. Suas aplicações são incontáveis e ocorrem em áreas como Física, Computação, Biologia, Teoria dos Jogos, Economia, entre outras.

Tais aplicações ocorrem desde problemas simples como os que apresentaremos aqui de forma acessível ao público do ensino médio, até problemas mais avançados que exigem técnicas e ferramentas de nível superior. Faremos uma apresentação introdutória das cadeias de Markov e mostraremos como elas podem ser utilizadas para contextualizar conceitos algébricos estudados no ensino médio, como multiplicação e potências de matrizes, algo geralmente percebido por alunos desse segmento como sem utilidade.

7.1 Noções sobre Sequências de Variáveis Aleatórias

Antes de apresentarmos conceito, definição e propriedades das cadeias de Markov, é importante compreender de forma intuitiva o conceito de sequência de variáveis aleatórias. Um exemplo simples pode ser obtido através de uma sequência de lançamentos consecutivos de um dado honesto de seis faces. Uma sequência possível de resultados obtidos pode ser dada por

$$(1, 2, 4, 6, 3, 4, 1, 6, 6, 2, 3, 4, 3, 4, 6, 4, 4, 3, 5, 2, 5, 5, 1, 4, 6, ...)$$

Observe que temos uma sequência de números reais obtidos de forma aleatória. Sequências desse tipo, conhecidas como sequências de variáveis aleatórias (V.A.s), são denotadas por $(X_n)_{n\geq 0}$ ou $(X_0, X_1, X_2, ...)$. Elas diferem de uma sequência de números reais, como $(x_n)_{n\geq 0}$ dada por $x_n = n^2$, pois seus termos $(0, 1, 4, 9, ...)$ são determinísticos e únicos a cada valor de n.

No caso do lançamento de um dado, a sequência de V.A.s tem a chamada propriedade *i.i.d.*, pois as variáveis $X_0, X_1, X_2, ...$ são *independentes e identicamente distribuídas*. Dizer que as variáveis X_n são identicamente distribuídas significa que as probabilidades de cada variável $X_0, X_1, X_2, ...$ assumir um determinado valor é sempre a mesma. No caso do dado honesto, cada um dos seis valores possíveis para as variáveis em qualquer um dos lançamentos tem probabilidade igual a $1/6$. Por exemplo, $P(X_3 = 5) = P(X_{11} = 5) = 1/6$.

De forma geral, dados quaisquer valores possíveis x e y para a sequência de variáveis aleatórias $(X_n)_{n\geq 1}$, a propriedade *i.i.d.*. de uma sequência de V.A.s pode ser descrita pelas igualdades

$$1)\ P(X_n = x|X_m = y) = P(X_n = x)\ \forall m, n \in \mathbb{N} \quad (Independência)$$
$$2)\ P(X_n = x) = P(X_0 = x)\ \forall n \in \mathbb{N} \quad (Distribuição\ Idêntica)$$

Problemas probabilísticos envolvendo sequências de variáveis aleatórias *i.i.d.* tiveram papel central no desenvolvimento da Teoria das Probabilidades e encontram incontáveis aplicações nas ciências modernas. Exemplos desse tipo de sequência foram abordados neste material no Capítulo 6, com a distribuição geométrica de probabilidades. A sequência de sucessos e fracassos que definem essa distribuição pode ser vista uma uma sequência de V.A.s *i.i.d.* que assume os valores 0 (fracasso) e 1 (sucesso).

7.2 Conceito e Exemplo Introdutório

As cadeias de Markov são sequências de V.A.s com uma propriedade mais geral que a propriedade *i.i.d.*. Elas são utilizadas para modelar problemas envolvendo sequências de variáveis aleatórias com a característica de que a probabilidade de determinado valor ocorrer no tempo n (ou seja, $P(X_n = x)$) depende condicionalmente apenas do valor assumido no tempo anterior da cadeia, ou seja, de X_{n-1}.

Com isso, as variáveis $X_0, X_1, X_2, ...$ não são independentes, pois a probabilidade de cada valor assumido depende do valor assumido anteriormente. E também não são identicamente distribuídas em geral, como ficará claro nos exemplos a seguir.

Antes de definir formalmente uma cadeia de Markov e estudar suas propriedades, vamos explorar um exemplo de forma intuitiva, utilizando conceitos de probabilidade já vistos nos capítulos anteriores, com o objetivo de familiarizar o leitor com as ideias descritas acima e com as definições que virão.

Observamos que alguns aspectos explorados no exemplo abaixo são direcionados mais ao público de professores de Matemática e têm como objetivo uma compreensão mais detalhada do conceito de cadeia de Markov. Por exemplo, o item c) do próximo exemplo e algumas justificativas mais formais feitas nos outros itens não são necessários para uma abordagem introdutória do tema em sala de aula do ensino básico.

Por outro lado, entendemos que os cálculos baseados em diagramas de árvores e matrizes compõem o principal aspecto de exploração didática do tema com alunos do ensino básico. Acreditamos que a introdução às cadeias de Markov pode ser feita de forma bastante intuitiva explorando esses aspectos.

Exemplo 7.1. *Considere uma locadora de automóveis que tem as lojas 1, 2 e 3 em uma determinada cidade. Um cliente que aluga um carro numa loja pode devolver em qualquer outra. O gerente da locadora estimou, através das frequências de devolução no ano passado, todas as probabilidades de que um carro seja devolvido na loja j, quando retirado na loja i. Ele organizou essas probabilidades na matriz $Q = (q_{ij})$ dada abaixo, onde cada elemento q_{ij} representa a probabilidade de devolução na loja j, dado que o carro foi retirado na loja i, para $i, j = 1, 2, 3$. Por exemplo, quando um cliente retira o carro na loja 3, a probabilidade de devolver na loja 1 é dada por $q_{31} = 40\%$.*

$$Q = \begin{bmatrix} 0,7 & 0,2 & 0,1 \\ 0,5 & 0,4 & 0,1 \\ 0,4 & 0,2 & 0,4 \end{bmatrix}$$

Note que a soma das linhas da matriz Q resulta em 1, pois se um carro é retirado na loja i, ele tem que ser devolvido ou na loja 1, ou na 2, ou na 3. Ou seja, $q_{i1} + q_{i2} + q_{i3} = 1$ para $i = 1, 2, 3$. Uma matriz com tal propriedade e com entradas não negativas é chamada de *matriz estocásitca*.

Fixado um carro dessa rede de locadoras, seja X_0 a loja em que ele se encontra hoje, X_1 a loja em que será devolvido após a próxima vez que for alugado e assim sucessivamente. Dessa forma, obtemos uma sequência de variáveis aleatórias $(X_0, X_1, X_2, ...)$ a cada aluguel e devolução do carro.

Note que a sequência acima é aleatória, pois em geral não podemos determinar em qual loja um carro estará após a n-ésima devolução. Mas, como veremos, podemos determinar as probabilidades de cada uma das alocações possíveis.

Sendo assim, representaremos em uma matriz linha 1×3 as probabilidades de um determinado carro estar nas lojas 1, 2 ou 3 após a n-ésima devolução (o que chamaremos de tempo n), denotando-a por

$$D_n = \begin{bmatrix} p_1 & p_2 & p_3 \end{bmatrix}$$

Por exemplo, se o carro em questão está na loja 1 no início do processo, temos:

$$D_0 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \quad \text{e} \quad D_1 = \begin{bmatrix} 0,7 & 0,2 & 0,1 \end{bmatrix}$$

pois estando na loja 1 no "tempo 0", pode ser devolvido nas lojas 1, 2 e 3 no "tempo 1" com probabilidades $0,7$, $0,2$ e $0,1$, respectivamente.

Segue-se daí um questionamento natural: temos como determinar D_2, D_3,...,D_n? Responderemos esta e as seguintes questões sobre essa situação:

a) É possível determinar a probabilidade de um carro ser devolvido na loja 2 após a primeira locação, ou seja, $P(X_1 = 2)$?

b) Qual é a probabilidade de um carro ser devolvido na loja 2 após a primeira locação, dado que foi retirado na loja 1? E de ser devolvido na loja 1 após a segunda locação, dado que foi retirado na loja 1 na primeira locação?

c) A sequência $(X_n)_{n\geq 0}$ definida acima é i.i.d.? Por quê?

d) Se um carro foi retirado na loja 2, a matriz probabilidades de ele estar em cada loja no tempo 0 é dada por $D_0 = [0 \quad 1 \quad 0]$. Qual é a matriz de probabilidades D_1 de devolução após a primeira locação? E matriz D_2, após a segunda locação? É possível obter uma expressão simplificada para obter D_n?

e) Suponha que as porcentagens de carros alocados nas lojas 1, 2 e 3 em um determinado momento sejam, respectivamente, 15%, 30% e 55%. Escolhendo aleatoriamente um carro da empresa, qual é a probabilidade de ele estar na loja 1 após cinco locações?

f) O que acontece a longo prazo? Será que a matriz D_n converge? Ou seja, as probabilidades de que um determinado carro esteja alocado em cada loja estabilizam-se quando n cresce? Essa possível convergência depende da matriz de probabilidades inicial D_0? Como essa informação pode ser usada de forma prática pelo gerente da empresa?

$a)$ De acordo com o enunciado, a probabilidade de um carro ser devolvido na loja 2 no tempo $n = 1$, ou seja, $P(X_1 = 2)$ não pode ser obtida diretamente da matriz Q, pois essa fornece apenas as probabilidades de devolução, **dada** a loja de retirada. Ou seja, na matriz Q temos apenas $P(X_1 = 2|X_0 = k)$, com $k = 1, 2, 3$.

Em outras palavras, a sequência $(X_0, X_1, X_2, ...)$ (nosso primeiro exemplo de cadeia de Markov) tem como principal característica o fato de que não temos como determinar probabilidades sobre o tempo 1, sem antes saber o que ocorreu no tempo 0 ou, no nosso caso, em qual loja estava o carro.

Em geral, para qualquer tempo n, a probabilidade de um carro estar na loja 2 dependerá de termos à disposição alguma informação sobre o passado. Assim, com as informações dadas não podemos determinar $P(X_1 = 2)$. Mas se soubéssemos que o carro foi retirado na loja 3, teríamos como determinar $P(X_1 = 2|X_0 = 3) = q_{32} = 0,2$.

Uma possibilidade adicional ocorreria se tivéssemos a informação de que um cliente escolheu uma loja de retirada aleatoriamente com probabilidade $1/3$ para cada, e observássemos a partir daí a cadeia de Markov com o carro retirado por esse cliente. Nesse caso, podemos obter a probabilidade pedida com a ajuda do Teorema da Probabilidade Total (Teorema 5.2):

$$\begin{aligned}
P(X_1 = 2) &= P(\{X_0 = 1\} \cap \{X_1 = 2\}) + P(\{X_0 = 2\} \cap \{X_1 = 2\}) \\
&+ P(\{X_0 = 3\} \cap \{X_1 = 2\}) \\
&= P(X_0 = 1)P(X_1 = 2|X_0 = 1) + P(X_0 = 2)P(X_1 = 2|X_0 = 2) \\
&+ P(X_0 = 3)P(X_1 = 2|X_0 = 3) \\
&= \frac{1}{3}q_{12} + \frac{1}{3}q_{22} + \frac{1}{3}q_{32} = \frac{1}{3}(0,2 + 0,4 + 0,2) = \frac{4}{15}
\end{aligned}$$

Note que, apesar da formalização acima ao aplicar Teorema da Probabilidade Total, o cálculo pode ser descrito de forma intuitiva: retirar na loja 1 e devolver na loja 2, ou retirar na loja 2 e devolver na loja 2, ou retirar na loja 3 e devolver na loja 2. Sugerimos que o leitor faça um diagrama de árvore para representar a situação.

$b)$ Ao contrário do item anterior, a resposta da primeira pergunta pode ser obtida diretamente da matriz Q, pois temos a informação sobre o tempo 0 e desejamos obter a probabilidade no tempo 1. Ou seja, desejamos obter $P(X_1 = 2|X_0 = 1) = q_{12} = 0,2$.

Por outro lado, a resposta da segunda pergunta não pode ser obtida diretamente. Temos a informação $X_0 = 1$ e desejamos obter uma probabilidade relacionada à variável X_2 (tempo 2). Para isso, precisamos analisar todas as possibilidades de transição entre a primeira locação e a segunda locação, passando pelos tempos 0, 1 e 2. Essas possibilidades estão ilustradas no diagrama de árvore na Figura 7.1.

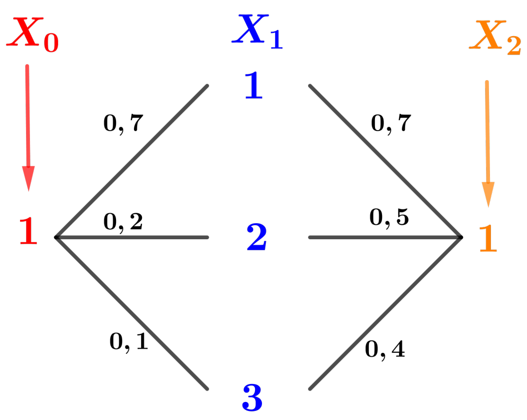

Figura 7.1: Diagrama de Árvore

Para justificar as probabilidades dadas na árvore, observe que, dada a informação $X_0 = 1$, temos que as probabilidades no tempo $n = 1$ (primeiro estágio da árvore) podem ser obtidas na primeira linha da matriz Q. Ou seja, são dadas pela matriz $D_1 = [0,7 \;\; 0,2 \;\; 0,1]$, pois essas são as probabilidades de o carro ser devolvido nas lojas 1, 2 e 3, dado que foi retirado na loja 1. O segundo estágio é obtido também na matriz Q, com as probabilidades de devolver o carro na loja 1, retirando em cada uma das outras três, ou seja, as entradas da coluna 1 de Q.

Por fim, observamos que calcular $P(X_2 = 1|X_0 = 1)$ é o mesmo que calcular $P(X_2 = 1)$. Essa afirmação segue do fato de que o evento $\{X_0 = 1\}$ tem probabilidade 1 e pode ser verificada a partir das propriedades da probabilidade[1]. Assim, usando novamente o Teorema da Probabilidade Total (ou diretamente pelo diagrama de árvore), podemos obter $P(X_2 = 1|X_0 = 1) = P(X_2 = 1)$:

$$\begin{aligned} P(X_2 = 1) &= P(X_1 = 1)P(X_2 = 1|X_1 = 1) + P(X_1 = 2)P(X_2 = 1|X_1 = 2) \\ &\quad + P(X_1 = 3)P(X_2 = 1|X_1 = 3) \\ &= 0,7 \cdot q_{11} + 0,2 \cdot q_{21} + 0,1 \cdot q_{31} \\ &= 0,7 \cdot 0,7 + 0,2 \cdot 0,5 + 0,1 \cdot 0,4 = 0,63. \end{aligned}$$

Note que a última linha da equação acima poderia ser obtida diretamente do diagrama de árvore, seguindo os três caminhos possíveis de locação e devolução. As primeiras linhas acima são a justificativa formal do cálculo utilizando o Teorema 5.2.

c) Vamos mostrar através desse exemplo que, como afirmado no início desta seção, a sequência $(X_n)_{n\geq 0}$ não é *i.i.d.*. Para constatar a dependência entre as variáveis, suponha novamente que $X_0 = 1$. Então, pelos resultados obtidos no item b), temos

$$P(X_2 = 1) = 0,63 \quad e \quad P(X_2 = 1|X_1 = 1) = q_{11} = 0,7$$

Ou seja,

$$P(X_2 = 1|X_1 = 1) \neq P(X_2 = 1)$$

Logo, X_2 é dependente de X_1.

Por fim, observamos que além de não serem independentes, as variáveis da sequência $X_0, X_1, ...$, também não são identicamente distribuídas, pois

$$P(X_0 = 1) = 1 \neq 0,63 = P(X_2 = 1)$$

Em outras palavras, as variáveis X_0 e X_2 assumem o valor 1 com probabilidades diferentes.

d) Dado que o carro foi retirado na loja 2, a segunda linha da matriz Q fornece-nos todas as probabilidades de devolução nas lojas 1, 2 e 3. Assim, a matriz pedida é $D_1 = [0,5 \;\; 0,4 \;\; 0,1]$. Observe que temos a seguinte relação:

$$D_1 = D_0 \cdot Q$$

[1] Se $P(B) = 1$, então $P(A|B) = P(A)$

Como veremos abaixo no cálculo da matriz D_2, essa relação não é uma coincidência.

Para obter a matriz D_2 vamos recorrer ao diagrama de árvore da Figura 7.2 para ilustrar todas as possibilidades de transição entre a primeira e a segunda locações. Lembramos mais uma vez que os cálculos feitos a partir do diagrama de árvore podem ser formalmente justificados pelo Teorema da Probabilidade Total.

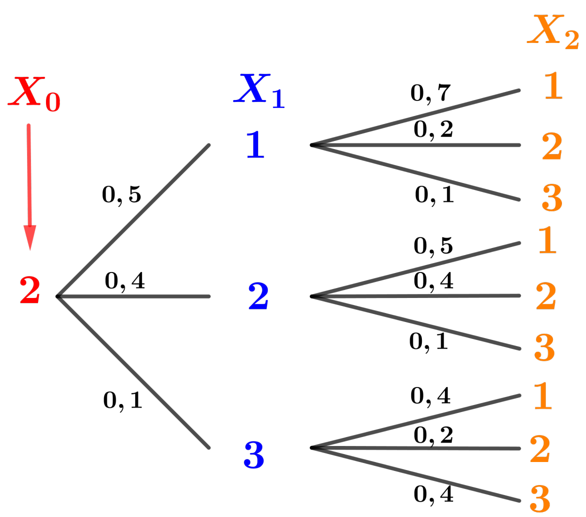

Figura 7.2: Diagrama de Árvore

Como a matriz D_2 é formada pelas probabilidades de devolução em cada loja no tempo 2, então devemos calcular $P(X_2 = k)$ para $k = 1, 2, 3$. Do diagrama de árvore, obtemos[2]:

$$\begin{aligned}
P(X_2 = 1) &= 0,5 \cdot 0,7 + 0,4 \cdot 0,5 + 0,1 \cdot 0,4 = 0,59 \\
P(X_2 = 2) &= 0,5 \cdot 0,2 + 0,4 \cdot 0,4 + 0,1 \cdot 0,2 = 0,28 \\
P(X_2 = 3) &= 0,5 \cdot 0,1 + 0,4 \cdot 0,1 + 0,1 \cdot 0,4 = 0,13
\end{aligned}$$
$$\Longrightarrow D_2 = \begin{bmatrix} 0,59 & 0,28 & 0,13 \end{bmatrix}$$

Observe que no cálculo de $P(X_2 = 1)$ temos os elementos da matriz linha $D_1 = [0,5 \;\; 0,4 \;\; 0,1]$ sendo multiplicados pelos elementos da primeira coluna da matriz Q, e, com isso, obtemos o primeiro elemento da matriz D_2.

$$\begin{bmatrix} 0,5 & 0,4 & 0,1 \end{bmatrix} \cdot \begin{bmatrix} 0,7 & 0,2 & 0,1 \\ 0,5 & 0,4 & 0,1 \\ 0,4 & 0,2 & 0,4 \end{bmatrix} = \begin{bmatrix} 0,59 & 0,28 & 0,13 \end{bmatrix}$$

[2]Note que as entradas de D_2 têm soma igual a 1.

Isso ocorre porque a cada produto dos elementos de D_1 pelos elementos da primeira coluna de Q, estão representados os caminhos: loja 1-loja 1, loja 2-loja 1 e loja 3-loja 1. Os elementos de D_1 representam as probabilidades de devolução na primeira locação, e a primeira coluna de Q contém as probabilidades de devolução na **loja 1** após a segunda locação. Ou seja, esse produto de linha por coluna contém todas as possibilidades de caminhos nos ramos da árvore entre as duas locações e que comecem na loja 2 e terminem na loja 1. Aplicando o mesmo argumento para as outras possíveis devoluções em D_2, obtemos:

$$D_2 = D_1 \cdot Q$$

De forma geral, usando argumentos semelhantes podemos deduzir que, para todo $n \geq 1$ temos:

$$D_n = D_{n-1}Q \Longrightarrow \begin{cases} D_1 & = & D_0 \cdot Q & & \\ D_2 & = & D_1 \cdot Q & = & D_0 \cdot Q^2 \\ \vdots & & & & \\ D_n & = & D_{n-1} \cdot Q & = & D_0 \cdot Q^n \end{cases}$$

Dessa forma, com a relação $D_n = D_0 \cdot Q^n$, dependemos apenas de D_0 para obter as probabilidades de devolução em cada loja após n locações. Isso reforça aquilo que foi dito no item a): as probabilidades no tempo n dependem termos informações sobre o início da cadeia, ou seja, o tempo 0.

e) Dadas as porcentagens do enunciado de carros em cada loja no instante inicial, um carro escolhido aleatoriamente tem $D_0 = [0,15 \quad 0,3 \quad 0,55]$ como matriz de probabilidades no tempo 0. Pela fórmula obtida no item anterior, podemos obter D_5 fazendo

$$D_5 = D_0 \cdot Q^5$$

O cálculo de Q^5 é bastante extenso para se fazer a mão, mas podemos recorrer a calculadores eletrônicas ou *softwares* matemáticos. Em particular, o GeoGebra permite fazer esse cálculo tanto pela Janela de Álgebra, quanto pela Janela CAS. Na Figura 7.3, temos as matrizes Q, D_0, Q^5 (nomeada de $Q5$) e $D_5 = D_0 \cdot Q^5$ obtidas na Janela de Álgebra do GeoGebra.

Para obter a matriz Q, basta digitar no campo de entrada o sequinte comando:

$$Q = \{\{0.7, 0.2, 0.1\}, \{0.5, 0.4, 0.1\}, \{0.4, 0.2, 0.4\}\}$$

Seguindo a mesma ideia, obtemos a matriz linha D_0, e com o comando $D_0 * Q^5$ obtemos a matriz D_5. Dessa matriz, deduzimos que a probabilidade de o carro sorteado estar na loja 1 após cinco locações é aproximadamente $60,61\%$.

f) Estamos interessados em obter o limite para matriz D_n quando n tende ao infinito. Não faz parte dos nossos objetivos, entretanto, formalizar e demonstrar a existência desse limite; tampouco entendemos que isso seja um impedimento para

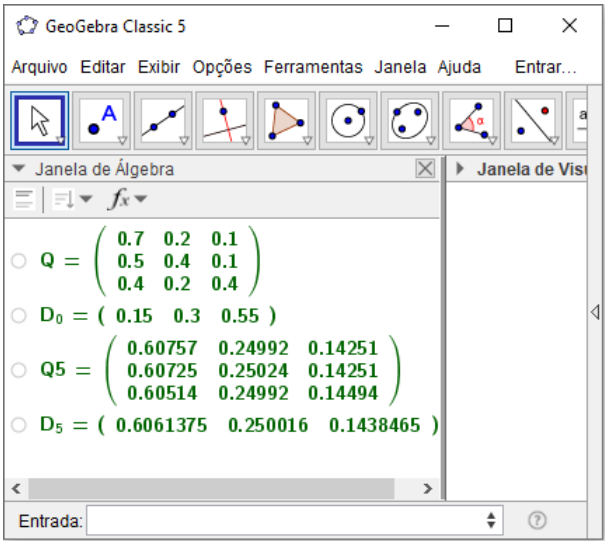

Figura 7.3: Cálculo matricial no GeoGebra

apresentar a ideia de convergência de forma intuitiva aos alunos do ensino médio. Postura semelhante é adotada quando trabalhamos com soma infinita dos termos de uma PG.

Informalmente, podemos analisar a convergência da matriz D_n usando novamente o GeoGebra. Com a ajuda de um controle deslizante, podemos analisar o comportamento de $D_n = D_0 \cdot Q^n$, quando n cresce.

Na Figura 7.4, temos o cálculo de D_n para n igual a 5 e 10, com aproximação de cinco casas decimais e utilizando como D_0 a matriz dada no item anterior. Manipulando o controle deslizante, é possível notar que a partir de $n = 10$, não há mais alteração nas cinco primeiras casas decimais. A partir de $n = 20$, todas as matrizes D_n são iguais até a décima casa decimal. Com 15 casas decimais, a estabilização acontece a partir de $n = 29$. O que indica uma convergência para uma matriz fixa E, com entradas próximas às da matriz D_n, quando n cresce.

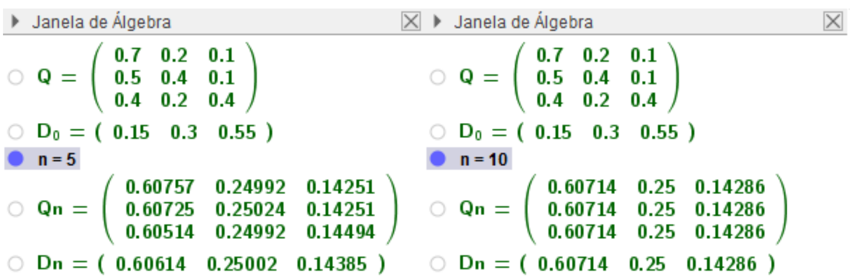

Figura 7.4: Convergência da matriz D_n

Se fizermos o mesmo cálculo alterando o estado inicial D_0, algo surpreendente ocorre. Sugerimos ao leitor substituir a matriz D_0 no cálculo do GeoGebra por qualquer outra[3]. Fazendo isso, será possível observar que D_n converge para a mesma matriz do cálculo anterior. Ou seja, a convergência não depende do estado inicial da cadeia, depende apenas da matriz Q.

Observe que, conforme mencionamos no item d), as probabilidades no tempo n, em geral, dependem do que ocorreu no tempo 0, ou seja, de D_0. Por outro lado, como vimos acima, conforme n cresce esse dependência do início da cadeia tende a desaperecer, ou seja, há uma perda de memória.

Por fim, observamos que a informação da convergência permite ao gerente da rede saber o que acontece a longo prazo (como vimos, não tão longo assim). As porcentagens de devolução entre as lojas vão se estabilizar, o que permite à empresa se planejar para a recepção e alocação dos carros em cada loja, de acordo com as porcentagens obtidas.

Comentários. O exemplo desenvolvido acima é bastante rico para o desenvolvimento em sala de aula de várias ideias relacionas às cadeias de Markov. Em particular, o processo que explica que a transição entre os tempos da cadeia está relacionada com um produto de matrizes utiliza uma das principais ferramentas do raciocínio probabilístico: o diagrama de árvores, que é sustentado pelo Teorema da Probabilidade Total.

As questões sobre dependência e probabilidade condicional também estão presentes em várias passagens do exemplo, bem como propriedades da probabilidade e o próprio significado de probabilidade em situações reais. Dessa forma, o trabalho introdutório com cadeias de Markov traz uma oportunidade rica para o desenvolvimento do raciocínio probabilístico por meio de um tópico moderno da área com aplicações em diversas áreas do conhecimento.

Além disso, o tópico ainda favorece a utilização de tecnologias e permite a integração e a contextualização entre áreas distintas da Matemática, com a utilização de conceitos como o cálculo matricial e a ideia de convergência. No caso das matrizes, é uma das poucas oportunidades de mostrar aos alunos na fase do ensino médio uma aplicação real desse conceito.

7.3 Definição e Propriedades

A partir dos conceitos desenvolvidos no Exemplo 7.1, algumas formalizações a respeito de cadeias de Markov tornam-se naturais. Vamos apresentar a seguir uma definição formal para cadeia de Markov e um importante teorema com os resultados deduzidos no referido exemplo.

[3] Qualquer outra matriz linha de probabilidades, ou seja, com entradas não negativas cuja soma seja igual a 1.

Definição 7.1. *Uma cadeia de Markov sobre o conjunto finito A com matriz de transição $Q = (q_{ij})$ é uma sequência de variáveis aleatórias $(X_n)_{n\geq 0}$ que assumem valores em A tais que $X_0 = a \in A$ e $P(X_n = j|X_{n-1} = i) = q_{ij}$ para todo $n \geq 1$, $i, j \in A$. Além disso, a distribuição de probabilidades de X_n é condicionalmente independente de X_m, se $m < n - 1$.*

Identificando a definição acima com o exemplo da seção anterior, o conjunto A de valores da cadeia naquele exemplo são os possíveis valores de X_n, ou seja, as lojas. Assim, temos $A = \{1, 2, 3\}$. O valor de $X_0 = a$ é a loja inicial onde está alocado o carro, que pode ser determinístico ou escolhido de forma aleatória.

A seguir, vamos enunciar um importante teorema sobre cadeias de Markov que estabelece propriedades fundamentais, algumas delas vistas no exemplo anterior. Não vamos demonstrar a parte do teorema relacionada à convergência. Tal demonstração pode ser encontrada em materiais mais avançados da área, entre os quais citamos [12].

Teorema 7.1. *Considere uma cadeia de Markov com matriz de transição Q e seja D_n a matriz linha de distribuição de probabilidades da variável X_n. Então:*

1) $D_n = D_{n-1}Q$ e $D_n = D_0 \cdot Q^n$ para todo $n \geq 1$.

2) Se Q^k tem todas as entradas positivas para algum $k \geq 1$, então

$$\lim_{n\to\infty} D_n = E \quad e \quad \lim_{n\to\infty} Q^n = S$$

onde E é a única matriz linha (de probabilidades) que satisfaz a equação $EQ = E$, chamada matriz de estado estacionário; e S é a matriz de mesma ordem de Q na qual todas as linhas são iguais a E.

Observações: 1) No exemplo da seção anterior fizemos um esboço da demonstração do item 1) do teorema. Sua justificativa depende apenas de conceitos básicos de matrizes e probabilidades, sem a necessiade de ferramentas avançadas.

2) A condição para a convergência no item 2) é a de que alguma potência da matriz Q tenha todas as entradas positivas. No caso do Exemplo 7.1, já tínhamos isso para Q^1. No exemplo que veremos mais à frente, precisaremos de uma potência maior para satisfazer essa condição. Uma matriz que satisfaz tal condição é chamada de matriz regular.

3) A conclusão do item 2) nos dá duas informações que não foram apresentadas até aqui. A primeira é um método para obter a matriz limite para a qual D_n converge: basta resolver o sistema linear $EQ = E$. Note que a equação $EQ = E$ tem apelo intuitivo, pois a matriz de estado estacionário é aquela em que mesmo após a transição de tempo (multiplicação por Q), as probabilidades permanecem iguais.

4) A segunda informação nova é que a potência Q^n converge para uma matriz cujas linhas são iguais a E. O leitor poderá verificar que isso justifica o fato de o limite de $D_n = D_0 \cdot Q^n$ não depender da distribuição inicial D_0, conforme vimos no item $f)$ do exemplo. Basta verificar o que acontece com o produto $D_0 \cdot S$, para qualquer D_0 arbitrária.

Exemplo 7.2. *Um conjunto de quatro* sites *está conectado por* links *que levam uns aos outros de acordo com as flechas indicativas no grafo da Figura 7.5. Suponha que, a cada 10 minutos, um usuário do sistema mude de* site *ou permaneça no mesmo com probabilidades iguais, de acordo com o número de possibilidades de mudança dadas pelo grafo.*

a) *Explicar por que a sequência de* sites *que um usuário visita pode ser modelada por uma cadeia de Markov, e determinar a sua matriz de transição a partir do grafo dado.*

b) *Supondo que o usuário entrou no sistema pelo* site *3, determinar, com a ajuda do GeoGebra, as matrizes de distribuição de probabilidades D_n, para $n = 1, 2, 3, 4$.*

c) *Determine, se existir, a matriz de estado estacionário com precisão de cinco casas decimais com o auxílio do GeoGebra. Caso existente, há outro modo de determinar tal matriz?*

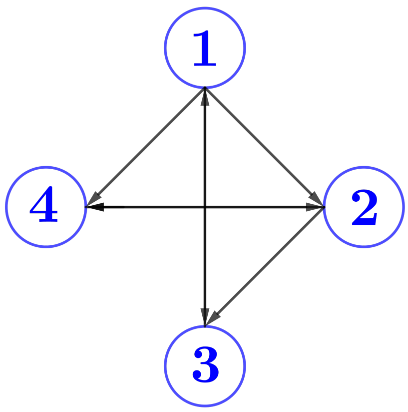

Figura 7.5: Grafo - Conexão entre *sites*

$a)$ Note que uma sequência de possíveis de *sites* visitados por um usuário é uma sequência $(X_n)_{n\geq 0}$ de V.A.s. Podemos ter, por exemplo, a sequência

$$(3, 2, 2, 4, 1, 3, 3, 2, 4, 4, 3, \ldots)$$

mas ela não é determinística. Tal sequência assume valores no conjunto $A = \{1, 2, 3, 4\}$.

De acordo com o grafo dado e com as informações do enunciado, as probabilidades de transição entre os *sites* a cada 10 minutos dependem do *site* anterior onde se encontra o usuário. Por exemplo, se o usuário está no *site* 4, ele pode ir para o *site* 2 ou permanecer no 4, com probabilidades iguais a $1/2$ para cada. Por outro lado, se ele estiver no *site* 1, ele pode ir para todos os outros ou permanecer no 1 com probabilidade $1/4$ para cada.

Assim, obtemos a matriz Q de transição do sistema de ordem 4×4 dada abaixo, onde q_{ij} é a probabilidade de mudar do *site* i para o *site* j.

$$Q = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

Note ainda que as probabilidades de transição não dependem do *site* que o usuário estava antes da última transição. Por exemplo,

$$P(X_6 = 2 | X_5 = 1, X_4 = 3) = P(X_6 = 2 | X_5 = 1) = \frac{1}{4}$$

Assim, todas as condições da Definição 7.1 estão satisfeitas, e a sequência $(X_n)_{n \geq 0}$ é uma cadeia de Markov.

$b)$ Se um usuário entrou no sistema pelo *site* 3, então temos $X_0 = 3$ e $D_0 = [0 \;\; 0 \;\; 1 \;\; 0]$. Assim, podemos determinar o valor de D_n para $n = 1, 2, 3, 4$ pela fórmula $D_n = D_0 \cdot Q^n$. Com a ajuda do GeoGebra, obtemos

$$D_1 = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix} \qquad D_2 = \begin{bmatrix} \frac{3}{8} & \frac{1}{8} & \frac{3}{8} & \frac{1}{8} \end{bmatrix}$$
$$D_3 = \begin{bmatrix} \frac{9}{32} & \frac{19}{96} & \frac{31}{96} & \frac{19}{96} \end{bmatrix} \qquad D_4 = \begin{bmatrix} \frac{89}{384} & \frac{271}{1152} & \frac{343}{1152} & \frac{271}{1152} \end{bmatrix}$$

Assim, por exemplo, um usuário que entrou no sistema pelo *site* 3, tem probabilidade de $31/96 \simeq 32,29\%$ de estar no mesmo *site* após 30 minutos.

$c)$ Em primeiro lugar, devemos avaliar a existência ou não da matriz de estado estacionário para a qual D_n converge. De acordo com o Teorema 7.1, devemos verificar se existe uma potência da matriz Q com entradas todas positivas.

De fato, usando o controle deslizante do GeoGebra para calcular Q^n para diversos valores de n, verificamos que Q^3 tem todas as entradas positivas. Logo, D_n converge para uma matriz de estado estacionário E.

Com o mesmo controle, podemos determinar $D_n = D_0 \cdot Q^n$ para vários valores de n. Ativando a opção de arredondamento para 5 casas decimais, verificamos que as entradas das matrizes D_n permanecem iguais (até a 5ª casa) a partir de $n = 19$.

Assim, uma aproximação da matriz de estado estacionário com precisão de cinco casas decimais é dada por

$$D_{19} = D_0 \cdot Q^{19} \simeq \begin{bmatrix} 0,18182 & 0,27273 & 0,27273 & 0,27273 \end{bmatrix}$$

Ainda de acordo com Teorema 7.1, podemos obter de forma precisa a matriz E resolvendo o sistema linear $EQ = E$. Assim, se $E = [x \;\; y \;\; z \;\; w]$, temos

$$\begin{bmatrix} x & y & z & w \end{bmatrix} \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} x & y & z & w \end{bmatrix}$$

$$\Longleftrightarrow \left\{ \begin{array}{rcl} \frac{x}{4} + \frac{z}{2} & = & x \\ \frac{x}{4} + \frac{y}{3} + \frac{w}{2} & = & y \\ \frac{x}{4} + \frac{y}{3} + \frac{z}{2} & = & z \\ \frac{x}{4} + \frac{y}{3} + \frac{w}{2} & = & w \end{array} \right. \Longleftrightarrow y = z = w = \frac{3x}{2}$$

Resolvendo o sistema possível e indeterminado acima, obtemos infinitas soluções em função de x, mas apenas uma delas satisfaz a condição necessária para a matriz E: a soma de suas entradas deve ser igual a 1, pois se trata de uma matriz de probabilidades. Logo,

$$x + 3 \cdot \frac{3x}{2} = 1 \Longleftrightarrow x = \frac{2}{11} \;\; e \;\; y = z = w = \frac{3}{11} \Longleftrightarrow E = \begin{bmatrix} \frac{2}{11} & \frac{3}{11} & \frac{3}{11} & \frac{3}{11} \end{bmatrix}$$

Note que a matriz E obtida acima é igual à matriz D_{19} até a quinta casa decimal (com arredondamento). O valor exato da matriz E pode sempre ser comparado com aproximações obtidas a partir de algum valor de n, com uma certa precisão de casas decimais.

Além disso, o resultado obtido para E pode ser interpretado como a proporção de tempo em que cada *site* do sistema é visitado, a longo prazo. Essa informação pode ser útil em termos de planejamento, propagandas, controle de *links* etc.

Por fim, observamos que é possível mostrar que, nas condições dadas no Teorema 7.1, o sistema $EQ = E$ sempre será possível e indeterminado, com conjunto solução de dimensão 1. Ao incluir a condição de que a soma das entradas de E tem que ser igual a 1, a solução passa a ser única.

Comentários. Como mostramos nos exemplos deste capítulo, o trabalho com cadeias de Markov, além de permitir uma possibilidade de inovação no tratamento de Probabilidade no ensino médio, tem a vantagem de integrar e contextualizar conceitos matemáticos importantes de diferentes áreas aparentemente desconexas para os alunos.

O raciocínio probabilístio é diretamente estimulado através da ideia de condicional, diagrama de árvore, probabilidade total, entre outros aspectos vistos acima. Por outro lado, a ideia intuitiva de convergência e conceitos de álgebra linear como cálculo matricial e sistemas lineares são aplicados nos problemas que resolvemos acima de forma acessível a alunos do ensino médio. Entre outras vantagens, essa integração pode ser usada para mostrar aos alunos uma justificativa para a multiplicação e potência de matrizes, cuja definição pode paracer em geral abstrata e sem utilizade.

7.4 Problemas Propostos

Em cada uma das situações apresentadas abaixo:

a) Identifique a sequência de variáveis aleatórias envolvida e justifique por que ela que define uma cadeia de Markov.

b) Determine a matriz de transição. Em seguida, escolha um estado para iniciar a cadeia e determine a distribuição de X_2 utilizando o diagrama de árvore.

c) Supondo X_0 escolhido aleatoriamente com igual probabilidade para todos os estados, determine a distribuição de X_1.

d) Com o auxílio do GeoGebra e com o D_0 do item anterior, determine os valores de D_n para $n = 1, 2, 3, 4$.

e) Identifique e justifique a existência da matriz de estado estacionário E. Se existir, determine no GeoGebra seu valor aproximado com cinco casas decimais e, depois, seu valor exato. Interprete o seu significado.

Problema 7.1. *Um país está dividido em três regiões demográficas. Constatou-se que todos os anos, entre os moradores da região 1, 5% muda-se para região 2 e 2% para a região 3. Entre os moradores da região 2, 12% mudam-se para a região 1 e 4% para a região 3. E entre os moradores da região 3, 8% vão para a região 1 e 1% para a região 2.*

Problema 7.2. *Na malha quadriculada* $\{(x, y) \in \mathbb{Z}^2, 0 \leq x, y \leq 2\}$ *há nove pontos. Numere-os de 1 a 9 e considere a cadeia na qual cada transição é feita para algum ponto vizinho na malha, com probabilidades iguais para todos os vizinhos.*

Problema 7.3. *No estudo de hereditariedade, sabe-se que os traços de um descendende é determinado pelo par de genes que recebe de seus pais. Os pares de genes (genótipos) possíveis são denotados por AA, Aa e aa, onde A e a, chamados de alelos, são as formas que cada gene pode assumir. Sabe-se ainda que se um dos pais tiver genótipo conhecido e o outro for aleatório, então é possível determinar as probabilidades de o descendente ter cada um dos genótipos. Na matriz abaixo, são dadas essas probabilidades associando às linhas 1, 2 e 3, respectivamente, os genótipos AA, Aa e aa do ascendente conhecido. As colunas estão associadas na* *mesma ordem aos genótipos do descendente.*

$$\begin{bmatrix} 0,5 & 0,5 & 0 \\ 0,25 & 0,5 & 0,25 \\ 0 & 0,5 & 0,5 \end{bmatrix}$$

Problema 7.4. *Em uma cidade americana* 35% *dos eleitores são adeptos do partido democrata,* 47% *são adeptos do partido republicano e o restante considera-se neutro em relação a ambos os partidos. Um cientista político verificou uma tendência anual de mudança de posição, com a matriz de probabilidades dada abaixo onde as linhas e colunas foram numeradas como: 1-Democratas, 2-Republicanos, 3-Neutros.*

$$\begin{bmatrix} 0,7 & 0,1 & 0,2 \\ 0,15 & 0,8 & 0,05 \\ 0,15 & 0,1 & 0,75 \end{bmatrix}$$

Problema 7.5. *Prove a afirmação feita na Observação 4 do Teorema 7.1: o limite de* $D_n = D_0 \cdot Q^n$ *não depende de* D_0.

Capítulo 8

Probabilidade na Geometria (ou o Contrário)

> *"... isso é natural no caso das probabilidades geométricas. O termo 'escolher' deve estar acompanhado de um procedimento, sem o qual essa palavra fica com significado vago."*
>
> – Eduardo Wagner

Neste capítulo, trataremos de um caso particular da definição mais geral de Probabilidade, vista na seção 3.3, para espaços amostrais infinitos e não enumeráveis. De forma geral, o cálculo de probabilidades nesses espaços ultrapassa os limites propostos para este material. Entretanto, no ensino básico, é possível explorar um de seus aspectos: a Probabilidade Geométrica. Muitos problemas de probabilidade geométrica podem ser tratados de forma intuitiva e em nível básico, sem a necessidade de aprofundamentos.

Entre os conjuntos não enumeráveis citados naquela seção, destacam-se aqueles que têm mais forte apelo intuitivo: as figuras geométricas elementares. O tópico conhecido como Probabilidade Geométrica propõe problemas de probabilidade envolvendo figuras geométricas contidas em espaços amostrais que são subconjuntos de $\mathbb{R}^n$, para algum $n \geq 1$. Dado o apelo intuitivo, esse tópico tem forte potencial para promover a contextualização e a integração entre duas áreas aparentemente desconexas da Matemática Básica: a Geometria e a Probabilidade.

Nesse contexto, os problemas contidos neste capítulo apresentam espaços amostrais infinitos não enumeráveis. Como já visto, a definição de probabilidade nesses espaços não pode ser feita no conjunto $\mathcal{P}(\Omega)$ das partes de Ω, mas em uma classe especial de subconjuntos de Ω conhecida como σ-álgebra.

Aqui, vamos nos limitar a dizer que a Definição 3.3 de Probabilidade pode ser adaptada para espaços amostrais infinitos e não enumeráveis geométricos. De forma mais direta, trabalharemos com subconjuntos reta $\mathbb{R}$, do plano $\mathbb{R}^2$ e do espaço $\mathbb{R}^3$ que sejam **mensuráveis**. Esse conceito, que tem uma definição precisa em teoria da medida, refere-se (no nosso contexto) a conjuntos em que é possível calcular seu comprimento, área ou volume nos casos da reta, plano ou espaço, respectivamente. Por exemplo, intervalos ou união de intervalos da reta, figuras geométricas elementares no plano, sólidos geométricos do espaço, entre outros.

8.1 Definição de Probabilidade Geométrica e suas Limitações

A definição clássica de probabilidades é baseada na ideia de atribuir à probabilidade de um evento ocorrer uma razão entre a cardinalidade entre dois conjuntos finitos: o dos casos favoráveis e o dos casos possíveis. Outra interpretação que se pode dar a essa ideia é a de que a probabilidade é a razão entre duas medidas de conjuntos. Essa é a ideia que sustenta a Probabilidade Geométrica. Esse tópico fundamenta-se em problemas em que o espaço amostral é uma figura geométrica do $\mathbb{R}^n$, $n \geq 1$.

A partir deste ponto, usaremos a palavra *medida* e sua respectiva notação μ para designar as medidas geométricas convencionais utilizadas nos espaços $\mathbb{R}^n$, para $n = 1, 2, 3$:

- Para um intervalo qualquer $[a, b] \subset \mathbb{R}$ da reta: $\mu([a, b]) = b - a$;
- Para uma região plana mensurável $A \subset \mathbb{R}^2$: $\mu(A) = \acute{a}rea(A)$;
- Para um sólido mensurável $S \subset \mathbb{R}^3$: $\mu(S) = volume(S)$.

Usando a propriedade aditiva, que afirma que $\mu(A \cup B) = \mu(A) + \mu(B)$ quando $A \cap B = \emptyset$, podemos medir também união de intervalos, de regiões planas e espaciais. E isso também pode ser estendido para uniões infinitas enumeráveis.

A fundamentação rigorosa da Probabilidade nesses espaços infinitos e não enumeráveis é feita em estudos avançados de probabilidade. Entretanto, restringindo os espaços amostrais e os eventos estudados àqueles em que podemos calcular sua medida, ou seja, seu comprimento em $\mathbb{R}$, sua área em $\mathbb{R}^2$ e seu volume em $\mathbb{R}^3$, podemos estabelecer a seguinte definição intuitiva:

Definição 8.1. *Seja $\Omega \subset \mathbb{R}^n$ um subconjunto com medida definida e finita $\mu(\Omega)$. Então, dado um conjunto mensurável $A \subset \Omega$, com medida $\mu(A)$, definimos a probabilidade de ocorrer A como*

$$P(A) = \frac{\mu(A)}{\mu(\Omega)}$$

O leitor deve pensar na razão da definição acima como a razão entre dois comprimentos, áreas ou volumes; com o objetivo de medir o tamanho do evento A em relação ao espaço amostral Ω. Assim, escolhendo um ponto aleatório de Ω a probabilidade de ele estar em A será dada por essa razão. Exatamente como fazemos na definição clássica de probabilidades.

Destacamos que é possível demonstrar que a definição acima satisfaz a Definição axiomática 3.3, se substituirmos $\mathcal{P}(\Omega)$ pela classe dos subconjuntos mensuráveis de $\mathbb{R}^n$. Assim, podemos dizer que a probabilidade geométrica é uma probabilidade no sentido axiomático e que satisfaz todas as propriedades demonstradas no Capítulo 4.

Por outro lado, observamos que essa definição conduz à seguinte interpretação: dois eventos $A, B \subset \Omega$ são equiprováveis se, e somente se, têm a mesma medida. Como consequência informal, a probabilidade de ocorrer um evento A não depende de sua localização em Ω. Como veremos no exemplo a seguir, isso pode trazer consequências importantes no uso dessa definição para a modelagem de problemas reais.

Exemplo 8.1. *Uma fonte com formato circular é famosa por realizar os desejos das pessoas que atiram moedas nela antes de fazer o pedido. Deseja-se obter as probabilidades de uma moeda atirada na fonte aleatoriamente cair em determinadas regiões. A probabilidade geométrica é uma boa ferramenta para modelar tais probabilidades? Se não, qual(is) seria(m) a(s) alternativa(s)?*

Representando por r o raio da fonte, uma possível modelagem para o problema é obtida tomando $\Omega = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 \leq r^2\}$ e utilizando a Definição 8.1, com μ sendo a medida de área, para determinar a probabilidade de uma moeda lançada cair em uma determinada região $A \subset \Omega$.

Assim, dados r, s tais que $0 < s < r$, se tomarmos

$$A = \{(x, y) \in \mathbb{R}^2; x^2+y^2 \leq s^2\} \quad \text{e} \quad B = \{(x, y) \in \mathbb{R}^2; r^2-s^2 \leq x^2+y^2 \leq r^2\},$$

teremos:

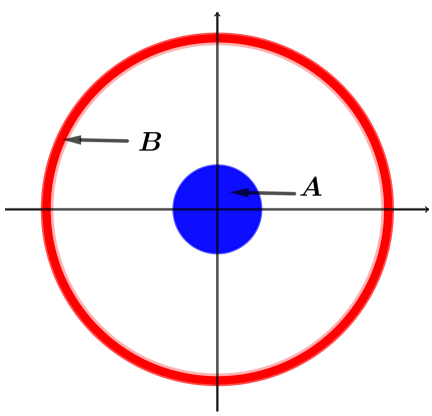

Figura 8.1: Problema da Fonte

$$P(A)=\frac{\mu(A)}{\mu(\Omega)}=\frac{\pi s^2}{\pi r^2}=\left(\frac{s}{r}\right)^2 \quad e$$
$$P(B)=\frac{\mu(B)}{\mu(\Omega)}=\frac{\pi r^2-\pi\left(r^2-s^2\right)}{\pi r^2}=\left(\frac{s}{r}\right)^2=P(A)$$

Por essa abordagem, somos levados a concluir que é igualmente provável uma moeda cair na região A ou na região B. Entretanto, como mostra a Figura 8.1, onde tomamos $r = 1$ e $s = 0,25$ para efeito de ilustração, tal conclusão não parece ser compatível com o problema real, visto que é bastante improvável que moedas atiradas aleatoriamente caiam na faixa estreita dada pela coroa circular B, ao contrário do círculo central A. Assim, a probabilidade geométrica não parece ser uma boa alternativa para modelar este problema.

Não faz parte de nossos objetivos tratar de ferramentas probabilísticas que mais se adequam a esse tipo de problema. Entretanto, para dar uma ideia de uma das abordagens possíveis, citamos a possibilidade da obtenção de uma função densidade de probabilidade sobre $\mathbb{R}^2$, que, ao contrário da definição geométrica, distribui as probabilidades entre as regiões de Ω de forma não uniforme.

Essa função pode ser obtida através da observação estatística do fenômeno, estabelecendo um modelo mais próximo da distribuição de moedas atiradas na fonte. Obtida a função, a probabilidade de uma determinada região A é dada pelo volume do sólido sob o gráfico da função e acima de A. Um exemplo de função que pode ser usada para esse fim é a distribuição normal bivariada, uma extensão da distribuição normal sobre a reta $\mathbb{R}$, conforme ilustra a Figura 8.2. Note que essa distribuição atribui maiores probabilidades aos pontos centrais da fonte.

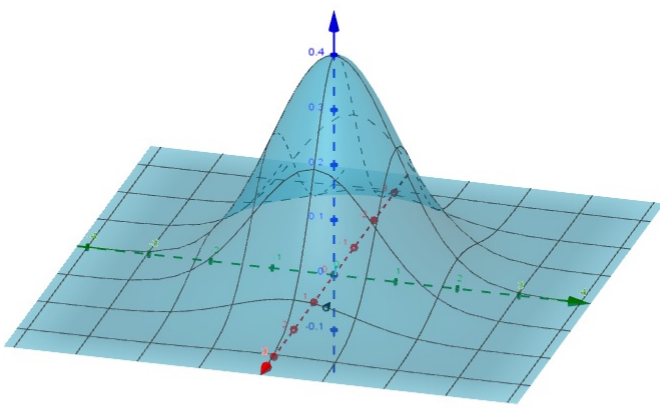

Figura 8.2: Distribuição Normal Bivariada

Comentários. A abordagem da Probabilidade Geométrica pode ser uma oportunidade para discussão em sala de aula sobre os limites da modelagem matemática perante a complexidade do mundo real, bem como as diferenças entre um problema abstrato de matemática pura e aplicação de ferramentas matemáticas com o intuito de solucionar um problema real.

O problema acima é rico para promover ambas as discussões. Qual abordagem mais adequada? A abordagem puramente geométrica é um exemplo de um problema matemático abstrato, importante para o desenvolvimento na área, mas se distancia do problema real. Há exemplos assim em outras áreas da Matemática? Essas e outras discussões que podem ser levantadas são importantes para a compreensão dos modelos estatísticos e probabilísticos, bem como da Matemática como um todo.

Adicionalmente, o problema acima pode ser aproveitado para introduzir as ideias de distribuição de frequências em espaços contínuos. Pode-se inclusive desenvolver um experimento físico que simule a situação apresentada, através do qual a coleta de dados a partir das simulações de lançamentos de moedas pode gerar gráficos de distribuição e uma ideia inicial de função densidade de probabilidade sobre $\mathbb{R}^2$.

8.2 Problemas Clássicos e Aspectos Didáticos

Nesta seção apresentaremos uma lista de problemas de Probabilidade Geométrica que tem interesse próprio, pelas situações curiosas que apresentam; mas que também trazem aspectos didáticos bastante interessantes para o trabalho em sala de aula, conforme explicitaremos nos comentários ao final de cada exemplo.

Exemplo 8.2 (Problema do Macarrão)**.** *Ao dividir aleatoriamente um segmento (fio de espaguete cru) em três partes, qual é a probabilidade de que os novos segmentos formem um triângulo?*

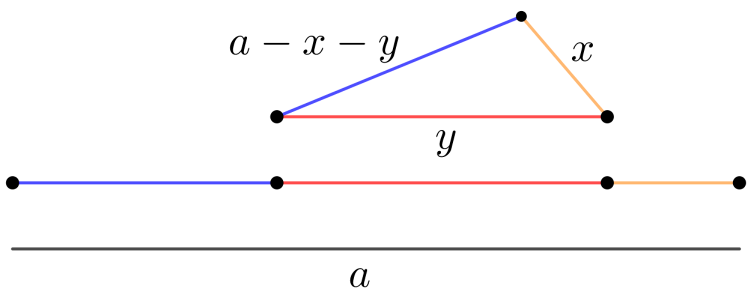

Figura 8.3: Problema do Macarrão

Consideramos que o segmento em questão tem comprimento $a > 0$ e que vamos dividi-lo em três partes. Assim, precisamos determinar duas medidas x e y tais que $0 < x < a$, $0 < y < a$ e $0 < x + y < a$, pois a terceira medida fica automaticamente determinada pelas duas primeiras. Assim, o espaço amostral do problema é dado pelos pares (x, y) satisfazendo as condições acima, ou seja,

$$\Omega = \left\{(x, y) \in \mathbb{R}^2; 0 < x < a; 0 < y < a; 0 < x + y < a\right\}$$

é dado pela região triangular do primeiro quadrante de $\mathbb{R}^2$ como mostra a Figura 8.4.

Por outro lado, nem todos os pares de Ω determinam um triângulo. Para isso, devem ser satisfeitas as condições de existência de um triângulo (desigualdade triangular). Ou seja, x e y devem satisfazer

$$\left\{\begin{array}{ccc} x & < & y + a - x - y \\ y & < & x + a - x - y \\ a - x - y & < & x + y \end{array}\right. \iff \left\{\begin{array}{ccc} x & < & \frac{a}{2} \\ y & < & \frac{a}{2} \\ x + y & > & \frac{a}{2} \end{array}\right.$$

Assim, a região de pontos de interesse é dada por

$$A = \left\{(x, y) \in \mathbb{R}^2; 0 < x < \frac{a}{2}; 0 < y < \frac{a}{2}; \frac{a}{2} < x + y < a\right\}$$

conforme ilustrado na Figura 8.4.

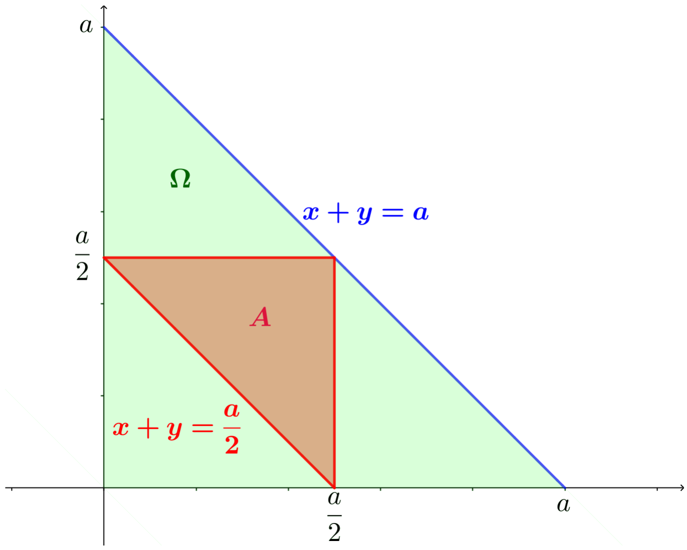

Figura 8.4: Problema do Macarrão

Portanto, a probabilidade buscada é dada por:

$$P(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{a^2/8}{a^2/2} = \frac{1}{4}$$

Comentários. Como o próprio nome dado ao problema sugere, há um experimento interessante que pode ser feito em sala de aula. Antes mesmo de apresentar o problema, pode-se distribuir aos alunos um fio de espaguete para cada, solicitando que cada um o parta em três pedaços de tamanho arbitrário. Em seguida, determina-se a porcentagem de alunos que conseguiram formar um triângulo com as três partes. Essa porcentagem, em tese, deveria ser próxima ao valor calculado para a probabilidade no problema. Será?

Pelo menos duas experiências podem ser citadas mostrando o contrário do esperado. No artigo [24] da *Revista do Professor de Matemática*, o autor conta que realizou a atividade com 60 professores, dos quais 41 ($\simeq 68\%$) conseguiram formar um triângulo com seus pedaços do espaguete. Em outra ocasião, em um minicurso que ministramos no III Simpósio da Formação do Professor de Matemática da Região Sul, em Chapecó-SC, havia 11 participantes da experiência, dos quais 9 ($\simeq 82\%$) conseguiram formar o triângulo. Ao apresentar o problema após o experimento, é claro que os resultados empíricos indicaram aos participantes que, a príncipio, a probabilidade pedida no enunciado deveria estar próxima à porcentagem obtida.

Tabela 8.1: Problema do Macarrão

Probabilidade Calculada	**Frequência Relativa Observada - Exp. 1**	**Frequência Relativa Observada - Exp. 2**
25%	$\simeq 68\%$	$\simeq 82\%$

Ao ser revelada a probabilidade teórica de 25% que calculamos acima, é natural que haja uma surpresa entre os participantes do experimento. Isso traz uma excelente oportunidade para questionar o porquê de o resultado empírico não estar sequer próximo do teórico. Um argumento possível é o de que a amostra não é grande o suficiente para aproximar a frequência relativa da probabilidade teórica; isso pode ser testado repetindo o experimento com um número maior de pessoas.

Mas há outra explicação possível e mais provável. Uma parte relevante dos casos considerados no conjunto Ω, nos quais x e/ou y são pequenos, dificilmente aparecerão na prática de partir um espaguete em três partes. Se, por exemplo, o espaguete tiver 30 cm, é difícil imaginar que as pessoas que não conhecem o experimento obtenham pedaços de 2 cm, 3 cm e 25 cm. Na prática, a partição fica naturalmente limitada a pedaços mais ou menos próximos a $1/3$ do total.

Ou seja, distribuir probabilidades uniformente ao longo do fio de espaguete, como faz a definição geométrica de probabilidade, pode não condizer com o experimento real de partir espaguetes. Daí fica a questão: é possível desenvolver um experimento aleatório cujos resultados aproximem-se da modelagem geométrica? E mais: há alguma semelhança com o exemplo anterior?

Por fim, note o problema integra conceitos de geometria plana, geometria analítica, sistemas de inequações lineares, probabilidade e, ao fazer o experimento com o macarrão, estatística. Uma oportunidade rica para o desenvolvimento integrado de conceitos para resolução de problemas.

Exemplo 8.3. *Considere um triângulo equilátero ABC cujo lado mede ℓ. Ao selecionar aleatoriamente um ponto P interno ao triângulo, qual é a probabilidade de que $A\widehat{P}B$ seja agudo?*

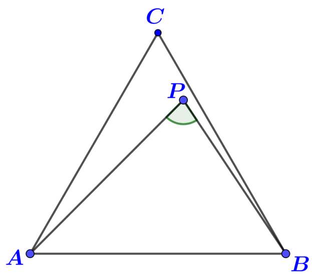

Figura 8.5: Região Aguda

Sabemos que se P pertence à circunferência de diâmetro AB, então $A\widehat{P}B$ é reto. Consequentemente, $A\widehat{P}B$ é obtuso se, e somente se, P pertence à região dada pela interseção do triângulo ABC com o círculo de diâmetro AB. Ou seja, enquanto o espaço amostral Ω é dado pela região interna ao triângulo ABC, nossa região de interesse R é o complementar daquela representada em vermelho na Figura 8.6.

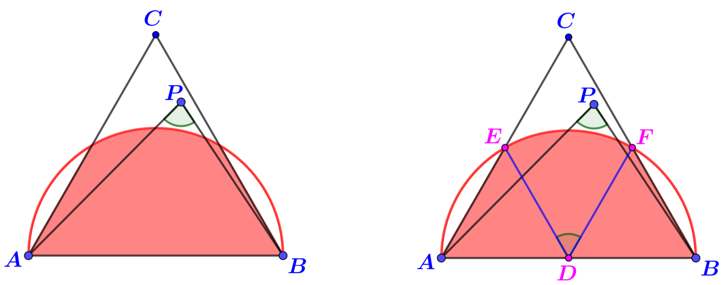

Figura 8.6: Região Aguda

Seja D o ponto médio de AB. No lado direito da Figura 8.6, note que os triângulos ADE e BDF são equiláteros, pois $AD = DE$ e $D\widehat{E}A = D\widehat{A}E = 60^o$, o que ocorre de forma análoga para o triângulo BDF. Da informação anterior, concluímos que o ângulo do setor circular $\sphericalangle EDF$ mede 60^o. Portanto,

$$\begin{aligned}\mu\left(R^C\right) &= \mu(ADE) + \mu(BDF) + \mu(\sphericalangle EDF) = 2 \cdot \frac{(\ell/2)^2\sqrt{3}}{4} + \frac{1}{6}\pi\left(\frac{\ell}{2}\right)^2 \\ &= \frac{\ell^2\sqrt{3}}{8} + \frac{\pi\ell^2}{24} = \frac{\ell^2(3\sqrt{3}+\pi)}{24}\end{aligned}$$

Logo,

$$\begin{aligned}P(R) = 1 - P\left(R^C\right) &= 1 - \frac{\mu\left(R^C\right)}{\mu(ABC)} = 1 - \frac{\frac{\ell^2(3\sqrt{3}+\pi)}{24}}{\frac{\ell^2\sqrt{3}}{4}} \\ &= 1 - \frac{3\sqrt{3}+\pi}{6\sqrt{3}} = \frac{3\sqrt{3}-\pi}{6\sqrt{3}} \simeq 19,77\%\end{aligned}$$

Comentários. No Capítulo 2, mostramos que em materiais didáticos tradicionais há uma abundância de problemas de probabilidades cujo foco é o raciocínio combinatório. Nesse tipo de problema, o conhecimento probabilístico exigido é apenas a identificação do espaço amostral e do evento de interesse e a definição clássica de probabilidade através da razão entre as cardinalidades dos conjuntos. Mas, em geral, o foco desse tipo de problema está no uso de técnicas de contagem da Análise Combinatória.

Observe que o problema que resolvemos acima tem, em essência, essa mesma característica com Geometria no lugar de Análise Combinatória. Todo o foco do problema está no raciocínio geométrico, ou seja, no conhecimento de propriedades da geometria euclidiana plana para obtenção das áreas dos conjuntos de interesse. O racicínio probabilístico é requisitado apenas no final, no uso da propriedade do complementar e no cálculo de probabilidade (geométrica) como razão. Entretanto, além de ser um problema interessante do ponto de vista geométrico, a questão aleatória envolvida parece ser bem curiosa e instigante para os alunos.

Nos comentários dos problemas anteriores, colocamos diversas questões sobre a natureza aleatória do experimento e sobre seu significado probabilístico. Uma forma de trazer para esse problema os mesmos elementos, seria propor experimentos que possam simular a escolha aleatória. Sejam experimentos físicos, como atirar pequenos objetos em um cercado triangular; sejam experimentos computacionais, onde é possível simular escolhas aleatórias de pontos no triângulo e calcular a frequência daqueles que formam um ângulo agudo com a base AB.

Exemplo 8.4 (O sumiço dos Racionais)**.** *Ao selecionar aleatoriamente um ponto de um intervalo fechado $[a, b]$ qualquer, com $a \neq b$, qual é a probabilidade de que sua abscissa seja racional?*

No espaço amostral $\Omega = [a, b]$, podemos utilizar a definição geométrica de probabilidade com μ sendo o comprimento de um intervalo (ou a soma de comprimentos). Ou seja, para um intervalo qualquer $[c, d] \subset [a, b]$ temos $\mu([c, d]) = d-c$.

Observe ainda que um conjunto unitário $\{x\}$ pode ser visto como um intervalo degenerado $[x, x]$ e, portanto, todo conjunto unitário tem medida nula, já que $\mu(\{x\}) = \mu([x, x]) = x - x = 0$.

Por outro lado, sabemos que o conjunto dos números racionais é enumerável e, portanto, o evento de interesse $A = \mathbb{Q} \cap [a, b]$ é enumerável. Assim, o conjunto dos números racionais do intervalo $[a, b]$ pode ser escrito como

$$A = \{r_1, r_2, r_3, ...\} = \bigcup_{i=1}^{\infty} \{r_i\}$$

Logo, segue do axioma da união da Definição 3.3 que

$$P(A) = P\left(\bigcup_{i=1}^{\infty} \{r_i\}\right) = \sum_{i=1}^{\infty} P(\{r_i\}) = \sum_{i=1}^{\infty} \frac{\mu(\{r_i\})}{\mu([a, b])} = \sum_{i=1}^{\infty} \frac{0}{b - a} = 0$$

Comentários. A resolução desse exemplo envolve conceitos mais avançados de Matemática, o que o torna inapropriado para o desenvolvimento em uma turma regular do ensino básico. Entretanto, há pelo menos dois aspectos interessantes do problema para professsores de Matemática.

Primeiro, que seu resultado contraintuitivo é consequência direta da definição mais geral de probabilidade, aquela para espaços amostrais infinitos não enumeráveis. O que nos traz de volta às questões elaboradas nos exemplos anteriores: que tipo de problema pode ser modelado em espaços desse tipo? É um problema apenas de matemática pura? É possível realizar o experimento de escolher um ponto aleatório de um intervalo? Um gerador aleatório comum de números, por exemplo, só gera números racionais. Como o experimento proposto poderia ser realizado na prática?

O segundo aspecto interessante do problema é que ele ilustra um conhecido resultado da teoria da medida: ao utilizar a medida geométrica em subconjuntos de $\mathbb{R}^n$, todo conjunto enumerável tem medida nula. No caso da reta real, esse resultado reforça a ideia contraintuitiva de que o conjunto dos números racionais, apesar de ser denso na reta (todo intervalo real contém infinitos racionais), tem medida nula. O que vai ao encontro do conhecido resultado de Análise que mostra que a cardinalidade dos números irracionais é maior que a dos racionais.

Exemplo 8.5 (Paradoxo de Bertrand)**.** *Escolhendo-se aleatoriamente uma corda de um círculo, qual é a probabilidade de que seu comprimento seja maior que o comprimento do lado de um triângulo equilátero inscrito?*

Solução 1. A primeira questão que devemos fazer é: de que forma podemos escolher uma corda aleatória em um círculo? Como comentaremos mais adiante, é essa escolha que faz do problema um aparente paradoxo.

Uma possível forma de escolher uma corda qualquer de um círculo é feita em duas etapas: fixamos uma direção para a corda e, em seguida, escolhemos uma distância da corda para o centro do círculo, como ilustra a Figura 8.7.

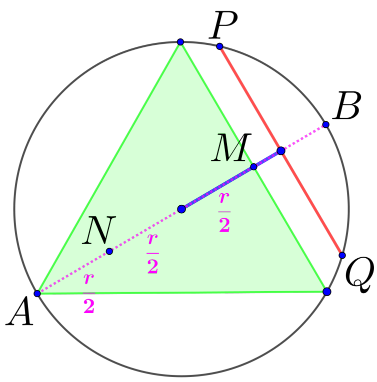

Figura 8.7: Paradoxo de Bertrand - Solução 1

Na figura acima, a direção da corda PQ foi fixada pelo diâmetro AB, perpendicular a ela. O baricentro do triângulo equilátero inscrito coincide com o centro do círculo, o que justifica as distâncias $r/2$ na figura. Por fim, note que o comprimento da corda é maior que o do lado do triângulo se, e somente se, a distância entre a corda e o centro do círculo for menor que $r/2$. Ou seja, podemos tomar o espaço amostral como sendo $\Omega = AB$ e o evento de interesse $E = MN$. Logo,

$$P(E) = \frac{\mu(MN)}{\mu(AB)} = \frac{r}{2r} = \frac{1}{2}$$

Solução 2. A corda é selecionada em duas etapas: fixamos um ponto A sobre a circunferência e, em seguida, escolhemos aleatoriamente o segundo ponto P, como ilustra a Figura 8.8.

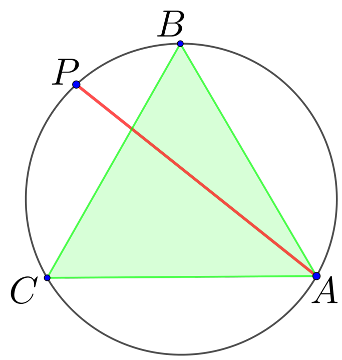

Figura 8.8: Paradoxo de Bertrand - Solução 2

Nesse caso, a corda PA será maior que o lado AB se, e somente se, o ponto P pertencer ao arco menor $\widehat{BC}$, que tem comprimento igual a $1/3$ da circunferência. Assim, nomeando a circunferência por $\mathcal{C}$, temos $\Omega = \mathcal{C}$ e o evento de interesse $E = \widehat{BC}$. Logo,

$$P(E) = \frac{\mu(\widehat{BC})}{\mu(\mathcal{C})} = \frac{2\pi r/3}{2\pi r} = \frac{1}{3}$$

Solução 3. Escolhe-se um ponto P qualquer interno ao círculo e tomamos a (única) corda AB cujo ponto médio é P, conforme ilustra a Figura 8.9.

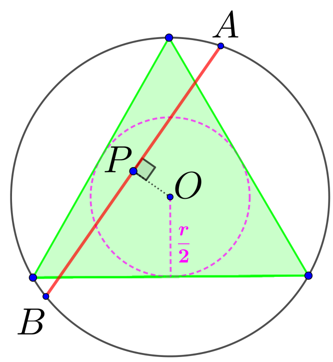

Figura 8.9: Paradoxo de Bertrand - Solução 3

Sejam $\mathcal{C}_1$ e $\mathcal{C}_2$ os círculos de cento O e raios r e $r/2$ respectivamente. Observe que a corda AB tem comprimento maior que o lado do triângulo equilátero inscrito se, e somente se, $P \in \mathcal{C}_2$. Assim temos $\Omega = \mathcal{C}_1$ e o evento de interesse $E = \mathcal{C}_2$. Logo,

$$P(E) = \frac{\mu(\mathcal{C}_2)}{\mu(\mathcal{C}_1)} = \frac{\pi(r/2)^2}{\pi r^2} = \frac{1}{4}$$

Comentários. Apesar do nome pelo qual o problema é conhecido, não se trata de um paradoxo de fato. Tal terminologia é utilizada para se referir ao fato de que o problema possui (pelo menos) três resultados diferentes. Entretanto, como mostram as soluções acima, a cada forma nova de se escolher uma corda aleatória tem-se um problema de probabilidade diferente, com espaço amostral e evento de interesse únicos. A mudança na forma de execução do experimento aleatório de escolha de uma corda qualquer faz com que o problema de probabilidade seja alterado também. Isso ocorre porque a frase “escolher uma corda aleatória” não explicita um procedimento bem definido pelo qual isso será feito.

A situação apresentada traz a oportunidade para discussão em sala de aula sobre a importância da descrição e compreensão detalhada e bem definida do experimento aleatório que dá base ao problema. O que significa escolher uma corda aleatória? De quais formas isso pode ser feito? Há outras maneiras além das apresentadas no problema? Trata-se de um problema geométrico abstrato ou pode-se desenvolver um experimento físico para reproduzi-lo?

No desenvolvimento desse problema em sala de aula, pode-se explorar entre os alunos formas de escolher uma corda qualquer, antes de se apresentar as três possibilidades aqui colocadas. Novas probabilidades podem surgir dessa exploração. Pode-se pensar ainda em outros problemas de probabilidade geométrica nessa mesma linha, com mais de uma escolha possível. Por exemplo, escolhendo um triângulo aleatório inscrito em um círculo, determinar a probabilidade de ele ser escaleno.

Exemplo 8.6 (Agulha de Buffon)**.** *Considere uma família de retas paralelas no plano cuja distância entre duas retas consecutivas quaisquer é d. Ao escolher um segmento de reta aleatório (ou lançar uma agulha) de comprimento $\ell \leq d$, qual é a probabilidade de que esse segmento intercepte uma das retas?*

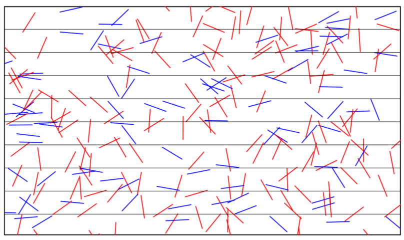

Figura 8.10: Agulha de Buffon - Fonte: [2]

Como as regiões entre as retas paralelas são idênticas, podemos analisar as posições possíveis para uma agulha lançada fixando a região entre as duas retas onde está localizado o ponto médio da agulha.

Dessa forma, uma agulha qualquer lançada no plano pode ter sua posição determinada por duas escolhas arbitrárias: a escolha da distância $y = MO$ (Figura 8.11) entre o ponto médio da agulha e a reta r mais próxima desse ponto; e a escolha do ângulo $\theta = M\widehat{N}P$ entre a reta suporte da agulha e a reta r, onde P está à esquerda de O e N. O caso em que agulha é paralela à r ocorre quando $\theta = 0$ ou $\theta = \pi$.

Note que as escolhas de y e θ devem satisfazer $0 \leq y \leq d/2$ e $0 \leq \theta \leq \pi$. Ou seja, o espaço amostral do problema é dado por

$$\Omega = \Big\{(\theta, y) \in \mathbb{R}^2; 0 \leq \theta \leq \pi, 0 \leq y \leq d/2\Big\}$$

Por outro lado, temos que a agulha intercepta a reta r se, e somente se

$$MN \leq \frac{\ell}{2} \iff \frac{y}{sen(\theta)} \leq \frac{\ell}{2} \iff y \leq \frac{\ell sen(\theta)}{2}$$

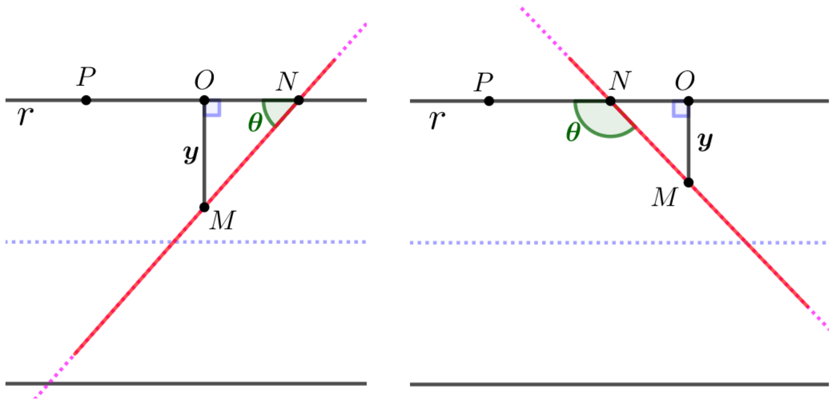

Figura 8.11: Agulha de Buffon

Assim, nosso evento de interesse é dado pelo conjunto

$$A = \left\{ (\theta, y) \in \Omega; y \leq \frac{\ell sen(\theta)}{2} \right\}$$

O conjunto Ω forma um retângulo no plano cartesiano de área igual a $\dfrac{\pi d}{2}$ e o evento A é a região sob o gráfico da função $y = \dfrac{\ell sen(\theta)}{2}$, para $0 \leq \theta \leq \pi$, e acima do eixo das abscissas, conforme ilustrado na Figura 8.12.

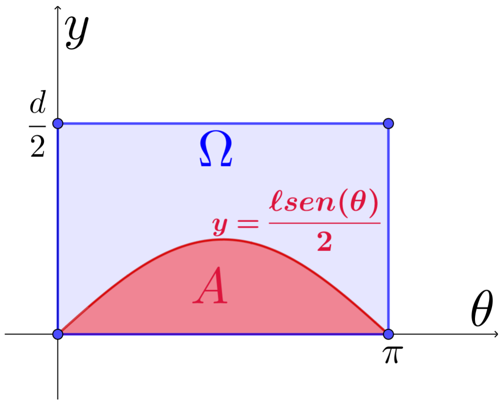

Figura 8.12: Agulha de Buffon

Assim, a área de A é dada por

$$\mu(A) = \int_0^\pi \frac{\ell sen(\theta)}{2} d\theta = -\frac{\ell cos(\theta)}{2}\bigg|_{\theta=0}^{\theta=\pi} = \ell$$

Logo,

$$P(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{2\ell}{\pi d} \tag{8.1}$$

Como os comprimentos ℓ e d são fixos, a fórmula obtida acima nos permite obter aproximações para π através de um experimento aleatório. Usando uma região no chão com retas paralelas (por exemplo, uma sala cujo chão é feito de ripas retangulares de madeira) e um conjunto com n agulhas satisfazendo as condições do problema, pode-se lançar as agulhas e contar o número n_A de agulhas que interceptam as retas. A razão $\frac{n_A}{n}$, se n for grande o suficiente, é uma boa aproximação para $P(A)$ e obtemos:

$$\frac{n_A}{n} \simeq P(A) = \frac{2\ell}{\pi d} \Longleftrightarrow \pi \simeq \frac{2n\ell}{dn_A}$$

De fato, diversos experimentos para obtenção de aproximações para π utilizando o problema da agulha de Buffon foram registrados ao longo da história, como mostra a Tabela 8.2, extraída da referência [23].

Tabela 8.2: Obtenção de π com a Agulha de Buffon

Autor (ano)	Razão $\frac{\ell}{d}$	Nº de lançamentos* (n)	Número de sucessos (n_A)	Aproximação $\pi \simeq \frac{2n\ell}{dn_A}$
Wolf (1850)	0,8	5000	2532	3,1596
Smith (1855)	0,6	3204	1218,5	3,1553
De Morgan (1860)	1,0	600	382,5	3,137
Fox (1864)	0,75	1030	489	3,1595
Lazzerini (1901)	0,83	3408	1808	3,1415929
Reina (1925)	0,5419	2520	859	3,1795
(*): Smith e De Morgan consideravam meio sucesso quando não ficava bem definida a interseção da agulha com a reta.				

Comentários. O problema da agulha de Buffon tem vários aspectos que consideramos interessantes de serem explorados no ensino básico. Apesar da necessidade da utilização do conceito de integral em uma das passagens da resolução, entendemos que isso não deve ser um impedimento. Quase todo raciocínio fundamental para a resolução do problema depende apenas de conceitos de matemática básica, como o posicionamento da agulha ao ser lançada, o intervalo de variação das duas variáveis, a condição para que a agulha intercepte uma das retas, a representação do espaço amostral e do evento no plano cartesiano e o cálculo da área da região de Ω.

Para o cálculo da região A, pode-se fixar valores para ℓ e d satisfazendo as condições do problema e utilizar o GeoGebra para obter um desenho como o apresentado na Figura 8.12. Em seguida, usar a ferramenta do GeoGebra (integral) que permite calcular a área sob uma curva dada por uma função. É importante observar que, apesar de não ser um impedimento, devemos esclarecer aos alunos que existem ferramentas matemáticas de nível superior que permitem obter a área em questão. Isso pode ser usado inclusive como justificativa para apresentar a fórmula (8.1) como generalização para o problema. Mais ainda, colocando o valor de ℓ como controle deslizante, obtém-se uma justificativa através do GeoGebra para $\mu(A) = \ell$, ainda que não seja uma demonstração formal.

De posse da fórmula e de todas as informações sobre o problema, a obtenção de aproximações para o π através de um experimento aleatório passa a ser outro atrativo com ricas discussões para o problema, incluindo os resultados exibidos na Tabela 8.2. Como planejar este experimento? Quantas agulhas devemos lançar para obter boas aproximações? Por que a razão ℓ/d é importante? Por que devemos ter $\ell \leq d$? É possível realizar esse experimento computacionalmente, além de fisicamente? As possibilidades de exploração didática do problema são muitas.

8.3 Problemas Propostos

Problema 8.1. *Ao selecionar aleatoriamente um ponto no círculo unitário $\mathcal{C}$ com centro na origem, qual é a probabilidade de que ele esteja: a) no primeiro quadrante? b) no quadrado com centro na origem e lado $\dfrac{\sqrt{\pi}}{2}$? c) no conjunto $\left\{(x,y) \in \mathcal{C}; |y| < \dfrac{|x|}{2}\right\}$. É possível imaginar um experimento físico modelado por esse problema?*

Problema 8.2. *Um computador seleciona aleatoriamente pontos do cubo com centro na origem do espaço tridimensional $\mathbb{R}^3$ e aresta de medida 2. Qual é a probabilidade de que um ponto selecionado esteja na esfera unitária inscrita no cubo? É possível reproduzir esse experimento com a ajuda do GeoGebra? Há limitações tecnológicas?*

Problema 8.3. *Considere uma família de retas paralelas no plano cuja distância entre duas retas consecutivas quaisquer é d. Ao lançar uma moeda de diâmetro*

$2r < d$ sobre o plano, qual é a probabilidade de que a moeda não intercepte nenhuma das retas?

Problema 8.4 (Problema do Encontro)**.** *Duas pessoas marcaram um encontro em um local entre 11 e 12 horas e combinaram que nenhuma delas esperará pelo outra mais do que 15 minutos. Supondo que ambas escolherão o horário de chegada aleatoriamente e de forma uniforme no intervalo acima, determinar a probabilidade de que o encontro ocorra.*

Problema 8.5. *Escolhe-se aleatoriamente e de forma uniforme três números reais x, y e z no intervalo $[0, a]$, com $a > 0$. Determinar a probabilidade de que $z^2 < x^2 + y^2$.*

Problema 8.6. *Um ponto P é escolhido uniformemente no triângulo ABC. Sejam K, L e M os pontos médios das três cevianas que passam por P. Calcule a probabilidade de que P seja interno ao triângulo KLM.*

Referências Bibliográficas

[1] ANTON, H.; RORRER, C. *Álgebra Linear com Aplicações*, 10ª ed. Porto Alegre: Bookman, 2012.

[2] BORTOLOSSI, H. Números (Pseudo) Aleatórios, Probabilidade Geométrica, Métodos de Monte Carlo e Estereologia. Projeto Klein de Matemática em Língua Portuguesa, 2012.

[3] BRASIL. Ministério da Educação. Parâmetros Curriculares Nacionais - Ensino Médio. Brasília. Ministério da Educação e Cultura, 1999.

[4] BRASIL. Base Nacional Comum Curricular. Brasília: MEC, 2018.

[5] DANTAS, C. A. B. *Probabilidade: um curso introdutório*, 3ª ed. São Paulo: Edusp, 2008.

[6] DANTE, L. R. *Matemática: Contexto e aplicações* - Vol. 2, 2ª ed. São Paulo: Ática, 2014.

[7] HAZZAN, S. *Fundamentos da Matemática Elementar, 5: combinatória, probabilidade*, 7ª ed. São Paulo: Atual, 2004.

[8] IEZZI, G. *et al. Matemática: Ciência e aplicações* - Vol. 2, 7ª ed. São Paulo: Saraiva, 2013.

[9] JAMES, B. R. *Probabilidade: um curso em nível intermediário*, 3ª ed. Rio de Janeiro: IMPA, 2011.

[10] LEON, S. J. *Álgebra Linear com aplicações*, 8ª ed. Rio de Janeiro: LTC, 2014.

[11] LEONARDO, F. M *et al. Conexões com a Matemática 2*, 2ª ed. SP: Moderna, 2013.

[12] LEVIN, D.A.; PERES, Y. WILMER, E.L. *Markov Chains and Mixing Times.* Providence, R.I.: American Mathematical Society, 2009.

[13] LIMA, C., MAGALHÃES, M. *Noções de Probabilidade e Estatística*, 7ª ed. São Paulo: Edusp, 2015.

[14] LOUZADA, F. *et al. Reflexões a respeito dos conteúdos de probabilidade e estatística na escola no Brasil: uma proposta.* Associação Brasileira de Estatística, 2015.

[15] MAES, J. Viés de sobrevivência: a falha de raciocínio que afeta a sociedade, 2019. Disponível em: <https://hypescience.com/vies-de-sobrevivencia/>. Acesso em: 11 de outubro de 2019.

[16] MAGALHÃES, M. *Probabilidade e Variáveis Aleatórias*, 3ª ed. São Paulo: Edusp, 2015.

[17] MEYER, P.L. *Probabilidade - Aplicações à Estatística*, 2ª ed. Rio de Janeiro: LTC, 2003.

[18] MLODINOW, L. *O Andar do Bêbado: como o acaso determina nossas vidas.* Rio de Janeiro: Zahar, 2009.

[19] MORGADO, A. *et al. Análise Combinatória e Probabilidade*, 9ª ed. Rio de Janeiro: SBM, 2006.

[20] MORGADO, A.; TEIXEIRA, R.C. *Introdução à Probabilidade.* Rio de Janeiro, 2011.

[21] ROSS, S. *Probabilidade: um curso moderno com aplicações*, 8ª ed. Porto Alegre: Bookman, 2010.

[22] ROSS, S. *Introduction to Probability Models*, 9ª ed. San Diego: Elsevier, 2007.

[23] TUNALA, N. Determinação de probabilidades por métodos geométricos. *Revista do Professor de Matemática*, SBM, vol. 20, Páginas 18-22, 1992.

[24] WAGNER, E. Probabilidade Geométrica: o problema do macarrão e um paradoxo famoso. *Revista do Professor de Matemática*, SBM, vol. 34, Páginas 28-35, 1997.

REALIZAÇÃO

ANPMat

Associação Nacional dos Professores de Matemática na Educação Básica

ORGANIZAÇÃO

SBM

SOCIEDADE BRASILEIRA DE MATEMÁTICA

ISBN 978-65-88013-10-6

9 786588 013106 >