O QUE É PROBABILIDADE?*

M.A. GAMA[†], A.L.C. SANTOS[‡], P.N. SILVA[§]

Resumo

Em geral, o ensino de probabilidade só se restringe a uma apresentação rápida de alguns conceitos e concentra-se em cálculos. Para resolver uma situação-problema, o método utilizado é a comparação com outros exercícios semelhantes já resolvidos. Como se fosse possível criar um catálogo. No entanto, isso acarreta em erros, pois é comum haver falha na interpretação do que se quer calcular. É fundamental compreender o que é probabilidade para evitar tais equívocos.

1 Introdução

Segundo [3], os experimentos podem ser de dois tipos: determinísticos ou aleatórios. "Um experimento é determinístico quando repetido em condições semelhantes conduz a resultados essencialmente idênticos." O ato de soltar objetos a uma determinada altura do solo faz com que o mesmo siga em sentido descendente. A natureza do objeto pode retardar o momento de chegada, mas não há dúvida que encontrará o chão. A ação da gravidade no planeta Terra garante que esse experimento é determinístico. Já os aleatórios são aqueles que ao serem repetidos nas mesmas condições podem gerar resultados diferentes, como o lançamento de um dado. Não existe a garantia de que irá sair um determinado número em um dado honesto[1]. O mesmo se pode dizer sobre os números na loteria e outros tantos experimentos sobre os quais não se tem certeza a respeito dos seus resultados. No entanto, em muitos casos, o conjunto dos possíveis resultados do experimento é conhecido previamente.

É de conhecimento geral que a origem da probabilidade foram os jogos de azar. Começou no século XVI com Gerolamo Cardano tentando maximizar seus ganhos em apostas. Segundo Boyer [2], Pierre Simon Laplace foi o primeiro a estruturar a teoria da probabilidade. Isso ocorreu no século XVIII. Para muitos, até hoje, a probabilidade se restringe a cálculos numéricos, utilizados para realizar escolhas sobre o lançamento de um determinado produto ou definir previsões metereológicas. Para além dessa visão pragmática, é preciso entender a probabilidade como uma área da matemática que não tem como objetivo apenas cálculos. A probabilidade não pode se resumir a contas. Não deve apenas buscar valores específicos para cada problema apresentado, mas sim modelos que podem ser aplicados em situações gerais.

* *Palavras chave:* Probabilidade, Bernoulli-Laplace, Von Mises, Kolmogorov, Jeffreys

† C.M. Rui Barbosa, mgamaxxi@gmail.com

‡ CEFET-RJ, andreluiz.cordeiro@gmail.com

§ UERJ, nunes@ime.uerj.br

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

DOI: https://doi.org/10.12957/cadmat.2019.46991

Segundo Feller [5], a probabilidade é semelhante à geometria. Deve ser regida por axiomas e partir de experiências mais simples, como o lançamento de dados para posteriormente ser generalizada em modelos abstratos. Os axiomas são verdades não questionáveis. A geometria é construída a partir de conceitos primitivos e axiomas. O ponto é um conceito primitivo para a geometria. Não se define ou demonstra o que é um ponto. Dizer que o ponto é feito quando se apoia a caneta no papel é apenas um recurso estilístico para passar a ideia do conceito. Não é uma definição. Em sua apresentação axiomática da probabilidade, Kolmogorov [10] considera evento aleatório e sua probablididade como conceitos primitivos.

Feller [5] entende haver três aspectos que devem ser considerados no estudo da probabilidade: (a) o conteúdo lógico formal, (b) o conhecimento intuitivo (experiências anteriores) e (c) as aplicações.

O conteúdo lógico formal é a maneira como a Matemática se estrutura. A geometria se baseia em axiomas e a partir deles demonstra teoremas. Fossa [6] diz que "axioma é uma proprosição aceita sem demonstração." São alicerces. A partir deles são construídas as teorias matemáticas.

Existem conceitos na Matemática que são muito abstratos. No entanto, a probabilidade e a geometria são temas que são mais intuitivos. Muitas vezes, é possível relacionar conceitos dessas áreas a situações cotidianas. Considere que um jogador vence se ao lançar um dado, sair o número 5. É comum ouvir que a chance dele ganhar é "uma em seis." Expressões como esta são vagas, mas são importantes para o conhecimento inicial de probabilidade. As pessoas entendem que a oferta de faces que garantem vitória é de apenas uma em um total de seis. Por ser um objeto familiar, percebem também que isso não significa uma garantia. Pode-se jogar o dado dez vezes e sair sempre o número dois, mesmo para um dado honesto.

O conhecimento intuitivo é importante para o estudo inicial da probabilidade, mas há outro fator central. Em suas aulas sobre o tema, o professor Augusto César Morgado dizia ser essencial que a pessoa que fosse resolver um problema de combinatória ou probabilidade se colocasse na posição de quem vai efetivamente lidar com aquela situação na prática. Exercícios sobre como acomodar pessoas em uma fila ou lançamento de moedas, por exemplo, devem ser tratados como se a pessoa fosse de fato organizar uma fila ou apostar em moedas.

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

Segundo Singh [12], não há uma definição formal universalmente aceita para a probabilidade. Há matemáticos que são mais categóricos. Azevedo [1] afirma que não se define probabilidade. Ela é um axioma. Singh [12] discute quatro tentativas de grandes matemáticos para formalizar a probabilidade: as de Bernoulli-Laplace, Von Mises, Kolmogorov e a de Jeffreys. Elas serão agora brevemente apresentadas.

2 Bernoulli-Laplace

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

Uma das críticas feitas a esta definição é o uso da ideia de equiprovável antes mesmo de se definir probabilidade. Ao se falar em equiprovável, já se utiliza o conceito de probabilidade. Singh [12] afirma que, para resolver isso, alguns autores retiram a palavra equiprovável. Porém isso gera outro problema. Causa confusão sobre os conjuntos de resultados possíveis nos quais os elementos aparecem mais de uma vez. No caso dos números sorteados na Mega Sena, cada um tem a mesma chance de ser sorteado. Diz-se que são resultados equiprováveis. Por outro lado, em uma caixa na qual há 4 bolas verdes e 1 branca, a probabilidade de sortear verde é maior do que sair branca. Ao se retirar a palavra equiprovável, não é possível diferenciar esses conjuntos de possíveis resultados.

A definição não contempla os conjuntos de resultados possíveis nos quais um caso tem probabilidade maior de ocorrer do que outro. Como exemplo, pode-se citar uma apólice de seguro de vida. Imagina-se que a seguradora liste n possíveis causas de morte. Um dos motivos de óbito poderia ser a pessoa estar viajando de avião e o mesmo cair. Neste caso, a probabilidade não é de $\frac{1}{n}$. Mortes por causas naturais ou acidentes de carro são mais comuns. Cada item da lista apresenta diferentes chances de ocorrer.

Azevedo [1] lembra outra crítica à definição clássica. Ela vale somente para conjuntos de resultados finitos. A teoria de Laplace apresenta ainda outro fator limitador. Para aplicá-la é preciso conhecer todos os possíveis resultados de um experimento. Ela foi criada em um contexto de jogos. Experiências sobre as quais se tem conhecimento sobre o que irá sair. No entanto, a teoria clássica não consegue determinar a probabilidade de um senhor de 50 anos viver até os 61.

3 Von Mises

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

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

Segundo Azevedo [1]

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

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

4 Kolmogorov

Segundo Corry [4], no Congresso de Matemática de 1900, Hilbert propôs 23 problemas matemáticos. O sexto deles se referia à axiomatização da probabilidade[2]:

> As investigações sobre os fundamentos da geometria sugerem o problema: tratar da mesma maneira, por meio de axiomas, aquelas ciências físicas nas quais a matemática desempenha um papel importante; no topo da lista estão as teorias das probabilidades e da mecânica [4, p. 84]

Kolmogorov definiu probabilidade a partir da teoria da medida. Concordando com Hilbert, Kolmogorov [10] defende que a probabilidade é uma disciplina da Matemática tal como a Geometria ou a Álgebra e precisa ser regida por axiomas. Eles têm que servir como alicerces para teoremas posteriores.

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

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

Narens [11] afirma que a teoria de Kolmogorov se baseia em seis axiomas, são eles:

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

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

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

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

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

5 Jeffreys

Howie [8] diz que para entender a probabilidade é preciso observar que há dois grupos principais de classificação. Para alguns matemáticos, a probabilidade é uma fração baseada em uma sequência de repetições de um experimento. Ou seja, uma vez que a probabilidade de lançar uma moeda e aparecer cara é de 50%, após uma sequência de jogadas tendendo ao infinito, a quantidade de caras será, aproximadamente, a metade delas.

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

Para Howie [8], o cálculo da probabilidade de jogos de azar é facilmente realizado devido ao conhecimento de todos os resultados possíveis. Essa abordagem não pode ser utilizada para o exemplo de chover em um determinado dia. Neste caso, há uma inversão do cálculo da probabilidade, devido à incerteza sobre os resultados possíveis. Chover deriva de uma conjuntura de fatores. Howie [8] afirma que a primeira menção à "probabilidade inversa" foi feita em um artigo de 1764 de Bayes. Seu método poderia ser aplicado ao sorteio de bolas de uma urna. Após se retirar cada bola, a mesma é reposta na urna. Há bolas pretas e brancas e não se conhece a proporção de cada cor. A partir da observação dos primeiros sorteios, verifica-se que saíram p bolas pretas e q brancas. A probabilidade da próxima ser preta[3] é de ${}^{p+1}/_{p+q+2}$.

Neste contexto é possível entender a perspectiva de Jeffreys. Para Jeffreys ([9] apud [8], 2002), o teorema de probabilidade inversa de Bayes "está para a probabilidade assim como o de Pitágoras está para a geometria" [4].

Howie [8] conta que no início da década de 1930, a teoria de Bayes estava associada à inferência científica e não era considerada útil para a análise de dados. Jeffreys mudou isso. Seu livro *Theory of Probability* foi o primeiro a aplicar a teoria de Bayes. Somente por volta de 1950 que as teorias de Bayes voltaram a ter seu valor e o status de estatístico foi atribuído novamente à Jeffreys. No entanto, Singh [12] afirma que a maioria dos matemáticos não aceita a abordagem do autor.

> A hipótese principal da teoria de Jeffreys é que a toda proposição p feita sobre dados q podemos atribuir uma única probabilidade e que probabilidades de diferentes proposições sobre diferentes dados são comparáveis. Mas ele não deu qualquer indicação de como esse número é calculado. Singh [12, p. 258][5]

6 Nos livros didáticos

Dentre essas abordagens sobre probabilidade, a mais utilizada é a de Kolmogorov, sendo a de Laplace considerada um caso particular para conjuntos finitos de resultados possíveis. A noção de probabilidade baseada na teoria de conjuntos está presente em textos didáticos. Para ilustrar este fato, vamos considerar a abordagem de Morgado [3] para probabilidade. Inicialmente, eles citam Cardano em linguagem moderna: o Espaço Amostral é o conjunto de todos os resultados possíveis de um determinado experimento. Já a probabilidade é o quociente da quantidade de casos favoráveis pelo total de resultados possíveis de um experimento aleatório. Para entender o que significam casos favoráveis, pode-se pensar em uma aposta. São os casos nos quais o jogador vence. A situação procurada, é chamada de "evento".

[3] Este resultado também foi obtido independentemente por Laplace em 1774 e é conhecido como Regra de Sucessão de Laplace. Para mais detalhes consulte Gorroochurn [7].

[4] o Theorem of Inverse Probability de Bayes is "to the theory of probability what Pythagoras's theorem is to geometry

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

Neste exemplo, é a vitória do participante. O evento é um subconjunto do espaço amostral. É importante lembrar que essa definição de probabilidade se refere aos espaços amostrais finitos e equiprováveis[6].

Mais formalmente, Morgado [3] consideram A um subconjuto de um espaço amostral S e definem:

> A probabilidade é uma função que associa a cada evento A um número $P(A)$ (a probabilidade do evento A ocorrer) de forma que:
>
> 1. Para todo evento A, $0 \leq P(A) \leq 1$
> 2. $P(S) = 1$
> 3. Se A e B são eventos mutuamente excludentes, isto é, eventos que não podem ocorrer simultaneamente (isto é, $A \cap B = \emptyset$) então: $P(A \cup B) = P(A) + P(B)$.

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

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

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

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

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

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

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

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

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

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

7 Considerações finais

A construção do conhecimento matemático requer definições criteriosas de seus conceitos e a associação, sempre que possível, de exemplos a situações do cotidiano para garantir o entendimento. O balanço entre abstração e intuição é fundamental no ensino de matemática, em particular, no de probabilidade. Valorizar o conhecimento intuitivo garante um ganho posterior sobre as generalizações. O raciocínio intuitivo é importante, mas não pode ser um fim em si mesmo. É preciso haver um jeito formal de explicar estes exemplos, ou seja, suas generalizações. É importante também privilegiar a compreensão dos conceitos através de problemas não operacionais. Eles permitem identificar melhor os obstáculos de entendimento e a garantir uma sólida construção do conhecimento.

Agradecimentos

Os autores agradecem o apoio da FAPERJ, CAPES e CNPq a este trabalho.

Referências

[1] AZEVEDO, Cecília O que é a probabilidade? Interpretações da probabilidade, DMAT – Comunicações, 2004, p. 1-14.

[2] BOYER, Carl Benjamin e MERZABACH, Uta Caecilia. História da Matemática, Blucher, 2012.

[3] CARVALHO, Paulo Cezar Pinto, MORGADO, Augusto Cezar de Oliveira, FERNANDEZ, Pedro e PITOMBEIRA, João Bosco. Análise Combinatória e Probabilidade, SBM, 2016.

[4] CORRY, Leo. David Hilbert and the Axiomatization of Physics (1894-1905), Archive for History of Exact Sciences, 51 (2), 1997, p. 83-198

[5] FELLER, William. An Introduction to Probability Theory and its Applications, John Wiley & Sons Inc., 1957.

[6] FOSSA, J.A. Introdução ás técnicas de demonstração na Matemática, Contextos da ciência, 2009.

[7] GORROOCHURN, P. Classic Topics on the History of Modern Mathematical Statistics: From Laplace to More Recent Times, Wiley, 2016.

[8] HOWIE, David. Interpreting Probability Controversies and Developments in the Early Twentieth Century Cambridge University Press, 2002.

[9] JEFFREYS, Harold. The Theory of Probability, OUP, 1950.

[10] KOLMOGOROV, Andrey Nikolaevich. Foundations of the Theory of Probability, Chelsea Publishing Company, 1956.

[11] NARENS, Louis. Theories of Probability: an Examination of Logical and Qualitative Foundations, World Scientific, 2007.

[12] SINGH, Jagjit. Theories of Probability Sankhyā: The Indian Journal of Statistics (1933-1960), 7(3), 1946, p. 257–262.