REVISEM
Revista Sergipana de Matemática e Educação Matemática

https://doi.org/10.34179/revisem.v5i2.13272

O CÁLCULO DE PROBABILIDADES SOB AS ABORDAGENS CLÁSSICA E FREQUENTISTA

Fernanda Vital de Paula
Universidade Federal do Tocantins
fernandavital@uft.edu.br

Resumo

O cálculo da probabilidade de um evento A é permitido por diversas abordagens. Entre elas, pode-se citar as abordagens clássica e a frequentista. Ambas se relacionam por meio de um importante teorema denominado Lei dos Grandes Números. A ideia principal deste trabalho é demonstrar essa relação. Além disso, evidencia-se a inviabilidade prática das repetições do experimento exigido no cálculo de probabilidade por meio da abordagem frequentista e indica-se o software R-project como uma poderosa ferramenta para a realização de tais repetições por meio de simulações.

Abstract

The computation of the probability of an event is allowed by several approaches. Among the possibilities, classical and frequentist approaches can be mentioned. Both are related through an important theorem called the Law of Large Numbers. The main idea of this work is to demonstrate it. In addition, the practical infeasibility of the repetitions of the experiment required in the calculation of probability through the frequentist approach is evidenced and the R-project software is indicated as a powerful tool for carrying out such repetitions through simulations.

1 Introdução

O cálculo de probabilidade de um evento é possibilitado por meio de algumas abordagens: clássica, frequentista, subjetivista, geométrica e axiomática. Neste trabalho, serão destacadas duas delas; a clássica e a frequentista.

Historicamente, a fórmula utilizada pela abordagem clássica para o cálculo de probabilidade de um evento se concretizou em 1654, quando os importantes matemáticos Pascal (1623-1662) e Fermat (1607-1665), por meio de cartas trocadas, expuseram um método de resolução para um problema proposto pelo famoso jogador Chevalier de Méré (1607-1684). A resolução continha a famosa razão utilizada para o cálculo de probabilidades nessa abordagem que será exposta na próxima seção. Por outro lado, na abordagem frequentista, Von Mises (1881-1973) definiu a probabilidade de um evento como um limite de frequências se apropriando de termos e conceitos de cálculo, em 1919.

A abordagem clássica entende o cálculo de probabilidade de um evento como a razão entre o número de casos favoráveis e o número total de casos possíveis, assumindo a equiprobabilidade de todos os elementos que compõem o espaço amostral associado a um experimento aleatório. Nesse caso, a probabilidade é calculada, geralmente, sem necessidade de realização de experiências.

Na abordagem frequentista, o cálculo de probabilidade de um evento é permitido por um processo de experimentação. Nesse caso, a probabilidade é compreendida como a frequência relativa de sucessos obtidos durante a realização de um experimento. Aqui, a probabilidade é obtida como uma aproximação que depende da quantidade de realizações do experimento e está amparada em um importante teorema conhecido como Lei dos Grandes Números.

Trabalhar com estas duas abordagens simultaneamente permite a confrontação de ambas possibilitando um melhor entendimento e domínio do cálculo de probabilidades. Além disso, as duas abordagens possibilitam a ampliação dos conceitos matemáticos que podem ser explorados no cálculo de probabilidade, dado que a abordagem clássica está relacionada aos conceitos de razão e análise combinatória, enquanto a frequentista está associada aos conceitos de proporção, limite e convergência.

Neste sentido, este trabalho tem como objetivos apresentar o cálculo de probabilidades conforme as abordagens clássica e frequentista, mostrar como os cálculos referentes a ambas as abordagens estão relacionados por meio da Lei dos Grandes Números e mostrar que o software livre R-project pode ser um grande aliado na simulação de experimentos que permitam o entendimento do cálculo de probabilidades por meio da abordagem frequentista.

2 Preliminares

Nesta seção serão apresentadas as definições que se referem ao cálculo de probabilidade de um evento pelas abordagens clássica [2] e frequentista [4] e como ambos se relacionam matematicamente.

2.1 Cálculo de probabilidades

Definição 2.1. Abordagem Clássica. *Considere um experimento aleatório onde todos os elementos do espaço amostral* Ω *tenham a mesma probabilidade de ocorrer e* *seja A um evento de interesse de Ω. Então, a probabilidade de ocorrência de A é dada por:*

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

onde $n(A)$ e $n(\Omega)$ correspondem ao número de elementos de A e Ω, respectivamente.

Definição 2.2. Abordagem Frequentista. *Suponha que o experimento foi repetido n vezes, sempre sob as mesmas condições, e que o evento A ocorreu m vezes entre essas n realizações do experimento. Então, a razão seguinte é uma boa aproximação para a probabilidade de A, se o número n for bastante grande:*

$$P(A) \approx \frac{m}{n}.$$

De outra maneira, a probabilidade de um evento A é o limite de sua frequência relativa em muitas tentativas

$$P(A) = \lim_{n \to \infty} \frac{m}{n}.$$

As dificuldades observadas no cálculo da probabilidade de um evento de interesse A, por meio da abordagem clássica, são a necessidade de existência de equiprobabilidade dos elementos do espaço amostral e a necessidade de que o número de elementos do espaço amostral seja finito. Por outro lado, a abordagem frequentista supera tais dificuldades da abordagem clássica, porém, impõe outras como um número grande de repetições do experimento sob as mesmas condições. Na prática, tal exigência no processo de repetição é inviável na maioria das vezes.

Para que um grande número de repetições de um experimento seja possível com manutenção das mesmas condições em cada repetição, uma alternativa viável é a simulação do experimento em softwares apropriados. Neste sentido, dado que o R-project é livre e gratuito, o mesmo pode ser um grande aliado no cálculo de probabilidades por meio da abordagem frequentista.

2.2 Relação entre as abordagens clássicas e frequentista

A Lei Fraca dos Grandes Números (LFGN) [1], um importante teorema da teoria probabilística, permite que se compreenda a obtenção do mesmo valor para a probabilidade de um determinado evento A independente da abordagem utilizada; clássica ou frequentista. Em outras palavras, a LFGN possibilita o entendimento de como a abordagem clássica e frequentista estão relacionadas.

Relembraremos alguns conceitos importantes da teoria de probabilidade: variável aleatória [1], modelo de Bernoulli [1], modelo Binomial [2] e esperança matemática [2].

Definição 2.3. Variável aleatória. *Uma variável aleatória é uma função de um espaço amostral* Ω *nos números reais. As variáveis aleatórias podem ser classificadas em discretas ou contínuas conforme a natureza numérica dos valores que assumem.*

Definição 2.4. Modelo de Bernoulli. *Uma variável aleatória* X *tem uma distribuição de Bernoulli se*

$$X = \begin{cases} 1, \textit{com probabilidade } p \\ 0, \textit{com probabilidade } 1-p \end{cases} .$$

Definição 2.5. Modelo Binomial. *A variável aleatória definida como o número de sucessos em* n *provas independentes de Bernoulli é chamada variável Binomial* X *e sua distribuição de probabilidades é:*

$$P(X) = C_n^x p^x q^{n-x}.$$

Indica-se $X \sim B(n, p)$*. Para o Modelo Binomial, tem-se* $\mathbb{E}(X) = np$*.*

Definição 2.6. Esperança Matemática. *Considere uma variável aleatória discreta* X *com conjunto de valores possíveis* D*. A esperança matemática de* X *é a soma de todos os produtos da variável aleatória pela sua respectiva probabilidade, desde que a soma exista.*

$$\mathbb{E}(X) = \sum_{x \in D} xP(X = x).$$

Teorema 2.7. Lei Fraca dos Grandes Números. *Sejam* $X_1, X_2, \ldots$ *variáveis aleatórias independentes e identicamente distribuídas (iid) com* $\mathbb{E}(X_i) = \mu$ *e* $V(X_i) = \sigma^2 < \infty$*. Definimos* $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$*. Então, para cada* $\epsilon > 0$*,*

$$\lim_{n \to \infty} P\left(|\overline{X}_n - \mu| < \epsilon\right) = 1,$$

isto é, $\overline{X}_n$ *converge em probabilidade para* μ*.*

O Teorema 2.7 apresenta o que se conhece como a Lei Fraca dos Grandes Números. Sob condições gerais, ele estabelece que a média amostral se aproxima da média da população à medida que $n \to \infty$.

Considere $\Omega = \{a_1, a_2, \ldots, a_n\}$ o espaço amostral finito equiprovável correspondente a um experimento aleatório E. Considere ainda o evento $A \subset \Omega$ com k elementos. Pela abordagem clássica, obtém-se $P(A) = \frac{k}{n} = p$.

Agora, considere N ensaios de Bernoulli de modo que

$$X_i = \begin{cases} 1, \text{se } a_i \in A \\ 0, \text{se } a_i \notin A \end{cases}.$$

Observe que $P(X_i = 1) = p$.

Seja $Y_N = \sum_{i=1}^{N} X_i$, a variável que representa o número de sucessos nos N ensaios. Segue que $Y \sim B(N, p)$. Logo, $\mathbb{E}(Y) = Np$ e, consequentemente, $\mathbb{E}(\overline{Y}) = p$.

Pela Lei Fraca dos Grandes Números,

$$\lim_{N \to \infty} P\left(|\overline{Y}_N - p| < \epsilon\right) = 1,$$

onde $\epsilon > 0$.

O resultado garante que a probabilidade de um evento A de interesse por meio da abordagem frequentista, fornecida pela frequência relativa de sucessos obtida nos N ensaios, se aproxima da probabilidade do evento A fornecida pela abordagem clássica, $P(A)$, à medida que $N \to \infty$.

Para mais detalhes sobre a Lei Fraca dos Grande Números, o leitor pode consultar [1]. Quanto à interpretação de resultados envolvendo limites, recomenda-se [3].

3 Resultados Principais

Nesta seção, iremos considerar o seguinte experimento aleatório E : lançamento de um dado equilibrado. Aqui, o fato de o dado ser equilibrado garante que as probabilidades associadas às ocorrências de cada face são as mesmas. Nesse caso, tem-se $\Omega = \{1, 2, 3, 4, 5, 6\}$. Considere os eventos A_i : ocorrência da face i, isto é, $A_i = \{i\},\ i = 1, 2, 3, 4, 5$ e 6. Pela abordagem clássica, tem-se

$$P(A_i) = \frac{n(A_i)}{n(\Omega)} = \frac{1}{6}.$$

Para calcular a probabilidade de ocorrência de A_i por meio da abordagem frequentista, é necessário que um grande número de lançamentos do dado equilibrado seja realizado sob as mesmas condições. Na prática, garantir as mesmas condições nas repetições de um experimento é impossível na maioria das vezes. Para isso, pode-se simular tais lançamentos em um software e verificar as frequências relativas de sucesso para cada evento.

A Tabela 1 exibe as probabilidades obtidas e aproximadas para cada evento A_i, conforme as abordagens clássica e frequentista, respectivamente. Para esta última, números de repetições diversos são considerados em simulações no R-project para que o leitor possa ter uma ideia da importância associada ao grande número de repetições do experimento necessário nessa abordagem. A título de simplificações, usa-se na tabela as seguintes abreviações: abordagem clássica ($A.C.$) e abordagem frequentista com n repetições ($A.F._n$). Quanto à forma numérica adotada, optou-se pela decimal para que o leitor visualize mais claramente o comportamento dos valores calculados.

Tabela 1: Probabilidades dos eventos A_i por meio das abordagens clássica e frequentista.

	Eventos					
Abordagens	A_1	A_2	A_3	A_4	A_5	A_6
$A.F._6$	0	$0,333333$	$0,166667$	$0,166667$	0	$0,333333$
$A.F._{60}$	$0,116667$	$0,150000$	0.233333	$0,166667$	$0,216667$	$0,116667$
$A.F._{600}$	$0,155000$	$0,156667$	$0,211667$	$0,155000$	$0,155000$	$0,166667$
$A.F._{6000}$	$0,170500$	$0,158833$	$0,177667$	$0,153667$	$0,165500$	$0,173833$
$A.F._{60000}$	$0,168267$	$0,165333$	$0,165583$	$0,166000$	$0,167467$	$0,167350$
$A.F._{600000}$	$0,166882$	$0,166540$	$0,166223$	$0,167315$	$0,167012$	$0,166028$
$A.C.$	$0,166667$	$0,166667$	$0,166667$	$0,166667$	$0,166667$	$0,166667$

Como esperado, observe que a frequência relativa do sucesso de obtenção dos eventos A_i se aproxima da probabilidade calculada para os mesmos conforme a abordagem clássica. Neste sentido, aproximando os valores obtidos para duas casas decimais obtém-se $A.F._{60000} = A.C.$ para todos os eventos A_i.

A simulação dos lançamentos do dado em software garante a manutenção das mesmas condições em cada repetição do experimento, conforme exige a abordagem frequentista. Para obter os valores exibidos na tabela, pela abordagem frequentista, uma rotina simples no R-project foi executada a fim de simular lançamentos do dado e calcular as frequências relativas de interesse. A mesma é exibida no Quadro 1. Observe que para cada número de lançamentos considerado, é necessário que o valor de n seja alterado. Na rotina exibida, $n = 6$, ou seja, seis lançamentos do dado foram simulados.

Quadro 1: Rotina executada no R-project para simulação de seis lançamentos do dado.

```
set.seed(1) #fixando o experimento
n=6 #número de lançamentos
faces=c(1,2,3,4,5,6) #possíveis faces do dado
lançamentos=sample(faces,n,replace=TRUE) #realizando os n lançamentos
table(lançamentos)/n #indica a frequência de obtenção de cada face
```

Outra forma de evidenciar a aproximação das frequências relativas obtidas para cada face, conforme o aumento do número de repetições, da probabilidade de $\frac{1}{6}$ obtida pela abordagem clássica, é por meio de uma análise gráfica. A Figura 1 exibe o comportamento das frequências relativas obtidas para cada face do dado conforme o número de repetições varia entre 1 e 60000 lançamentos. A linha vermelha está fixada em $\frac{1}{6}$.

Pelos gráficos da Figura 1, observa-se que as frequências relativas obtidas para cada face começam a se estabilizar em torno de $\frac{1}{6}$ a partir de cerca de 30000 lançamentos. Este número de lançamentos evidencia a dificuldade em estimar a probabilidade de um evento por meio da abordagem frequentista na vida real.

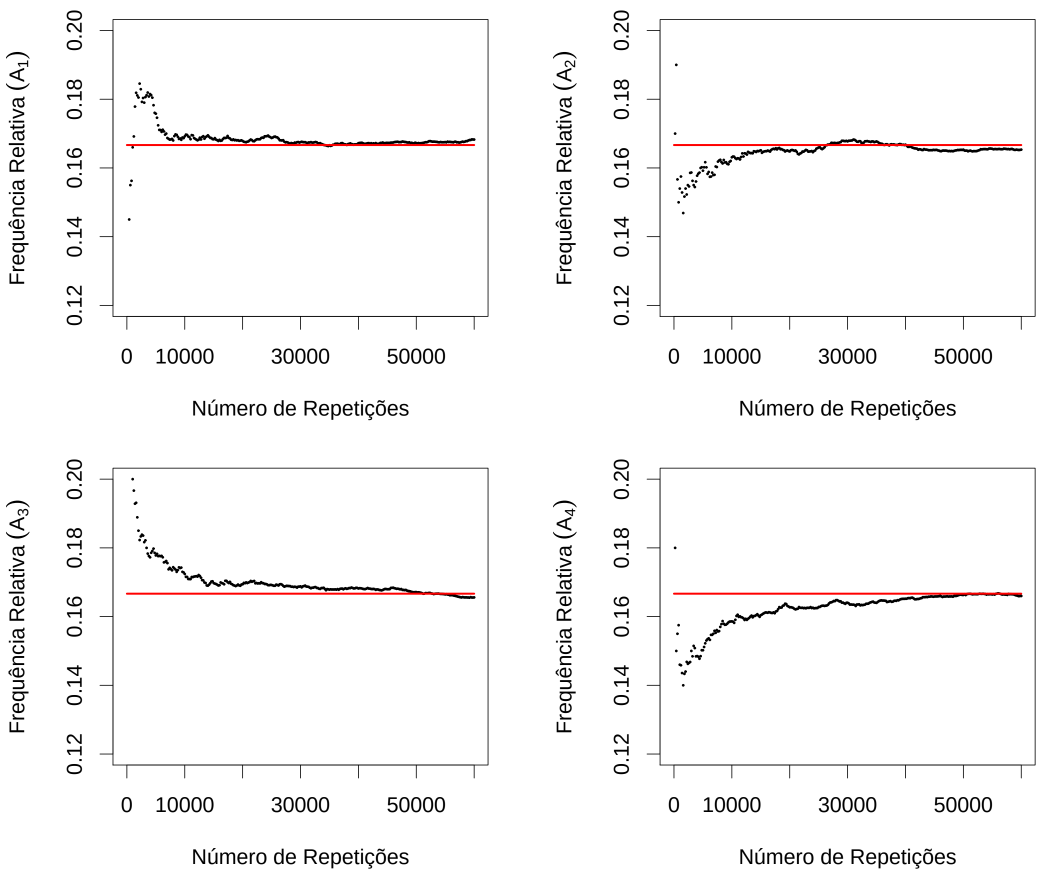

Figura 1: Frequências relativas obtidas para os eventos A_i.

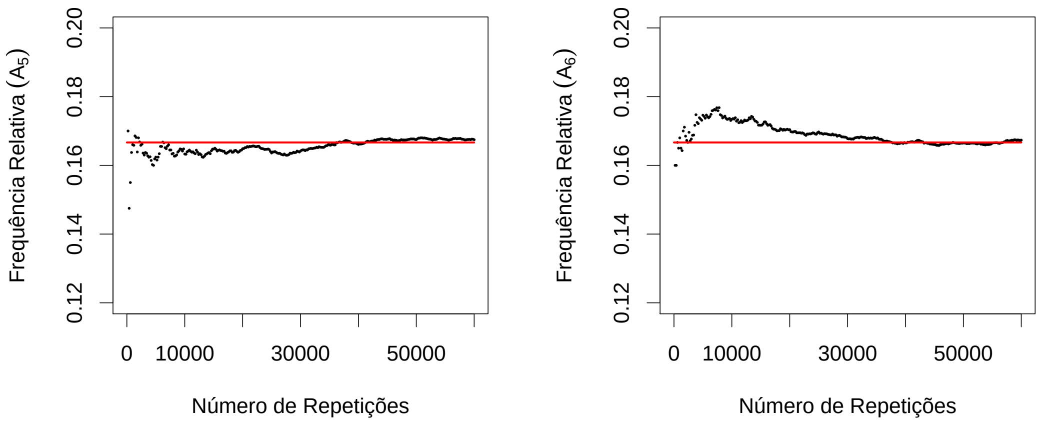

Figura 2: Frequências relativas obtidas para os eventos A_i.

4 Considerações Finais

A respeito do cálculo de probabilidade de um evento, é necessário conhecer as diversas abordagens possíveis para tal e refletir sobre os possíveis obstáculos associados a cada uma delas considerando as diversas formas de ensino e aprendizagem.

O trabalho vem ao encontro desta necessidade apresentando o cálculo de probabilidades de um evento por meio das abordagens clássica e frequentista e demonstrando a relação existente entre ambas.

Na prática, a dificuldade em calcular a probabilidade por meio da abordagem frequentista é evidenciada. Dessa forma, apresenta-se o R-project como importante ferramenta para transpor essa dificuldade.

Espera-se que este trabalho possa motivar os leitores envolvidos com o ensino e aprendizagem de matemática a utilizarem as ideias apresentadas no ensino de probabilidade e inspire os leitores curiosos na utilização do R-project para simulação de outros experimentos além do exemplo apresentado.

Referências

[1] G. Casella R.L. Berger, *Inferência Estatística*. São Paulo: Cengage Learning, 2010.

[2] G. O. Costa, *Curso de Estatística Inferencial e Probabilidades: Teoria e Prática.* São Paulo: Atlas, 2012.

[3] L. Leithold, *O cálculo com geometria analítica, vol. 1.* Rio de Janeiro: Editora Harbra Ltda, 1994.

[4] J. I. D. Pinheiro, S. B. Cunha, S. R. Carvajal e G. C. Gomes, *Estatística Básica: a Arte de Trabalhar com Dados.* Rio de Janeiro: Elsevier, 2009.

Submetido em 21 de Março de 2020.
Aceito em 03 de Agosto de 2020.