

A LEI FRACA DOS GRANDES NÚMEROS E A FALÁCIA DO JOGADOR EM UM MODELO PROBABILÍSTICO PARA A MEGA SENA

Fernanda Vital de Paula
Universidade Federal do Tocantins
fernandavital@uft.edu.br
Kevellyn Samara Lima da Silva
Universidade Federal do Tocantins
kevellyn.samara@uft.edu.br

Resumo

Os jogos de azar estão presentes no cotidiano há milhares de anos. No Brasil, a Mega-Sena é o maior jogo de azar, na modalidade lotérica. Para quantificar o quão provável é uma vitória na Mega-Sena, modelos probabilísticos podem ser utilizados. Este trabalho parte de um modelo proposto na literatura com o principal objetivo de mostrar que uma crença bastante comum dos apostadores ao realizarem seus jogos, denominada falácia do jogador, está relacionada à uma interpretação equivocada da Lei Fraca dos Grandes Números. Para tal, além de evidências teóricas, resultados empíricos são apresentados, via simulação no *software R-project*.

Palavras-chave: Jogos de Azar. Modelo Probabilístico. Simulação.

Abstract

Gambling has been present in everyday life for thousands of years. In Brazil, Mega-Sena is the biggest game of chance in the lottery mode. To quantify how likely a victory at the Mega-Sena is, probabilistic models can be used. This work is based on a model proposed in the literature with the main objective of showing that a very common belief of gamblers when playing their games, called the gambler's fallacy, is related to a misinterpretation of the Weak Law of the Great Numbers. For this, in addition to theoretical evidence, empirical results are presented, via simulation in the R-project software.

Keywords: Gambling. Probabilistic Model. Simulation.

1 Introdução

Há milhares de anos, o ser humano se relaciona através dos jogos de azar. Gadelha [5] faz referência a pinturas feitas em 3.500 a.C. em tumbas egípcias mostrando pessoas jogando uma espécie de dados com quatro faces feitos de um osso do calcanhar de nome astrágalo. Com a evolução das tecnologias, novos jogos surgiram e os antigos evoluíram e adaptaram-se à cada civilização.

Hoje, muitos jogos de azar são populares entre apostadores que acreditam na sorte e desejam faturar prêmios. No caso do Brasil, a maior modalidade lotérica dessa categoria é a Mega-Sena.

Para quantificar o quão provável é a vitória na Mega-Sena, conceitos de Análise Combinatória e Probabilidade são necessários. Apesar disso, não existe uma fórmula que garanta a vitória de um apostador. Todavia, algumas pessoas possuem suas próprias crenças em sequências que acreditam ser mais prováveis e não ousam se arriscar em apostas que acreditam ser menos prováveis. Tais crenças, geralmente, não possuem argumentação matemática válida.

No que diz respeito a estudos matemáticos, Nomelini *et al.* [9] propuseram um modelo probabilístico para a Mega-Sena que, posteriormente, foi utilizado por Araújo [1], em sua Monografia de Graduação em Licenciatura em Matemática.

Este trabalho parte do referido modelo com o principal objetivo de mostrar que uma crença bastante comum dos apostadores ao realizarem seus jogos, denominada *Falácia do Jogador*, está relacionada à uma interpretação equivocada da *Lei Fraca dos Grandes Números* (LFGN). Para tal, além de evidências teóricas, resultados empíricos são apresentados, via simulação no *software* R-project.

Para facilitar a compreensão do leitor, na próxima seção são apresentados algumas definições e um teorema necessários para o entendimento dos resultados deste artigo.

2 Preliminares

Algumas técnicas de contagem, fornecidas pela Análise Combinatória, são necessárias para o estudo de jogos de azar. No caso da Mega-Sena, por exemplo, a Combinação apresenta-se como pré-requisito indispensável. A Combinação é uma consequência do *Princípio Fundamental da Contagem (P.F.C)*, denominado também como *Princípio da Multiplicação*. Ambas as técnicas são descritas a seguir.

- **P.F.C.** Considere r conjuntos A_1, A_2, \dots, A_r com n_1, n_2, \dots, n_r elementos, respectivamente. Então, o número de r -uplas ordenadas (sequências de r elementos) do tipo (a_1, a_2, \dots, a_r) , onde $a_i \in A_i$, $a_2 \in A_2, \dots, a_r \in A_r$ é dado por $n_1.n_2. \dots .n_r$.

- **Combinação.** Seja M um conjunto com m elementos, isto é, $M = \{a_1, a_2, \dots, a_m\}$. Chamamos de combinações dos m elementos, tomados r a r , aos subconjuntos de M constituídos de r elementos. Nesse caso, o número de combinações é dado por C_m^r de modo que

$$C_m^r = \frac{m!}{r!(m-r)!} \quad \forall m, r \in \mathbb{N}^* \quad e \quad r \leq m,$$

onde $!$ é o símbolo de *fatorial* definido por

$$m! = m \cdot (m-1) \cdot (m-2) \cdot \dots \cdot 3 \cdot 2 \cdot 1 \quad m \geq 2,$$

de modo que $0! = 1$ e $1! = 1$.

Para um maior aprofundamento em Análise Combinatória, pode-se consultar Haz-zan [2], autor no qual estão baseadas as três definições seguintes.

Além dos conceitos de Análise Combinatória, um teorema importante para este artigo, é a LFGN. Para o entendimento da LFGN, faz-se necessário o conhecimento de alguns conceitos que serão apresentados a seguir.

Cálculo de Probabilidade pela Abordagem Clássica. Seja Ω um espaço amostral equiprovável com n elementos e $D \subset \Omega$ um evento contendo d elementos. A probabilidade de ocorrência de D é dada pela razão entre as quantidades de elementos em D e Ω , isto é,

$$P(D) = \frac{n(D)}{n(\Omega)} = \frac{d}{n}. \quad (2.1)$$

Definição 2.1. 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.*

Considerando novamente o experimento aleatório “Três moedas são lançadas e as faces que ficam viradas para cima são observadas”, pode-se definir a variável aleatória discreta X como sendo “o número de vezes que a face cara é obtida”. Nesse caso, $X = \{0, 1, 2, 3\}$. Tem-se:

x	0	1	2	3
$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Definição 2.2. Esperança Matemática. *A esperança matemática de uma variável aleatória discreta X é a soma dos produtos de cada valor da variável aleatória por sua respectiva probabilidade.*

$$\mathbb{E}(X) = \sum_{i=1}^n x_i P(x_i).$$

Para o Exemplo da Definição 2.1, tem-se

$$\mathbb{E}(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = 1,5.$$

Apesar de não ser possível a obtenção de 1,5 caras, esse é o valor que se espera obter, em média, para o número de caras na repetição do experimento.

Definição 2.3. Variáveis aleatórias mutuamente independentes. *Sejam X_1, \dots, X_n variáveis aleatórias com função densidade de probabilidade (fdp) ou função de probabilidade (fp) conjunta $f(x_1, \dots, x_n)$. Suponhamos que $f_{X_i}(x_i)$ denote a fdp ou fp marginal de X_i . Então X_1, \dots, X_n são chamadas de variáveis aleatórias mutuamente independentes se, para cada (x_1, \dots, x_n) ,*

$$f(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i).$$

Para o caso em que a variável aleatória é contínua na Definição 2.2, calcula-se a esperança utilizando integração e, tendo em vista que a Definição 2.3 envolve os conceitos de fp e fdp marginal e conjunta, não apresentados neste trabalho, recomenda-se [3] para um maior entendimento.

Teorema 2.4. LFGN. *Sejam X_1, X_2, \dots variáveis aleatórias independentes e identicamente distribuídas (iid) com $\mathbb{E}(X_i) = \mu$ e $V(X_i) = \sigma^2 < \infty$. Definimos $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Então, para cada $\epsilon > 0$,*

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < \epsilon) = 1,$$

isto é, \bar{X}_n converge em probabilidade para μ .

O Teorema 2.4 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 $n \rightarrow \infty$.

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

Para mais detalhes sobre a LFGN, o leitor pode consultar Casella e Berger [3], autores nos quais estão baseadas as quatro últimas definições. Quanto à interpretação de resultados envolvendo limites, recomenda-se Leithold [7].

Definição 2.5. Falácia do jogador. *Crença de que a ocorrência de desvios no comportamento esperado para uma sequência de eventos independentes de algum processo aleatório implica uma maior probabilidade de se obter, em seguida, desvios na direção oposta.*

Conforme a Definição 2.5, existe a falsa crença de que um evento aleatório é menos provável de ocorrer, em um determinado momento, se o evento ocorreu recentemente. Neste sentido, Tversky e Kahneman [12] sugerem que muitas pessoas acreditam que pequenas amostras devem se comportar como grandes amostras e que uma amostra que se desvia da proporção esperada para grandes amostras deve reverter rapidamente para essa proporção.

A Falácia do Jogador é tipicamente manifestada pela crença de que um evento aleatório tem maior probabilidade de ocorrer por não ter acontecido por um determinado período de tempo [8].

Uma exemplificação dessa ideia é dada pelo experimento E : lançamento de uma moeda não-viciada. O termo não-viciada garante que a probabilidade de obter cara e coroa em um lançamento são iguais, ou seja, de 0,5. Após 30 lançamentos seguidos da moeda, com obtenção de 30 caras, o jogador tende a acreditar que a probabilidade de uma coroa ser obtida no próximo lançamento é maior que 0,5. Essa falácia é gerada pela falta de atenção ao fato de que cada jogada é uma realização independente [12].

Clotfelter e Cook [4] utilizaram jogos de loterias para evidenciar que a Falácia do Jogador influencia os padrões de apostas em determinados números. Os autores descobriram que a quantidade de dinheiro apostada em um determinado número tende a cair drasticamente após o mesmo ter sido sorteado. Tal fato evidencia a ocorrência da Falácia do Jogador no cotidiano.

Por meio deste contexto de jogos de loteria, faremos algumas reflexões a respeito da LFGN e a Falácia do Jogador em um estudo realizado a partir do modelo probabilístico para a Mega-Sena proposto por Nomelini *et al.* [9].

2.1 Problema motivador

A probabilidade de acerto na Mega-Sena é pequena, principalmente quando se faz uma única aposta simples. Com o intuito de aumentar a probabilidade de acerto, muitas pessoas buscam encontrar uma lógica ou determinar um padrão para os números sorteados. Para isso, alguns consultam resultados de concursos anteriores para tentar prever os números de concursos futuros e aumentarem a probabilidade de vitória. Outras pessoas utilizam a estratégia de aposta nos números, que acreditam ser, seus números da sorte. No entanto, em sua maioria, não existe uma fundamentação teórica para explicar o possível aumento da probabilidade de vencer nessas situações. Teoricamente, a utilização destes métodos garante a mesma probabilidade de vitória de uma aposta em uma sequência qualquer, sem se basear em padrões de escolha.

Do ponto de vista científico, modelos probabilísticos podem ser utilizados para capturar aspectos quantitativos do comportamento de um experimento aleatório como, por exemplo, o padrão de comportamento dos números sorteados na Mega-Sena.

Neste sentido, Nomelini *et al.* [9] apresentaram um modelo que se baseia no pensamento dos possíveis resultados da Mega-Sena não como combinações individuais, mas como grupos com o mesmo comportamento. Para isso, os autores classificaram as possibilidades numéricas da Mega-Sena (números naturais de 1 a 60) nos seguintes intervalos de dezenas:

$$[1, 10], [11, 20], [21, 30], [31, 40], [41, 50] \text{ e } [51, 60].$$

Conforme essa classificação, os autores definiram 11 gabaritos possíveis de acordo com a Tabela 1.

É possível calcular a probabilidade de ocorrência de cada gabarito utilizando o número de combinações possíveis para cada um e o número total (NT) de combinações da Mega-Sena, que é dado por

$$NT = C_{60}^6 = 50.063.860.$$

1. Gabarito A

O número de combinações possíveis do gabarito A é definido por $n(A)$. Como são seis números pertencentes a dezenas diferentes, de forma que cada número pode ser escolhido entre 10 possibilidades, tem-se

$$n(A) = 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 10^6 = 1.000.000.$$

Assim, a probabilidade de ocorrência do gabarito A , conforme (2.1), é dada por

$$P(A) = \frac{n(A)}{NT} = \frac{1.000.000}{50.063.860} \approx 0,01997.$$

Tabela 1: Possíveis gabaritos (gab.) conforme proposta de Nomelini *et al.* [9].

Gab.	Descrição	Exemplo
<i>A</i>	Seis números em dezenas diferentes.	01 – 12 – 21 – 33 – 45 – 51
<i>B</i>	Dois números em uma dezena e os demais em outras dezenas diferentes entre si.	12 – 15 – 23 – 38 – 45 – 52
<i>C</i>	Dois números em uma dezena, dois números em outra dezena e os demais em dezenas diferentes entre si.	04 – 09 – 25 – 34 – 36 – 48
<i>D</i>	Três pares de números em três dezenas diferentes entre si.	02 – 06 – 25 – 28 – 54 – 60
<i>E</i>	Três números em uma dezena, dois em outra dezena e o último em uma terceira dezena.	12 – 15 – 16 – 36 – 42 – 45
<i>F</i>	Três números em uma dezena e três números em dezenas diferentes entre si.	25 – 28 – 30 – 36 – 41 – 52
<i>G</i>	Três números em mesma dezena e os demais em outra dezena.	04 – 07 – 09 – 31 – 35 – 39
<i>H</i>	Quatro números em uma dezena e os outros em dezenas diferentes entre si.	11 – 13 – 15 – 19 – 35 – 54
<i>I</i>	Quatro números em uma mesma dezena e os demais em outra dezena.	22 – 25 – 26 – 30 – 41 – 46
<i>J</i>	Cinco números em uma mesma dezena e o último em uma dezena diferente.	31 – 32 – 35 – 38 – 40 – 59
<i>K</i>	Seis números em uma mesma dezena.	21 – 22 – 25 – 27 – 28 – 30

2. Gabarito *B*

No gabarito *B*, os dois primeiros algarismos podem ser escolhidos entre seis dezenas e os outros quatro números de dez maneiras diferentes entre as cinco dezenas restantes. Segue que o número de combinações possíveis para *B* é dado por

$$n(B) = C_6^1 \cdot C_{10}^2 \cdot C_5^4 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 6 \cdot 45 \cdot 5 \cdot 10000 = 13.500.000.$$

Portanto,

$$P(B) = \frac{n(B)}{NT} = \frac{13.500.000}{50.063.860} \approx 0,26966.$$

3. Gabarito C

Para o gabarito C , os dois primeiros algarismos podem ser escolhidos entre seis dezenas, o segundo par entre cinco dezenas e os demais podem ser escolhidos entre as quatro restantes e de dez formas diferentes cada um. O número de combinações existentes, nesse caso, é dado por

$$n(C) = C_6^2 \cdot C_{10}^2 \cdot C_{10}^2 \cdot C_4^2 \cdot 10 \cdot 10 = 15 \cdot 45 \cdot 45 \cdot 6 \cdot 100 = 18.225.000.$$

Assim,

$$P(C) = \frac{n(C)}{NT} = \frac{18.225.000}{50.063.860} \approx 0,36404.$$

4. Gabarito D

Como o gabarito D é composto por três pares de algarismos em três dezenas diferentes entre si, tem-se

$$n(D) = C_{6,3} \cdot C_{10,2} \cdot C_{10,2} \cdot C_{10,2} = 20 \cdot 45 \cdot 45 \cdot 45 = 1.822.500.$$

Logo,

$$P(D) = \frac{n(D)}{NT} = \frac{1.822.500}{50.063.860} \approx 0,03640.$$

5. Gabarito E

No gabarito E , os três primeiros números podem ser escolhidos entre seis dezenas, dois em uma das cinco dezenas restantes e o último número entre quatro dezenas. Logo, o número de combinações possíveis é

$$n(E) = C_6^1 \cdot C_{10}^3 \cdot C_5^1 \cdot C_{10}^2 \cdot C_4^1 \cdot 10 = 6 \cdot 120 \cdot 5 \cdot 45 \cdot 4 \cdot 10 = 6.480.000.$$

Logo,

$$P(E) = \frac{n(E)}{NT} = \frac{6.480.000}{50.063.860} \approx 0,12943.$$

6. Gabarito F

Como no gabarito F os três primeiros números podem ser escolhido entre seis dezenas e os três restantes entre cinco dezenas diferentes entre si de dez formas diferentes cada um, obtém-se

$$n(F) = C_6^1 \cdot C_{10}^3 \cdot C_5^3 \cdot 10 \cdot 10 \cdot 10 = 6 \cdot 120 \cdot 5 \cdot 45 \cdot 4 \cdot 10 = 7.200.000.$$

Assim,

$$P(F) = \frac{n(F)}{NT} = \frac{7.200.000}{50.063.860} \approx 0,14382.$$

7. Gabarito G

No gabarito G os três primeiros números podem ser escolhidos entre seis dezenas e os demais em uma outra dezena entre as cinco dezenas restantes. Assim, o número de combinações é dado por

$$n(G) = C_6^2 \cdot C_{10}^3 \cdot C_{10}^3 = 15 \cdot 120 \cdot 120 = 216.000.$$

Portanto,

$$P(G) = \frac{n(G)}{NT} = \frac{216.000}{50.063.860} \approx 0,00431.$$

8. Gabarito H

No gabarito H , os quatro primeiros números podem ser escolhidos entre seis dezenas e os dois restantes em outras duas dezenas diferentes entre si de dez maneiras diferentes cada um. Então,

$$n(H) = C_6^1 \cdot C_{10}^4 \cdot C_5^2 \cdot 10 \cdot 10 = 6 \cdot 210 \cdot 10 \cdot 10 \cdot 10 = 1.260.000.$$

Assim,

$$P(H) = \frac{n(H)}{NT} = \frac{1.260.000}{50.063.860} \approx 0,02517.$$

9. Gabarito I

Como no gabarito I , os quatro primeiros números podem ser escolhidos entre seis dezenas e os demais em uma das cinco dezenas restantes, tem-se

$$n(I) = C_6^1 \cdot C_{10}^4 \cdot C_5^1 \cdot C_{10}^2 = 6 \cdot 210 \cdot 5 \cdot 45 = 283.500.$$

Portanto,

$$P(I) = \frac{n(I)}{NT} = \frac{283.500}{50.063.860} \approx 0,00566.$$

10. Gabarito J

Como no gabarito J temos cinco números em uma mesma dezena que pode ser escolhida entre seis e o número restante entre as cinco restantes de dez formas diferentes, segue que

$$n(J) = C_6^1 \cdot C_{10}^5 \cdot C_5^1 \cdot 10 = 6 \cdot 252 \cdot 5 \cdot 10 = 75.600.$$

Assim,

$$P(J) = \frac{n(J)}{NT} = \frac{75.600}{50.063.860} \approx 0,00151.$$

11. Gabarito K

O gabarito K consiste de seis números pertencentes a uma mesma dezena. Assim,

$$n(K) = C_6^1 \cdot C_{10}^6 = 6 \cdot 210 = 1.260.$$

Logo,

$$P(K) = \frac{n(K)}{NT} = \frac{1.260}{50.063.860} \approx 0,00003.$$

Sob condições gerais, a LFGN garante que a probabilidade de um evento A de interesse por meio da abordagem frequentista, fornecida pela frequência relativa de sucessos obtida em n repetições, se aproxima da probabilidade do evento A fornecida pela abordagem clássica, $P(A)$, à medida que $n \rightarrow \infty$.

Dessa forma, é possível prever os tipos de gabaritos que possuem maior probabilidade de ocorrência em determinados sorteios, por meio da análise dos sorteios realizados até o momento? Sabendo que $P(F) \approx 0,14382$, se a frequência relativa de obtenção do gabarito F encontra-se abaixo de $0,14382$, é mais indicado apostar em um gabarito do tipo F ? O gabarito F , nesse caso, teria maior probabilidade de ocorrência no próximo sorteio?

Muitos leitores e apostadores diriam que sim, dada a Falácia do Jogador, conforme apresentado. Buscando esclarecer teoricamente, por meio da LFGN, o comportamento de ocorrência dos gabaritos apresentados de acordo com as repetições ocorridas, a próxima seção apresenta algumas discussões a respeito de tais questões.

3 Resultados Principais

Inicialmente, a fim de responder às perguntas levantadas, apresentamos três contagens realizadas dos gabaritos considerados neste trabalho, por meio do levantamento de sorteios já realizados da Mega-Sena.

- **1º Momento:** Contagem realizada por Nomelini *et al.* [9].

No artigo em questão, os autores consideraram os primeiros 565 sorteios da Mega-Sena. Na ocasião, foram observadas as seguintes frequências (FO) e as frequências esperadas (FE) para cada gabarito:

Para obtenção da FE do gabarito A , por exemplo, multiplica-se $P(A)$ pelo número de casos considerados. Assim, obtém-se

$$FE = P(A) \cdot 565 = 0,01997 \cdot 565 \approx 11,28559 \approx 11.$$

O mesmo raciocínio é seguido para o cálculo da FE dos demais gabaritos.

Tabela 2: Número de gabaritos conforme contagem realizada por Nomelini *et al.* (2005).

Gabaritos	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>	<i>K</i>
FO	10	164	198	19	72	91	2	7	2	0	0
FE	11	152	206	21	73	81	2	14	3	1	0

- **2º Momento:** Contagem realizada por Araújo [1].

Araújo [1], em sua Monografia de conclusão de curso, adicionou 235 sorteios à contagem realizada por Nomelini *et al.* [9], totalizando 800 sorteios. O autor obteve os seguintes resultados:

Tabela 3: Número de gabaritos conforme contagem realizada por Araújo (2013).

Gabaritos	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>	<i>K</i>
FO	14	239	279	28	105	117	3	12	3	0	0
FE	16	215	292	30	103	115	4	20	4	1	0

- **3º Momento:** Contagem realizada pelas autoras.

Para este artigo, foram analisados 1.420 sorteios além dos 800 considerados no segundo momento, totalizando 2.220 sorteios. O último sorteio incluído na contagem foi o da chamada Mega da virada de 2019. Obteve-se as seguintes frequências:

Tabela 4: Número de gabaritos conforme contagem realizada pelas autoras.

Gabaritos	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>	<i>K</i>
FO	48	621	792	88	271	332	7	46	13	2	0
FE	44	599	808	81	287	319	10	56	13	3	0

Considerando as Tabelas 2-4, o jogador poderia ter um raciocínio parecido com o apresentado na seção anterior para o exemplo do lançamento das moedas quando acredita-se que a probabilidade de obter coroa no 31º lançamento da moeda pode ser maior que 0,5 dada a sequência anterior de 30 obtenções seguidas de caras.

Na Tabela 2, por exemplo, observa-se $FO < FE$ para o gabarito *C*. Alguns jogadores pensariam: "como esperava-se um maior número de gabaritos do tipo *C*, os gabaritos do tipo *C* têm maior probabilidade de saírem no próximo sorteio. Apostarei no tipo *C*".

Outra suposição poderia ser a respeito das frequências observadas para o gabarito *F*: "como nos três momentos considerados, $FE < FO$, a probabilidade de obtenção

do gabarito F em um próximo sorteio é menor. Apostas no gabarito F devem ser evitadas”.

Esses raciocínios caracterizam a Falácia do Jogador cuja ideia é de que desvios no comportamento esperado na realização de um experimento implica em uma maior probabilidade de se obter, em seguida, desvios na direção oposta. Neste sentido, a falácia já pode ser observada na Tabela 4 onde, apesar de ter sido levada em consideração a contagem de outros 1.420 sorteios, a FO ainda não havia atingido a FE do gabarito C .

A Falácia do Jogador ocorre, de certa forma, por uma má interpretação da LFGN que assegura que as frequências observadas se aproximam das frequências esperadas conforme o número de repetições do experimento tende a ∞ . A falácia ocorre quando desconsidera-se que o número de repetições deve ser grande para a estabilização das frequências relativas obtidas e faz-se conclusões para um número pequeno de repetições.

A fim de enfatizar tal falácia, a Tabela 5 pode ser analisada. Utiliza-se a seguinte notação: MOx representa a frequência relativa observada no x -ésimo momento e TE, a frequência relativa teórica obtida pelos cálculos apresentados na seção anterior. Se a LFGN fosse verificada para qualquer número de repetições, as frequências relativas observadas em cada momento se aproximariam de TE, consoante o número de sorteios aumentasse.

Tabela 5: Frequências relativas observadas e esperadas.

	MO1	MO2	MO3	TE
A	0,01770	0,01750	0,02162	0,01997
B	0,29027	0,29875	0,27973	0,26966
C	0,35044	0,34875	0,35676	0,36404
D	0,03363	0,03500	0,03964	0,03640
E	0,12743	0,13125	0,12207	0,12943
F	0,16106	0,14625	0,14955	0,14382
G	0,00354	0,00375	0,00315	0,00431
H	0,01239	0,01500	0,02072	0,02517
I	0,00354	0,00375	0,00586	0,00566
J	0,00000	0,00000	0,00090	0,00151
K	0,00000	0,00000	0,00000	0,00003

Para o gabarito E , por exemplo, observa-se que, embora o número de sorteios analisados tenha sido maior no 2º momento, MO2 se distancia mais (0,00182 a mais) de TE quando comparada a MO1 (0,00200 a menos). Para o gabarito G , observa-se que

a frequência relativa de MO2 é a mais próxima de TE quando comparada com MO3. Dessa forma, observa-se que a crença da Falácia do Jogador não é observada para as contagens realizadas, ou seja, por mais que novos sorteios da Mega-Sena ocorram, a falta de estabilização das frequências relativas observadas para cada gabarito impossibilita uma previsão mais segura do resultado de um determinado sorteio.

Neste sentido, embora na Tabela 4 observa-se $FE > FO$ para o gabarito C , no sorteio de número 2.221 e nos demais sorteios, a probabilidade de obtenção do gabarito C continuou e continuará sendo igual a, aproximadamente, 0,36404.

Por outro lado, a LFGN garante que a frequência relativa observada se aproxima da esperada para um número grande de sorteios. Mas, quão grande deve ser o número de sorteios para que as frequências observadas se aproximem, de fato, das esperadas? Afinal, 2.220 é um número razoavelmente grande de sorteios.

Para uma discussão a respeito, realizamos algumas simulações utilizando a plataforma computacional R [11]. As simulações permitem observar o comportamento das frequências dos gabaritos considerados para um número bem maior de sorteios comparado aos que já foram realizados. Podemos simular, por exemplo, 1.000.000 de sorteios e observar o que ocorre para cada gabarito, embora na prática, o sorteio de número 1.000.000 ocorrerá em um futuro consideravelmente distante, se mantida a frequências dos sorteios.

Para obtenção da estimativa do ano em que ocorrerá o sorteio de número 1.000.000, pode ser considerado o seguinte: um ano tem aproximadamente 365 dias e uma semana tem 7 dias. Assim, a quantidade média de semanas por ano é

$$\frac{365}{7} \approx 52.$$

Para saber quantos sorteios ocorrem por ano, aproximadamente, basta multiplicar a quantidade de semanas existentes em um ano pelo número de sorteios realizados por semana, ou seja, por 2. Logo, $52 \cdot 2 = 104$ sorteios por ano. Quanto ao ano que será realizado o sorteio de número 1.000.000, serão desconsiderados os sorteios de número 1 ao de número 2.220 (realizado em 31/12/2019) e calculada a quantidade de anos necessários para a realização dos 997.780 sorteios restantes, a partir do sorteio de número 2.221, obtendo

$$\frac{997.780}{104} \approx 9.594.$$

Somando essa quantidade de anos com 2.019, ano do último sorteio contabilizado neste trabalho, tem-se

$$9.594 + 2.019 = 11.613.$$

Assim, o sorteio de número 1.000.000 ocorrerá por volta do ano 11.613. Como não é possível aguardar esse ano para levantar as informações e analisar como a frequência de

obtenção de cada gabarito se comportou ao longo do tempo, a simulação dos sorteios é uma importante ferramenta para observamos tal comportamento.

Para a simulação, adotaremos a seguinte estratégia: no modelo probabilístico considerado para a Mega-Sena, o que importa é o grupo de gabaritos no qual um determinado sorteio se classifica. Como são 11 gabaritos possíveis, os sorteios se comparam ao experimento E : retirada de 1 bola de uma urna contendo 2.503.193 bolas, enumeradas pelos naturais de 1 a 11, de tal forma que a correspondência descrita na Tabela 6 será estabelecida. Tal metodologia será adotada na simulação visto que, solicitar ao *software* a ação de reconhecer cada tipo de gabarito, conforme cada resultado simulado, não é algo trivial.

Tabela 6: Probabilidade de obtenção de cada bola, conforme numeração.

Gabarito(Numeração da bola)	Quantidade de bolas	Probabilidade de ocorrência
A (1)	50.000	0,01997
B (2)	675.000	0,26966
C (3)	911.250	0,36404
D (4)	91.125	0,03640
E (5)	324.000	0,12943
F (6)	360.000	0,14382
G (7)	10.800	0,00431
H (8)	63.000	0,02517
I (9)	14.175	0,00566
J (10)	3.780	0,00151
K (11)	63	0,00003
Total	2.503.193	1

Na Tabela 6, embora tenha sido considerado um número de bolas que representa $1/20$ do número total de combinações da Mega-Sena, a fim de simplificar o problema no processo de simulação, observa-se que a probabilidade de ocorrência associada a cada numeração de bola, se mantém igual à probabilidade teórica de obtenção de cada gabarito, cujos valores foram representados na última coluna da Tabela 5. Dessa forma, é matematicamente coerente o uso da estratégia apresentada para realização da simulação.

Adotada essa estratégia, o código em linguagem R para verificar o comportamento do gabarito B , representado pela bola numerada pelo 2, é exibido a seguir para exemplificação da simulação.

A Figura 1 exibe os gráficos obtidos para os gabaritos B , C , F e K , representados pelas bolas numeradas por 2, 3, 6 e 11, respectivamente.

Nos gráficos, a linha vermelha indica a frequência esperada para cada um dos gabaritos considerados.

Sobre a convergência das frequências relativas observadas para a frequência esperada, garantida pela LFGN, observa-se que a convergência é demorada. Embora tenham sido considerados 1.000.000 sorteios, as frequências do gabarito F parecem ser as que mais se aproximam da frequência esperada, apesar de que esse comportamento pode se alterar ao considerar outros sorteios.

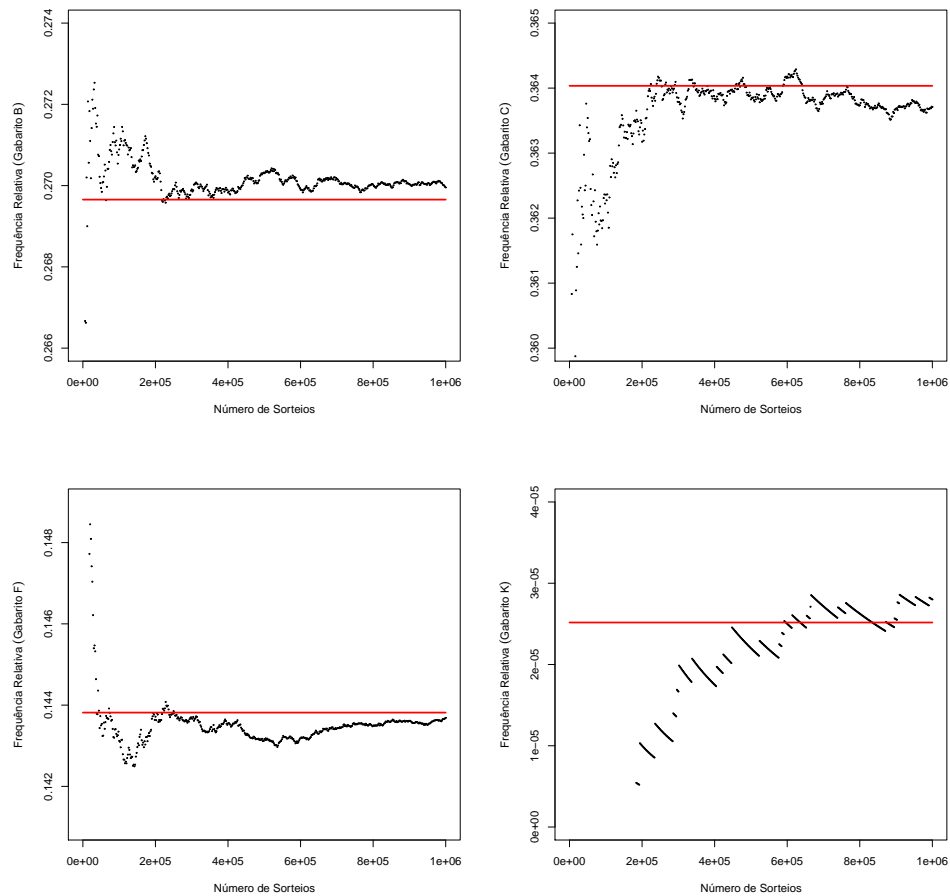


Figura 1: Frequências relativas obtidas para os gabaritos B , C , F e K .

```
set.seed(1) #fixando o experimento
Númerodesorteios=1000000 #número de sorteios
gabaritos <-c(rep(1,50000),rep(2,675000),rep(3,911250),rep(4,91125),
rep(5,324000),rep(6,360000),rep(7,10800),rep(8,63000),rep(9,14175),
rep(10,3780), rep(11,63))
#possíveis gabaritos e suas respectivas quantidades

sorteios<-sample(gabaritos,Númerodesorteios,replace=TRUE)
#realização dos sorteios
table(sorteios) # tabela dos resultados
cont=array(0,c(Númerodesorteios,1))
#criando um contador para o número de obtenções da bola 2
(gabarito B)

for(i in 1:Númerodesorteios)
{
if (sorteios[i]==2) cont[i,1]=1
}
Sn<-cumsum(cont)
n<-1:Númerodesorteios
Média<-rep(0,Númerodesorteios)

for(i in 1:Númerodesorteios)
{
if(n[i] %% 2000==0) Média[i]<-Sn[i]/n[i]
}

for(i in 1:Númerodesorteios)
{
if(Média[i]==0) Média[i]<-NA
}
#frequências relativas acumuladas da bola 2 para cada sorteio

plot(n, Média, type = "p",pch=20, cex = 0.3,xlim=c(0,1000000),
ylim=c(0.266,0.274),xlab="Número de Sorteios",
ylab="Frequência Relativa (Gabarito B)")
segments(1,13500000/50063860, Númerodesorteios,13500000/50063860,
col = "red",lwd =2.5)
#gráfico com os resultados
```


Nota-se que, apesar das frequências relativas observadas estarem próximas das esperadas, principalmente a partir de 200.000 sorteios, o comportamento das frequências observadas é bastante imprevisível.

Sobre a Falácia do Jogador, de fato, os quatro gráficos apontam que quando as frequências observadas aumentam, em algum momento, elas diminuirão, e, vice-versa, buscando uma estabilização em torno da frequência esperada. Porém, não é possível prever o momento em que essa mudança de comportamento ocorrerá. A probabilidade de obtenção de um determinado gabarito é fixa para cada sorteio. Tal fato desmistifica a crença da Falácia do Jogador de que a probabilidade de obtenção de um determinado gabarito em um certa jogada é maior ou menor que a frequência relativa esperada.

De maneira geral, utilizar argumentos lógicos para justificar padrões de comportamentos em fenômenos aleatórios, a fim de garantir certos resultados, é algo bastante ousado e desafiador.

4 Conclusão

Neste artigo, por meio do modelo probabilístico proposto por Nomelini *et al.* [9] para a Mega-Sena, mostrou-se a relação existente entre a LFGN e a Falácia do Jogador. Foi possível identificar a Falácia do Jogador como um exemplo do risco da interpretação errônea de resultados matemáticos como a LFGN e como estratégias baseadas exclusivamente em senso comum, sem uma confirmação matemática, podem levar a resultados equivocados.

Para tal, foi feita uma análise do comportamento do modelo em resultados observados da Mega-Sena e em resultados de sorteios simulados. Neste sentido, além de fornecer um suporte teórico para o entendimento da discussão realizada por meio definições e teoremas, o trabalho também proveu um suporte tecnológico ao leitor, disponibilizando o código utilizado para realização de simulações utilizadas para a obtenção de resultados empíricos, via *software R-project*.

Espera-se que este trabalho possa motivar os leitores envolvidos com o ensino e aprendizagem de probabilidade a utilizarem as ideias apresentadas e a realizarem simulações de outros experimentos no *R-project*.

Referências

- [1] J. C. G. Araújo, *A mega-sena sob uma visão probabilística*. Monografia (Graduação) - Graduação em Licenciatura em Matemática - Universidade Federal do Tocantins, Araguaína, 73f, 2013.
- [2] S. Hazzan, *Fundamentos de Matemática Elementar: Combinatória e Probabilidade*. 7.ed. São Paulo: Atual, 2004.

- [3] G. Casella e R. L Berger, *Inferência Estatística*. São Paulo: Cengage Learning, 2010.
- [4] T. C. Clotfelter e P. J. Cook, The “gambler’s fallacy” in lottery play. *Management Science*, v. 39, n. 12, p. 1521–1525, 1971.
- [5] A. Gadelha, *Uma pequena história da probabilidade*. Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2004. Disponível em: <https://bit.ly/2SshLKb>. Acesso em: 05 out. de 2020.
- [6] G. G. O. Costa, *Curso de Estatística Inferencial e Probabilidades: Teoria e Prática*. São Paulo: Atlas, 2012.
- [7] L. Leithold, *O cálculo com geometria analítica, vol. 1*. Rio de Janeiro: Editora Harbra Ltda, 1994.
- [8] J. Lyons, D. J. Weekes e D. Elliott, The gambler’s fallacy: a basic inhibitory process?. *Frontiers in Psychology*, v. 4, n. 72, p. 1–11, 2013.
- [9] Q. S. S. Nomelini, W. C. J. Jesus, M. A. S. SANTOS, E. G. Silva, H. D. Silva, e A. J. S. Júnior, Modelo Probabilístico para o Sorteio da Mega-Sena. In: XIV Congresso dos Pós-Graduandos da Universidade Federal de Lavras. Lavras, 2005. v. 1.
- [10] 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.
- [11] R Core Team (2018), *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <https://www.R-project.org/>.
- [12] A. Tversky, e D. Kahneman, Belief in the law of small numbers. *American Psychological Association*, v. 76, n. 2, p. 105–110, 1971.