Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 12, n. 1, 2026.

Trabalho apresentado no XLIV CNMAC, Fundação Getulio Vargas - Rio de Janeiro - RJ, 2025.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Brincando com Distribuição de Benford

Henrique N. Silva[1] Eduardo Guéron[2]
UFABC, Santo André, SP

O astrônomo Simon Newcomb e o físico (engenheiro elétrico) Frank Benford perceberam, com uma diferença de quase 60 anos, que folhas das antigas tábuas de logaritmos possuíam suas páginas associadas a números iniciados em 1 bem mais desgastadas do que a maioria. Supuseram, então, que alguns conjuntos de dados de certos fenômenos observados, por uma razão ainda não muito clara para ambos, seguiam uma distribuição para seus primeiros dígitos, implicando maior frequência de números iniciados em 1 em relação a números iniciados em 8 ou 9. Essa distribuição ficou conhecida como Lei de Newcomb-Benford (NB)[9, 1].

Com o tempo, observou-se que um número enorme de conjuntos de dados segue NB, dando um caráter meio 'mágico' à distribuição [6]. A partir principalmente das análises de Mark Nigrini [10], NB passou a ser aplicada na detecção de fraudes contábeis, entre outras. Recentemente, a partir da invariância por mudança de base em NB [4], começamos a estudar possíveis fraudes (eleitorais) utilizando NB em outras bases numéricas [2, 8, 3].

$$P_b(d) = \log_b\left(1 + \frac{1}{d}\right). \tag{1}$$

A equação acima descreve a frequência esperada (probabilidade) $P(d)$ de números iniciados com o dígito d em um conjunto de dados escritos na base b que segue a distribuição de Newcomb-Benford (NB).

Neste trabalho, observamos a emergência da distribuição de Newcomb-Benford a partir de simulações simples em sorteios de um conjunto discreto de números, em outras palavras, a partir de um dado de n lados com uma distribuição de probabilidade qualquer, conseguimos chegar a NB.

A ideia básica é sortear um conjunto de dados que comece com dígito 1 a $b-1$ escritos na base b. Cada dado é obtido a partir do produto de n sorteios de modo que o resultado final será sempre dividido por b^j garantindo que o resultado esteja no intervalo $I = [1, b-1)$.

A seguir, um exemplo prático na base 3.

Como primeiro exemplo, fizemos o sorteio de dois números, 1 ou 2 com respectivas probabilidades p e $1-p$, algo como um cara e coroa. Esse sorteio se repete N vezes e, ao número obtido, multiplica-se o anterior. Cada vez que a soma ultrapassa 3, dividimos o resultado por 3. O procedimento pode ser entendido como (partindo de $R = 1$):

1. Sorteia-se o k-ésimo número $x(k)$; **2.** Multiplica-se ao resultado acumulado $R = R * x(k)$; **3.** Se $R \geq 3$ então $R = R/3$ - garante $R \in [1, 3)$; **4.** Retorna para $k+1$ enquanto $k < n$; **5.** Quando $k = n$, escreve-se R final e extrai-se o primeiro dígito;

Obtemos, com o procedimento acima, um número do nosso conjunto de dados partindo das condições iniciais $p(x(k) = 1) = 0,2$ e $p(x(k) = 2) = 0,8$. A rotina foi repetida 750 vezes e os dados plotados na Fig.1. A abscissa corresponde ao número n do procedimento descrito. As linhas tracejadas correspondem às proporções esperadas segundo Eq.1. Percebe-se, ainda, que quase todos os dados estão dentro de duas vezes o desvio padrão, representado pelas linhas vermelhas pontilhadas.

[1]h.nascimento@ufabc.edu.br
[2]eduardo.gueron@ufabc.edu.br

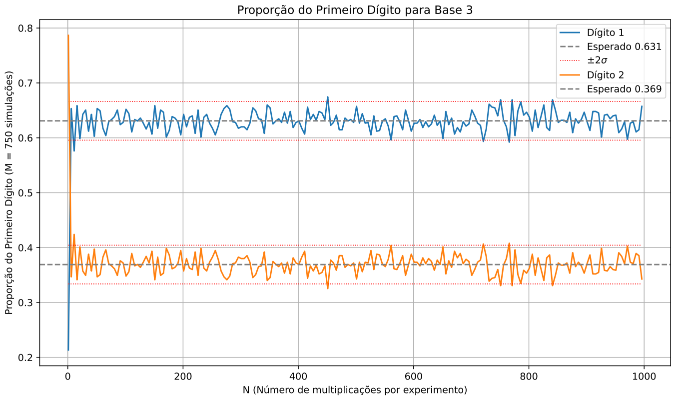

Figura 1: Evolução da Proporção do Dígito 1 na base 3. Fonte: Original dos autores.

Tal processo foi repetido em outras bases com a regra de sortear sempre os números $1, \cdots, b-1$ e estudar o problema na base b. Para estudar um dado de 6 lados fazemos, naturalmente, o procedimento acima na base 7. Resultados obtidos em outras bases foram incluídos no poster apresentado no CNMAC - 2025. A simulação como um todo representa, de maneira lúdica, a emergência da NB conforme previsto no chamado Teorema do Limite Central - Módulo 1 [7, 5].

Referências

[1] F. Benford. "The law of anomalous numbers". Em: **Proceedings of the American philosophical society** (1938), pp. 551–572.

[2] E. Gueron e J. Pellegrini. "Application of Benford–Newcomb law with base change to electoral fraud detection". Em: **Physica A: Statistical Mechanics and its Applications** 607, 128208 (2022). DOI: `10.1016/j.physa.2022.128208`.

[3] E. Gueron, J. Pellegrini e B. Aristimunha. "Distribuição de Newcomb-Benford, Mudanças de Bases e Aplicações Eleitorais". Em: **Proceeding Series of the Brazilian Society of Computational and Applied Mathematics**. 2023. DOI: `https://doi.org/10.5540/03.2023.010.01.0042`.

[4] T. P. Hill. "Base-invariance implies Benford's law". Em: **Proceedings of the American Mathematical Society** 123.3 (1995), pp. 887–895.

[5] T. P. Hill. "The significant-digit phenomenon". Em: **The American Mathematical Monthly** 102.4 (1995), pp. 322–327.

[6] W. Hurlimann. "Benford's Law from 1881 to 2006: a bibliography". preprint arXiv:math/0607168. 2006. URL: `https://arxiv.org/abs/math/0607168`.

[7] S. Miller e M. Nigrini. "The modulo 1 central limit theorem and Benford's law for products". Em: **International Journal of Algebra** 2.13-16 (2008), pp. 615–631.

[8] H. N. Silva. "Uso de Invariância de Base em Aplicações da Lei de Newcomb-Benford". Dissertação de mestrado. Universidade Federal do ABC, 2023.

[9] S. Newcomb. "Note on the frequency of use of the different digits in natural numbers". Em: **American Journal of mathematics** 4.1 (1881), pp. 39–40.

[10] M. Nigrini e J. Wells. **Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection**. John Wiley & Sons, 2012. ISBN: 978-1-118-15285-0.