Lei de Benford e aplicações

Alexandre Cano Teixeira - alexandre.teixeira@usp.br

Prof.ª Dra. Elisabeti Kira - betikira@ime.usp.br

Instituto de Matemática e Estatística – USP

Objetivo

Esse trabalho tem por objetivo introduzir, de forma intuitiva, o que se conhece como a "*Lei de Benford*" para a distribuição de probabilidade do primeiro dígito de números de um conjunto de dados.

Ilustraremos a lei para alguns conjuntos de dados como, por exemplo, a sequência dos números de Fibonacci, sequência dos números primos e para dados da CVM[4].

Introdução

Quando solicitamos a uma pessoa que diga qual é a probabilidade associada para o primeiro dígito de um conjunto de números, em geral, ela deve fornecer intuitivamente o valor 1/9, ou seja, a mesma probabilidade de ocorrência para cada um dos nove dígitos.

Newcomb[1] foi o primeiro a perceber que essa uniformidade não era válida em várias situações e concebeu uma distribuição de probabilidade para o primeiro dígito de números obtidos de várias fontes. Benford[2] mostrou formalmente a lei, e encontramos resultados e propriedades mais gerais da lei em Hill[3].

Consideremos o conjunto de dados fornecidos pela sequência dos números de Fibonacci: 1, 2, 3, 5, 8, 13, 21, …. Tomamos o primeiro dígito de cada número, considerado da esquerda para a direita, e contamos a sua frequência de aparição. Por exemplo, para os números acima mencionados verificamos que o primeiro dígito desses números são, respectivamente, 1, 2, 3, 5, 8, 1, 2, … . Considerando os 300 primeiros números de Fibonacci temos a frequência relativa para o primeiro dígito dada pela Tabela 1:

Primeiro dígito	1	2	3	4	5	6	7	8	9	Total
Porcentagem (%)	30,00	17,67	13,00	9,00	8,33	6,33	5,67	5,67	4,33	100,00

Tabela 1: Tabela de frequência relativa para o primeiro dígito dos números da sequência de Fibonacci.

Notamos que para essa amostra dos números de Fibonacci a distribuição do primeiro dígito está longe de ser uniforme.

A lei

Como proposto por Newcomb, o primeiro dígito dos dados acima segue uma distribuição logarítmica. O modelo teórico foi demonstrado por Benford[2], onde a probabilidade de ocorrência do dígito *d*, é na forma do logaritmo:

$$P(d)=\log\left(1+\frac{1}{d}\right),\quad d=1,2,\dots,9 \qquad (1)$$

A Figura 1 ilustra a Lei de Benford juntamente com os dados da Tabela 1:

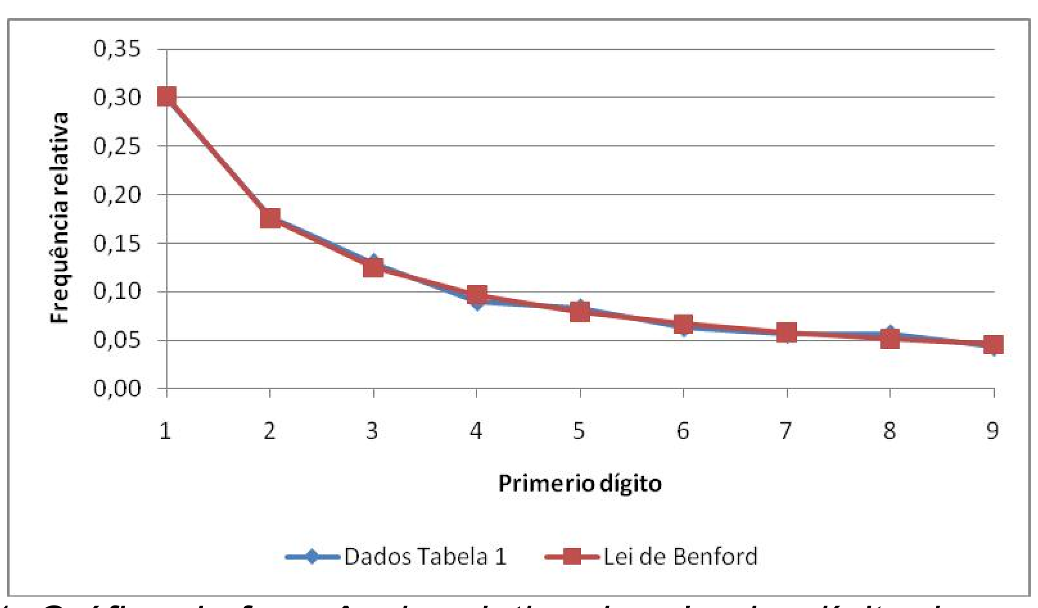

Figura 1: Gráfico da frequência relativa do primeiro dígito da sequência dos números de Fibonacci e a Lei de Benford.

Observamos que, para os dados de Fibonacci, descritivamente, o ajuste do modelo foi bom. Esse resultado foi validado pelo teste Qui-Quadrado de aderência (a um nível de significância de 5%).

Outra fonte de dados que consideramos são os valores (em reais) de 1.274 notas fiscais oriundos da CVM[4]. Contamos a frequência relativa do primeiro dígito do valor encontrado em cada uma dessas notas. Os resultados indicam que a Lei de Benford se ajustou bem aos dados da CVM conforme ilustra o gráfico na Figura 2:

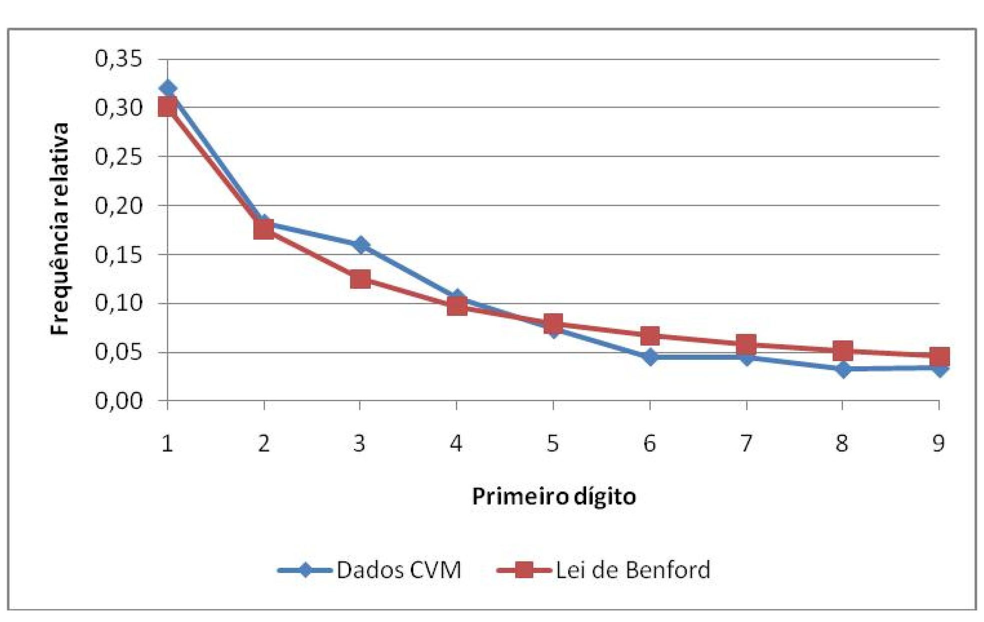

Figura 2: Gráfico da frequência relativa do primeiro dígito dos dados da CVM.

Outros dados considerados foram a sequência dos números primos, onde ilustramos na Figura 3 a frequência relativa do primeiro dígito (até 100.000 há 9.593 primos):

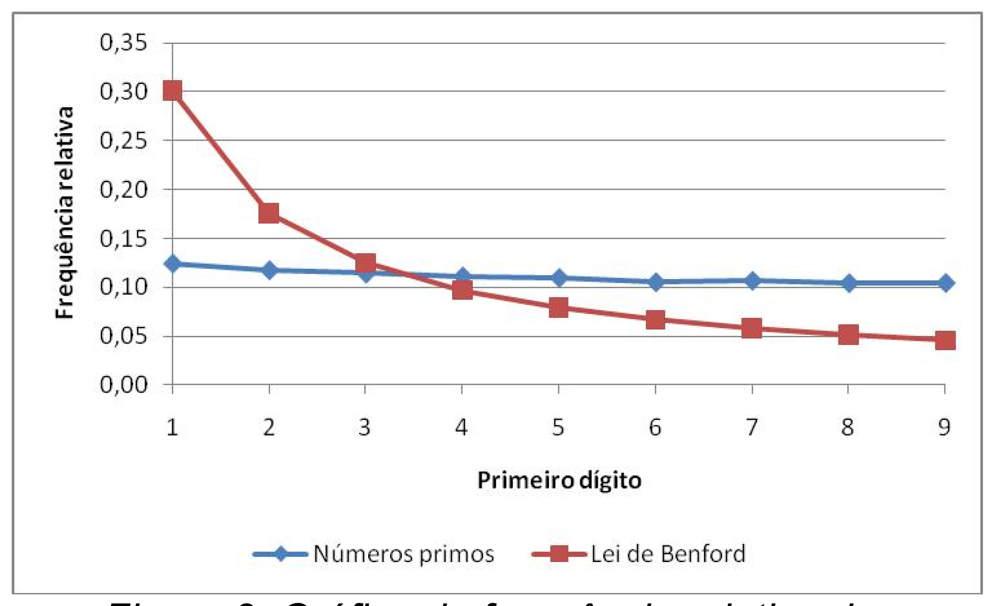

Figura 3: Gráfico da frequência relativa do primeiro dígito da sequência dos números primos.

Neste exemplo podemos verificar que o modelo de Benford não se ajustou aos dados experimentais obtidos da sequência dos números primos.

Assim, observamos que a distribuição de probabilidade para o primeiro dígito pode ser verificada em alguns tipos de dados como, por exemplo, os números da sequência de Fibonacci e as notas fiscais da CVM que seguem a Lei de Benford. Entretanto, em alguns dados como os números primos, a lei não se mostrou adequada para a distribuição do primeiro dígito.

Hill em seu artigo considerou uma generalização da lei dos dígitos não só para o primeiro dígito mas para os demais dígitos. Por exemplo, formulando a lei para o segundo dígito temos que denotando d_2 o valor do segundo dígito:

$$P(d_2=k)=\sum_{d_1=1}^{9}\log\left(1+\frac{1}{\underline{d_1d_2}}\right),\quad k=0,1,2,\dots,9 \qquad (2)$$

Com $\underline{d_1d_2}$ denotando o número cujo primeiro dígito é d_1 e segundo dígito igual a d_2.

Por exemplo, para *k*=3,

$$P(d_2=3)=\log\left(1+\frac{1}{13}\right)+\log\left(1+\frac{1}{23}\right)+\dots+\log\left(1+\frac{1}{93}\right), \qquad (3)$$

Neste momento, temos a inclusão do algarismo 0 (zero) como mais uma categoria que pode ser assumida pelo segundo dígito.

Verificamos então a distribuição da frequência relativa do segundo dígito dos números da sequência de Fibonacci onde a contagem inicia-se à partir do sexto número (13) para que possa haver segundo dígito, temos a Tabela 2:

Segundo dígito	0	1	2	3	4	5	6	7	8	9	Total
Porcentagem (%)	11,67	12,00	9,00	11,00	9,00	9,00	8,00	11,00	9,33	10,00	100,00

Tabela 2: Tabela de frequência para o segundo dígito dos números da sequência de Fibonacci.

A Figura 4 mostra que o modelo de Benford gerou uma curva quase uniforme o que não se adaptou bem aos dados do segundo dígito dos números de Fibonacci:

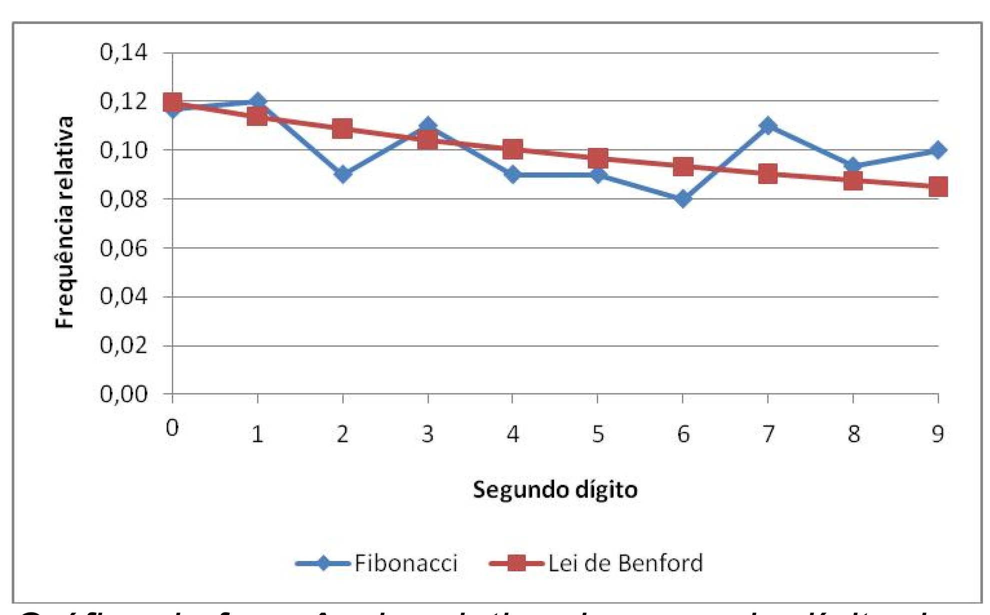

Figura 4: Gráfico da frequência relativa do segundo dígito da sequência dos números de Fibonacci e a Lei de Benford.

Conclusões

Nem sempre a distribuição uniforme é a que melhor se ajusta para descrever o primeiro dígito de um conjunto de dados. A Lei de Benford, com forma logarítmica, é aplicável em algumas situações e mostra que existe um privilégio para o dígito 1 frente aos demais dígitos.

Hill também considerou a lei estendida para o segundo, terceiro e demais dígitos de números que "tende" a distribuição uniforme.

Referências Bibliográficas

[1] Newcomb, Simon, "Note on the frequency of use of the different digits in natural numbers", *Amer. Journ. of Math.* 4, p.39-40, 1881.

[2] Benford, Frank, "The law of anomalous numbers", *Proceedings of The American Philosophical Society*, vol. 78, p.551-572, 1938.

[3] Hill, T. P., "The Significant-Digit Phenomenon", *The American Mathematical Montly*, Vol. 102, nº 4, p.322-327, 1995.

[4] CVM - Comissão de Valores Mobiliários, www.cvm.gov.br (Acesso em março/2009).