

UNIVERSIDADE FEDERAL DE SANTA CATARINA

CAROLINA SAVI SERAFIM FALCÃO

**A LEI DE BENFORD PARA A DISTRIBUIÇÃO DOS PRIMEIROS
DÍGITOS**

FLORIANÓPOLIS

2014

CAROLINA SAVI SERAFIM FALCÃO

**A LEI DE BENFORD PARA A DISTRIBUIÇÃO DOS PRIMEIROS
DÍGITOS**

Trabalho de Conclusão de Curso submetido à Universidade Federal de Santa Catarina como parte dos requisitos necessários para a obtenção do Grau de Licenciada em Matemática.

Orientador: Prof. Marcelo Sobottka

FLORIANÓPOLIS

2014

A Deus

À minha filha Luiza

Ao meu marido Christian

Agradecimentos

A Deus, por ter me capacitado e me dado forças todos os dias.

A esta Universidade e seu corpo docente que, desde o primeiro dia, fez com que eu tivesse certeza de que este era o curso que eu me formaria.

Ao meu orientador, professor Dr. Marcelo Sobottka, pelo suporte, paciência e pelo tema deste trabalho.

Ao meu amigo e professor Aldrovando, pelos dias e dias dedicados a me fazer compreender cada mínima coisa e que, sem o qual, este trabalho não teria sido concluído.

À minha filha, Luiza, que mesmo tão pequena compreendeu a ausência da mamãe em casa durante as noites. É por você que estou aqui hoje. Te amo!

Aos meus pais e meu irmão. Obrigada pela paciência, pelo incentivo, pela força e principalmente pelo carinho. Valeu a pena todo o sofrimento, todas as renúncias... Valeu a pena esperar! Hoje estamos colhendo, juntos, os frutos desse empenho! Esta vitória não é só minha, é nossa!!!

Ao meu marido, Christian. Meu melhor amigo, pessoa com quem amo partilhar a vida. Com você, até as coisas mais difíceis se tornam leves e fáceis. Obrigado pelo amor, paciência e pela sua capacidade de me trazer paz na correria de cada semestre. Você acreditou em mim até quando eu mesma não acreditei e isso me fez ter força para continuar.

Aos meus amigos que me incentivaram, empurraram, apoiaram e, por muitas vezes, me fizeram esquecer essa correria que é a vida acadêmica e me fizeram rir. Vocês são parte fundamental na minha história, perto de vocês o mau humou não existe e os tempos difíceis passam mais rápido. Amo-os!

Sumário

1	Introdução	6
2	Teoria da probabilidade	8
2.1	Experimento aleatório	8
2.2	Eventos	9
2.3	Probabilidade	9
2.4	Probabilidade Condicional	11
2.5	Probabilidade Total	12
3	Variáveis aleatórias e distribuições de probabilidade	14
3.1	Variável Aleatória	14
3.2	Distribuição de Probabilidade	14
4	Teoria da medida	17
4.1	σ -álgebras	17
4.2	Espaços com medida	19
4.3	Propriedades elementares da medida	20
4.4	Medida de Lebesgue em \mathbb{R}	22
4.5	Medida de Probabilidade	23
5	A Lei de Benford	24
5.1	A Lei dos Primeiros Dígitos	24
5.2	A σ -álgebra \mathcal{M}	26

5.3	Invariância na base e invariância na escala	27
5.4	Amostras aleatórias de distribuições aleatórias	28
5.5	Lei do limite logarítmico para dígitos significativos	30
6	Conclusão	33
	Referências	34

1 *Introdução*

Quando pensamos na distribuição dos primeiros dígitos de uma série de números, intuitivamente concluímos que a probabilidade de ocorrência do dígito 1 é igual à do dígito 9, ou de qualquer outro dígito. Por serem 9 possíveis dígitos, espera-se que a probabilidade da ocorrência de cada um é a mesma, a saber, $\frac{1}{9}$. Porém, é verificado que em muitas situações isso não ocorre, e este é o tema central deste TCC.

O primeiro a perceber indícios da diferença de probabilidade de ocorrência dos algarismos nos primeiros dígitos de um número foi o matemático Simon Newcomb, em 1881, ao observar as páginas das tábuas de logaritmos (ver Capítulo 5). Muitos anos mais tarde, em 1938, o físico Frank Benford foi quem formalizou a lei dos primeiros dígitos. Porém, apenas em 1995 é que foi publicada uma demonstração rigorosa desta lei.

A Lei de Benford afirma que se você coletar números de forma aleatória e calcular as frequências de seus primeiros dígitos significativos, então os números com primeiro dígito significativo igual a 1 aparecerão cerca de 30% das vezes, enquanto que o 9, por exemplo, aparecerá somente 4.5% das vezes. Esta regra é observada em muitos conjuntos de números, como potências de 2 e os números de Fibonacci. O primeiro dígito significativo de um número positivo é o dígito não nulo mais à esquerda em sua representação decimal. Por exemplo, o primeiro dígito significativo de π é 3, o de 2371,5 é 2 e o de 0,00563 é 5. Uma outra maneira de defini-lo é escrever cada número real positivo x como um número $m \in [1, 9)$ multiplicado por uma potência de 10:

$$x = m10^n, \quad n \in \mathbb{Z}.$$

Nesse contexto, o primeiro dígito significativo de x é a parte inteira de m , que será denotada por $[m]$. O número m é denominado mantissa de x . Afirmamos que se você coletar números de forma aleatória e calcular a frequência $D(d)$ do primeiro dígito significativo d , então o valor de $D(d)$ será dado aproximadamente por $\log_{10}(1 + \frac{1}{d})$. Isto nos dá a seguinte tabela de frequências:

d	1	2	3	4	5	6	7	8	9
D(d)	0.3010	0.1761	0.1249	0.0969	0.0792	0.0669	0.0580	0.0511	0.0458

A Lei de Benford é fascinante: ela desafia nossa intuição e é algo que poderia ser adaptada para uma excelente atividade em sala de aula. Pode parecer uma mera curiosidade, mas ela é atualmente usada como uma ferramenta para detectar fraudes. Mas, o primeiro dígito significativo não é a única coisa com a qual se é preciso preocupar. A Lei de Benford Generalizada permite obter uma lei para o segundo dígito significativo, para o terceiro dígito significativo, etc.

2 Teoria da probabilidade

Em nosso cotidiano é comum nos depararmos com situações que não sabemos, efetivamente, qual será seu desfecho. Contudo, podemos prever a probabilidade de ocorrência de alguns eventos com base nos seus possíveis resultados. O objetivo da teoria da probabilidade é estudar estes fenômenos ocorridos ao acaso, isto é, aqueles cujos resultados não podem ser previstos com exatidão. Neste capítulo faremos uma introdução à teoria da probabilidade, visto que a Lei de Benford trata diretamente deste assunto.

2.1 Experimento aleatório

Definição 1. *Um experimento aleatório é um experimento que pode ser repetido tantas vezes quanto desejarmos e cujos resultados não podem ser previstos. O conjunto de todos os resultados de um experimento aleatório é chamado de espaço amostral, designado por Ω .*

Definição 2. *Espaços amostrais podem ser discretos ou contínuos.*

- I. *Espaços amostrais discretos podem ter um número finito ou infinito de possíveis resultados, desde que estes resultados possam ser colocados em uma relação de um para um com o conjunto dos números inteiros positivos.*
- II. *Espaços amostrais contínuos são os que contêm um ou mais intervalos.*

Exemplo 1.

- I. *Se um dado honesto é lançado e o número que aparece é indicado, então $\Omega = \{1, 2, \dots, 6\}$. Porém, se rolarmos este dado até um “6” ser obtido e anotarmos o número de vezes que o dado foi rolado antes do primeiro “6”, então $\Omega = \{0, 1, 2, \dots\}$*
- II. *Se um dado é lançado e marcamos o tempo até ser obtido um “6”, então temos que $\Omega = (0, \infty)$*

2.2 Eventos

Definição 3. Um evento é um conjunto de resultados. Isto é, um subconjunto do espaço amostral Ω . Em particular, cada resultado é um evento, assim como o espaço amostral em si.

Observação 1.

- I. Um resultado também pode ser chamado de evento simples. Porém, um evento composto é formado por pelo menos dois resultados.
- II. Para sermos precisos, devemos distinguir entre o resultado ω , que é um elemento de Ω e o evento $\{\omega\} \subset \Omega$.

Definição 4. Dois eventos, A e B , são ditos incompatíveis (ou mutuamente exclusivos) se sua intersecção é vazia, isto é, se $A \cap B = \emptyset$.

Exemplo 2. No lançamento de um dado, os eventos $A = \{n \leq 2\}$ e $B = \{n \text{ múltiplo de } 3\}$ são incompatíveis, ou seja, os dois não podem ocorrer ao mesmo tempo pois a ocorrência de um exclui a ocorrência do outro.

2.3 Probabilidade

Definição 5. A probabilidade de um evento $A \subset \Omega$, denotado por $P[A]$, é um número real obtido quando aplicamos em A uma função P que possui as seguintes propriedades:

1. $0 \leq P[A] \leq 1$
2. Se $A = \Omega$, então $P[A] = 1$
3. Se $A = \bigcup_{n \in \mathbb{N}} A_n$, onde $\{A_n\}_{n \in \mathbb{N}}$ é uma família de eventos incompatíveis, então temos que

$$P[A] = \sum_{i=1}^n P[A_i]$$

Proposição 1.

- I. Seja A' o evento complementar de A , temos que

$$P[A] + P[A'] = 1$$

o que significa que A' ocorre se, e somente se, A não ocorre.

- II. $P[\emptyset] = 0$

III. $P[A] \leq P[B]$ se $A \subset B$

Prova.

I. Sabemos que $A \cap A' = \emptyset$ e $A \cup A' = \Omega$, então $P(A \cup A') = P(\Omega) = 1$. Portanto, $P(A) + P(A') = 1$

II. Observe que $\Omega \cap \emptyset = \emptyset$. Logo pelo item I. tem-se

$$1 + P(\emptyset) = P(\omega) + P(\emptyset) = p(\Omega) = 1 \Rightarrow P(\emptyset) = 0$$

III. Temos que $A \subseteq B$, então $B = A \cup (B \setminus A)$, mas $A \cap (B \setminus A) = \emptyset$. Portanto, $P(B) = P(A) + P(B \setminus A) \geq P(A)$

Observação 2.

I. Uma maneira “ingênua” de definir a probabilidade de um evento é usando a frequência relativa com que esse evento ocorre. A frequência relativa do evento $\{k\}$ é a quantidade $f_{\{k\}}(n)$ definida por

$$f_{\{k\}}(n) = \frac{N_{\{k\}}(n)}{n}$$

onde $N_{\{k\}}(n)$ é o número de vezes que o resultado k ocorre em n experimentos. Portanto, a probabilidade de um evento fundamental $\{k\}$ pode ser obtida tomando o limite de $f_{\{k\}}(n)$ com n tendendo a infinito:

$$P[\{k\}] = \lim_{n \rightarrow \infty} f_{\{k\}}(n)$$

Mais formalmente, a probabilidade é definida a partir da teoria de medida (ver Capítulo 4), e demonstra-se a relação acima entre a frequência relativa e a probabilidade do evento através da Lei dos Grandes Números, que declara: Se um evento de probabilidade p é observado repetidamente em ocasiões independentes, a proporção da frequência observada deste evento em relação ao total número de repetições converge em direção a p à medida que o número de repetições se torna arbitrariamente grande [8]. Isso quer dizer que se uma experiência é repetida várias vezes, a probabilidade observada nas repetições se aproxima da probabilidade calculada.

II. Algumas vezes, quando o espaço amostral é finito, a probabilidade de um resultado é simplesmente igual a 1 dividido pelo número total de resultados. Neste caso, os resultados são ditos equiprováveis, ou seja, têm a mesma probabilidade de ocorrer. Por exemplo, quando rolamos um dado não viciado temos que $P[\{1\}] = P[\{2\}] = \dots = P[\{6\}] = 1/6$.

Se os r_i resultados não forem equiprováveis, podemos usar a seguinte fórmula:

$$P[A] = \sum_{r_i \in A} P[\{r_i\}]$$

Entretanto, esta fórmula só tem aplicação quando conhecemos a probabilidade de todos os resultados r_i que constituem o evento A .

2.4 Probabilidade Condicional

Definição 6. A probabilidade condicional de um evento A , dado que um evento B ocorra, é definida por:

$$P[A|B] = \frac{P[A \cap B]}{P[B]} \quad \text{se } P[B] > 0$$

Da definição acima, obtemos a regra da multiplicação:

$$P[A \cap B] = P[A|B]P[B] \quad \text{se } P[B] > 0$$

Definição 7. Sejam A e B dois eventos tais que $P[A]P[B] > 0$. Dizemos que A e B são eventos independentes se

$$P[A|B] = P[A]$$

Deduzimos, a partir da regra da multiplicação, que A e B são independentes se, e somente se,

$$P[A \cap B] = P[A]P[B]$$

Na verdade, esta equação é a definição geral de eventos independentes A e B , onde podemos ter $P[A]P[B] = 0$. Em geral, os eventos A_1, A_2, \dots, A_n são independentes se, e somente se,

$$P \left[\bigcap_{j=1}^k A_{i_j} \right] = \prod_{j=1}^k P[A_{i_j}]$$

para qualquer escolha de subíndices $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$

Observação 3. Se A e B são eventos incompatíveis, então eles não serão independentes, a menos que $P[A]P[B] = 0$. Portanto, no caso em que $P[A]P[B] > 0$, temos:

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{P[\emptyset]}{P[B]} = \frac{0}{P[B]} = 0 \neq P[A]$$

2.5 Probabilidade Total

Sejam B_1, B_2, \dots, B_n eventos incompatíveis. Isto é:

$$B_i \cap B_j = \emptyset \text{ se } i \neq j \text{ e } \bigcup_{i=1}^n B_i = \Omega.$$

Dizemos que os eventos B_i constituem uma partição do espaço amostral Ω . Segue que

$$P\left[\bigcup_{i=1}^n B_i\right] = \sum_{i=1}^n P[B_i] = P[\Omega] = 1$$

Teorema 1 (Teorema da Probabilidade Total). *Seja B_1, B_2, \dots, B_n uma partição de Ω com $P(B_i) > 0$ para $i = 1, 2, \dots, n$. Seja A um evento qualquer. Então*

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i).$$

Prova. Como B_1, B_2, \dots, B_n é uma partição de Ω temos

$$A = \bigcup_{i=1}^n A \cap B_i \quad \text{onde esta união é disjunta.}$$

Logo

$$P(A) = \sum_{i=1}^n P(A \cap B_i),$$

e assim,

$$P[A] = \sum_{i=1}^n P[A \cap B_i] = \sum_{i=1}^n P[A|B_i]P[B_i]$$

Teorema 2. *Seja A um evento tal que $P(A) > 0$. Sejam B_1, B_2, \dots, B_n uma partição de Ω com $P(B_i) > 0$ para todo $i = 1, 2, \dots, n$. Então*

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^n P(A|B_i)P(B_i)} \quad \text{para todo } j = 1, 2, \dots, n.$$

Esta fórmula é chamada de Fórmula de Bayes.

Prova. Segue da definição de probabilidade condicional que

$$P(B_j|A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(A|B_j)P(B_j)}{P(A)}.$$

Pelo teorema da probabilidade total, obtemos

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i).$$

Portanto

$$P(B_j|A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(A|B_j)P(B_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^n P(A|B_i)P(B_i)}.$$

3 Variáveis aleatórias e distribuições de probabilidade

3.1 Variável Aleatória

Definição 8. Uma variável aleatória é uma função $X : \Omega \rightarrow \mathbb{R}$ que associa um número real a cada elemento do espaço amostral e para cada intervalo $J \subset \mathbb{R}$ se tem que $X^{-1}(J)$ é um evento.

Exemplo 3. Considere o experimento que corresponde a jogarmos uma moeda três vezes. Um espaço amostral associado é

$$\Omega = \{(ccc), (kcc), (ckc), (cck), (kkc), (kck), (ckk), (kkk)\}. \quad \text{Onde } c \text{ é coroa e } k \text{ é cara.}$$

Seja a variável aleatória X que expressa o número de coroas obtido. Então os valores assumidos pela variável aleatória são:

$$X(kkk) = 0$$

$$X(kkc) = X(kck) = X(ckk) = 1$$

$$X(kcc) = X(ckc) = X(cck) = 2$$

$$X(ccc) = 3$$

3.2 Distribuição de Probabilidade

No exemplo acima, a variável X pode assumir valores no conjunto $\{0, 1, 2, 3\}$. Considere então os seguintes subconjuntos de Ω : para cada $x \in \{0, 1, 2, 3\}$, seja $X^{-1}(\{x\})$. Sendo este um subconjunto de Ω temos uma probabilidade atribuída a ele. A saber

$$P(X^{-1}(\{x\}))$$

que se denota por $P(X = x)$. Assim a cada número $x \in \{0, 1, 2, 3\}$ estamos atribuindo uma probabilidade. A função que assim resulta denomina-se de função densidade de probabilidade da variável X .

Exemplo 4. *Seja um experimento cujos possíveis resultados são SUCESSO ou FRACASSO e que as probabilidades correspondentes são p, q respectivamente. Agora, repetimos este experimento n vezes. Seja a variável aleatória que expressa o número de SUCESSOS obtidos nas n repetições, vamos determinar a densidade de probabilidade associada a esta variável. Os resultados possíveis desta variável são $\{0, 1, 2, \dots, n\}$. A densidade de probabilidade é a função*

$$P(k) = P(X = k) = \binom{n}{k} p^k q^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}.$$

Observe que se somarmos todos os valores da densidade obtemos

$$\sum_{k=0}^n P(k) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1^n = 1$$

Uma variável cuja densidade de probabilidade é do tipo acima se chama de variável binomial.

Definição 9. *Dada uma variável aleatória $X : \Omega \rightarrow \mathbb{R}$, uma função densidade para X é uma função não negativa $f_X : \mathbb{R} \rightarrow [0, 1]$ que relaciona cada valor de X com a sua probabilidade de ocorrência. Se a variável aleatória é discreta, então a função densidade associada a ela também será discreta e, para cada $x \in \mathbb{R}$, o valor $f_X(x)$ representa exatamente a probabilidade do evento $\{X = x\}$. Se a variável aleatória for contínua, então a probabilidade do evento $\{X = x\}$ pode ser nula para todo $x \in \mathbb{R}$. Nesse caso, a função densidade é dita contínua e temos $P[a \leq X \leq b] = \int_a^b f_X(x) dx$*

Definição 10. *Dada uma variável aleatória $X : \Omega \rightarrow \mathbb{R}$, uma **distribuição de probabilidade** é uma função $F_X : \mathbb{R} \rightarrow [0, 1]$ que associa cada valor $x \in \mathbb{R}$ à probabilidade de que a variável aleatória X assuma um valor inferior ou igual a x . No caso de uma variável aleatória discreta, tomando valores em \mathbb{Z} temos:*

$$F_X(x) = \sum_{i=1}^{\lfloor x \rfloor} f_X(X = i),$$

Para uma variável aleatória contínua temos:

$$F_X(x) = \int_{-\infty}^x f_X(s) ds$$

Definição 11. *Se X é uma variável aleatória com função distribuição F_X e densidade f_x , então*

a sua *esperança* é dada por:

$$E(x) = \int_{\Omega} X dP = \int_{\Omega} X dF_X = \int_{\Omega} X(x) f_X(x) dx$$

Observação 4. A integral acima corresponde à integral de Lebesgue. Para mais detalhes a respeito da integral de Lebesgue, ver [5, Capítulo 4]

4 Teoria da medida

Uma medida em um conjunto X é uma função que atribui um número real positivo para subconjuntos de X . Ela pode ser interpretada como área, comprimento, massa, volume, ou qualquer grandeza aditiva, isto é, a medida da união de dois conjuntos disjuntos é igual à soma de suas medidas. Porém, certas medidas não podem ser definidas sobre todos os subconjuntos possíveis (por exemplo, para a medida de Lebesgue sobre os reais é possível demonstrar que existirão conjuntos reais não mensuráveis). Por isso, temos que considerar uma coleção especial de subconjuntos de X onde a medida está definida, a chamada σ -álgebra de subconjuntos de X . Elementos da σ -álgebra são chamados de conjuntos mensuráveis. A teoria da medida dará a formulação exata do que é uma probabilidade.

4.1 σ -álgebras

Como dito anteriormente, a σ -álgebra é uma família de subconjuntos de X a qual podemos definir uma medida. Mais exatamente:

Definição 12. *Uma σ -álgebra de um conjunto X é uma família Σ de subconjuntos de X tais que:*

1. $\emptyset \in \Sigma$;
2. Para todo $A \in \Sigma$, seu complemento $A^C = X \setminus A \in \Sigma$
3. Para toda sequência $(E_n)_{n \in \mathbb{N}}$ em Σ , $\bigcup_{n \in \mathbb{N}} E_n \in \Sigma$

Observação 5. *Uma álgebra de conjuntos é uma família de subconjuntos fechada com respeito às operações de complementação e união **finita**. A σ -álgebra também é fechada pela união enumerável.*

Teorema 3. *Propriedades de uma σ -álgebra: Se Σ é uma σ -álgebra de subconjuntos de X , então, para todo $E, F \in \Sigma$, tem-se:*

I. Se $(E_n)_{n \in \mathbb{N}}$ é uma seqüência de Σ , então $\bigcap_{n \in \mathbb{N}} E_n \in \Sigma$

II. $E \setminus F \in \Sigma$

Prova. Primeiro vamos provar a seguinte igualdade de conjuntos:

$$\bigcap_{i \in I} E_i = \left(\bigcup_{i \in I} E_i^c \right)^c.$$

Para provarmos a igualdade acima seja

$$\begin{aligned} x \in \bigcap_{i \in I} E_i &\Leftrightarrow x \in E_i \quad \forall i \in I \Leftrightarrow \\ x \notin E_i^c \quad \forall i \in I &\Leftrightarrow x \notin \bigcup_{i \in I} E_i^c \Leftrightarrow \\ x \in \left(\bigcup_{i \in I} E_i^c \right)^c & \end{aligned}$$

I. Sejam $E, F \in \Sigma$. Então $F^c \in \Sigma$. Segue pelo item 1, já provado, que $E \cap F^c \in \Sigma$. Como

$$E \setminus F = E \cap F^c$$

II. Seja $(E_n)_{n \in \mathbb{N}}$, com $E_n \in \Sigma$ para todo $n \in \mathbb{N}$. Então $E_n^c \in \Sigma$ para todo $n \in \mathbb{N}$. Logo $\bigcup_{n \in \mathbb{N}} E_n^c \in \Sigma$ e finalmente

$$\bigcap_{n \in \mathbb{N}} E_n = \left(\bigcup_{n \in \mathbb{N}} E_n^c \right)^c \in \Sigma.$$

segue que $E \setminus F \in \Sigma$.

Teorema 4. Seja $S = (\Sigma_i)_{i \in I}$ uma família (não vazia) de σ -álgebras de subconjuntos de X . Então:

$$\bigcap_{i \in I} \Sigma_i = \{E \in \Sigma_i; \forall i \in I\},$$

ou seja, a interseção de todas as σ -álgebras que pertencem a S é uma σ -álgebra de X .

Prova. Vamos verificar os axiomas de Σ -álgebra:

1. $\emptyset \in \Sigma_i$ para todo $i \in I$. Logo $\emptyset \in \bigcap_{i \in I} \Sigma_i = \Sigma$
2. Seja $E \in \Sigma$, então $E \in \Sigma_i$ para todo $i \in I$. Segue que $E^c \in \Sigma_i$ para todo $i \in I$, isto é, $E^c \in \Sigma$.
3. Seja $\langle E_n \rangle_{n \in \mathbb{N}}$, com $E_n \in \Sigma$ para todo $n \in \mathbb{N}$. Então para cada $i \in I$, $E_n \in \Sigma_i$ para todo $n \in \mathbb{N}$. Como Σ_i é uma σ -álgebra segue que

$$\bigcup_{n \in \mathbb{N}} E_n \in \Sigma_i \quad \text{para todo } i \in I,$$

isto é $\bigcup_{n \in \mathbb{N}} E_n \in \Sigma$.

Definição 13. A σ -álgebra gerada pela família de abertos de \mathbb{R} (ou \mathbb{R}^n) é conhecida como σ -álgebra de Borel. Seus elementos são os **conjuntos de Borel** ou **borelianos**.

4.2 Espaços com medida

Definição 14. Dizemos que uma sequência $\{E_n\}_{n \in \mathbb{N}}$ de subconjuntos de X é disjunta se nenhum ponto pertence a mais do que um E_n , isto é, $E_m \cap E_n = \emptyset$ para todos $m, n \in \mathbb{N}$ distintos. Generalizando, se $\{E_i\}_{i \in I}$ é uma família de conjuntos indexada por um conjunto arbitrário I , então E_i é disjunta se $E_i \cap E_j = \emptyset$ para todos $i, j \in I$ disjuntos.

Usualmente, definimos funções no intervalo infinito *aberto*. Porém, para estudar medida, precisamos definir uma função no intervalo infinito *fechado*, ou seja, no intervalo $[0, \infty]$. Este conjunto é a união do elemento ∞ com o intervalo $[0, \infty[\subset \mathbb{R}$. Em medida, representamos comprimento, área ou volume infinito com o símbolo ∞ . Este é um novo significado para este símbolo na Matemática.

Definição 15. Vamos definir as operações básicas envolvendo ∞ :

1. Adição (+): $\infty + \infty = \infty + a = a + \infty = \infty$ para todo $a \in \mathbb{R}$;
2. Subtração (-): $\infty - a = \infty$ para todo $a \in \mathbb{R}$, mas $\infty - \infty$ não está definido;
3. Multiplicação (\cdot): $\infty \cdot \infty = a \cdot \infty = \infty \cdot a = \infty$ para todo $a > 0$ e convencionamos $0 \cdot \infty = \infty \cdot 0 = 0$

A relação de ordem, incluindo ∞ fica definida da seguinte maneira: $a < \infty$ para todo $a \in \mathbb{R}$. Com isto, definimos, também, supremo e ínfimo de subconjuntos de $\mathbb{R} \cup \{\infty\}$.

Definição 16. $\inf \emptyset = \infty$

Dado $(x_i)_{i \in I}$ com $x_i \in [0, \infty]$, definimos

$$\sup \left\{ \sum_{i \in J} x_i; \quad J \subset I \quad \text{é finito} \right\} = \sum_{i \in I} x_i$$

Se $I = \emptyset$, então definimos $\sum_{i \in I} x_i = 0$

Definição 17. Um **Espaço de medida** é uma tripla (X, Σ, μ) onde:

1. X é um conjunto;
2. Σ é uma σ -álgebra de subconjuntos de X ;
3. $\mu : \Sigma \rightarrow [0, \infty]$ é uma função tal que:
 - (a) $\mu(\emptyset) = 0$;
 - (b) se $\{E_n\}_{n \in \mathbb{N}}$ é uma sequência disjunta em Σ , então $\mu(\bigcup_{n \in \mathbb{N}} E_n) = \sum_{n=0}^{\infty} \mu(E_n)$.

Os elementos de Σ são chamados de **conjuntos mensuráveis** (ou μ -mensuráveis), e μ é chamado de uma **medida em X** . A propriedade (b) é chamada de **σ -aditividade** ou **aditividade contável**.

Observação 6. Uma medida definida em uma σ -álgebra de Borel é conhecida como **medida de Borel**.

4.3 Propriedades elementares da medida

Teorema 5. Seja (X, Σ, μ) um espaço de medida.

1. Se $E, F \in \Sigma$ e $E \subset F$, então $\mu(E) \leq \mu(F)$
2. $\mu(E \cup F) \leq \mu(E) + \mu(F)$ para todo $E, F \in \Sigma$.
3. Se $\{E_n\}_{n \in \mathbb{N}}$ é uma sequência em Σ , então $\mu(\bigcup_{n \in \mathbb{N}} E_n) \leq \sum_{n=0}^{\infty} \mu(E_n)$.
4. Se $\{E_n\}_{n \in \mathbb{N}}$ é uma sequência não-decrescente em Σ (isto é, $E_n \subset E_{n+1}$ para todo $n \in \mathbb{N}$, então

$$\mu\left(\bigcup_{n \in \mathbb{N}} E_n\right) = \lim_{n \rightarrow \infty} \mu(E_n) = \sup_{n \in \mathbb{N}} \mu(E_n).$$

5. Se $\{E_n\}_{n \in \mathbb{N}}$ é uma sequência crescente em Σ (isto é, $E_{n+1} \subset E_n$ para todo $n \in \mathbb{N}$ e se algum $\mu(E_n)$ é finito, então

$$\mu\left(\bigcap_{n \in \mathbb{N}} E_n\right) = \lim_{n \rightarrow \infty} \mu(E_n) = \inf_{n \in \mathbb{N}} \mu(E_n).$$

Prova. 1. Como $E \subset F$, então $F = (F \setminus E) \cup E$.

Assim, $\mu(F) = \mu(F \setminus E) + \mu(E) \geq \mu(E)$

2. $E \cup F = (E \setminus F) \cup F$. Então, $\mu(E \cup F) = \mu(E \setminus F) + \mu(F)$.

Como $E \setminus F \subset E$, por (1) acima, $\mu(E) \geq \mu(E \setminus F)$. Assim, $\mu(E \cup F) \leq \mu(E) + \mu(F)$

3. Seja $F_0 = E_0, F_n = E_n \setminus \bigcup_{i < n} E_i$ para $n \geq 1$; então $\langle F_n \rangle_{n \in \mathbb{N}}$ é uma sequência disjunta em Σ .

Além disso

$$\bigcup_{n \in \mathbb{N}} E_n = \bigcup_{n \in \mathbb{N}} F_n.$$

Como os F_n são disjuntos segue que

$$\mu \left(\bigcup_{n \in \mathbb{N}} F_n \right) = \sum_{n=0}^{\infty} \mu(F_n).$$

Mas como $F_n \subset E_n$ segue que $\mu(F_n) \leq \mu(E_n)$ para todo $n \in \mathbb{N}$. Logo

$$\mu \left(\bigcup_{n \in \mathbb{N}} E_n \right) = \mu \left(\bigcup_{n \in \mathbb{N}} F_n \right) = \sum_{n=0}^{\infty} \mu(F_n) \leq \sum_{n=0}^{\infty} \mu(E_n).$$

4. Seja $F_0 = E_0, F_n = E_n \setminus E_{n-1}$ para $n \geq 1$; então $\{F_n\}_{n \in \mathbb{N}}$ é uma sequência disjunta em Σ e $\bigcup_{n \in \mathbb{N}} F_n = \bigcup_{n \in \mathbb{N}} E_n$. Consequentemente $\mu(\bigcup_{n \in \mathbb{N}} E_n) = \sum_{n=0}^{\infty} \mu(F_n)$. Mas uma indução

fácil em n , usando (1) para o passo indutivo, mostra que $\mu(E_k) = \sum_{n=0}^k \mu(F_n)$ para todos k .

Então,

$$\sum_{n=0}^{\infty} \mu(F_n) = \lim_{k \rightarrow \infty} \sum_{n=0}^k \mu(F_n) = \lim_{k \rightarrow \infty} \mu(E_k).$$

Finalmente, $\lim_{n \rightarrow \infty} \mu(E_n) = \sup_{n \in \mathbb{N}} \mu(E_n)$ porque (por (1)) $\mu(E_n) = \sum_{m=0}^n \mu(F_m)$ é não-decrescente.

5. Suponha que $\mu E_k < \infty$. Defina $F_n := E_k \setminus E_{k+n}$ para $n \in \mathbb{N}$, $F = \bigcup_{n \in \mathbb{N}} F_n$; então $\{F_n\}_{n \in \mathbb{N}}$ é uma sequência não-decrescente em Σ e $\mu(F) = \lim_{n \rightarrow \infty} \mu(F_n)$, por (5) acima. Temos que $\mu(F_n) + \mu(E_{k+n}) = \mu(E_k) < \infty$, podemos escrever que $\mu(F_n) = \mu(E_k) - \mu(E_{k+n})$, e, portanto

$$\mu(F) = \lim_{n \rightarrow \infty} (\mu(E_k) - \mu(E_{k+n})) = \mu(E_k) - \lim_{n \rightarrow \infty} \mu(E_n).$$

Agora, $F \subset E_k$, então $\mu(F) + \mu(E_k \setminus F) = \mu(E_k)$, e (novamente pois $\mu(E_k)$ é infinito), $\mu(F) = \mu(E_k) - \mu(E_k \setminus F)$. Portanto, temos que $\mu(E_k \setminus F) = \lim_{n \rightarrow \infty} \mu(E_n)$. Mas $E_k \setminus F$ é somente $\bigcap_{n \in \mathbb{N}} E_n$. Finalmente, $\lim_{n \rightarrow \infty} \mu(E_n) = \inf_{n \in \mathbb{N}} \mu(E_n)$ pois $\{\mu(E_n)\}_{n \in \mathbb{N}}$ é não-crescente.

Definição 18. Seja (X, Σ, μ) um espaço de medida. Um conjunto $A \subset X$ possui **medida nula** se existe um conjunto $E \in \Sigma$ tal que $A \subset E$ e $\mu(E) = 0$.

Observação 7. Um conjunto de medida nula não precisa ser mensurável, embora esteja contido em um conjunto mensurável de medida nula.

Definição 19. Espaços com medidas em que todos os conjuntos de medida nula não mensuráveis é chamado de espaço **completo**.

4.4 Medida de Lebesgue em \mathbb{R}

A medida de Lebesgue, além de ser a mais importante para aplicações, é também a medida sobre a qual os resultados da Teoria Geral da Medida foram desenvolvidos.

Definição 20. Seja $I = [a, b) \subset \mathbb{R}$ um intervalo semiaberto. Definimos seu **comprimento** $\lambda(I)$ por

$$\lambda(\emptyset) = 0, \quad \lambda([a, b)) = b - a \text{ se } a < b$$

Definição 21. Definimos $\theta^* : \mathcal{P}(\mathbb{R}) \rightarrow [0, \infty]$, a **medida exterior de Lebesgue** por

$$\theta^*(A) = \inf \left\{ \sum_{j=0}^{\infty} \lambda(I_j); \{I_j\}_{j \in \mathbb{N}} \text{ é uma sequência de intervalos semiabertos tal que } A \subset \bigcup_{j \in \mathbb{N}} I_j \right\}$$

Definição 22. Dizemos que $A \subset \mathbb{R}$ tem **medida de Lebesgue nula** se, para todo $\varepsilon > 0$, existe uma sequência $(E_n)_{n \in \mathbb{N}}$ de intervalos abertos e limitados tal que

$$A \subset \bigcup_{n=1}^{+\infty} E_n \quad \text{e} \quad \sum_{n=1}^{+\infty} |E_n| \leq \varepsilon,$$

sendo que $|I|$ representa o comprimento de I , ou seja, $|I| = b - a$ se $I = (a, b)$

Definição 23. Dizemos que $E \subset \mathbb{R}$ é **mensurável a Lebesgue** se para cada $A \subseteq \mathbb{R}$, tem-se

$$\theta^*(A) = \theta^*(A \cap E) + \theta^*(A \cap E^c).$$

Se E é um conjunto mensurável segundo Lebesgue, então a medida de Lebesgue de E , que denotaremos por $\mu(E)$, é a medida exterior de E , isto é

$$\mu(E) = \theta^*(E).$$

Observação 8. Como a medida exterior de Lebesgue é contavelmente subaditiva, isto é,

$$\theta^* \left(\bigcup_i E_i \right) \leq \sum_i \theta^*(E_i)$$

segue que para provar que um conjunto E é mensurável basta provar que

$$\theta^*(A) \geq \theta^*(A \cap E) + \theta^*(A \cap E^c)$$

é válida para todo $A \subseteq \mathbb{R}$.

Teorema 6. Todo conjunto de Borel de \mathbb{R} é mensurável a Lebesgue.

Prova. Como a σ -álgebra de Borel é a menor σ -álgebra que contém os abertos, basta provar que todo aberto é mensurável. Como todo aberto de \mathbb{R} é uma união disjunta e contável de intervalos abertos, basta provar que um intervalo do tipo (a, b) é mensurável. Seja $a \in \mathbb{R}$. Provaremos que (a, ∞) é mensurável, o que é suficiente. Vamos provar que

$$\theta^*(A \cap (a, \infty)) + \theta^*(A \cap (-\infty, a]) \leq \theta^*(A).$$

Esta desigualdade é óbvia se $\theta^*(A) = +\infty$, logo podemos assumir que $\theta^*(A) < \infty$. Seja $\varepsilon > 0$. Então existe uma sequência de intervalos abertos $\{I_k\}_{k \in \mathbb{N}}$ tal que $A \subseteq \bigcup_k I_k$ e $\sum_k \lambda(I_k) < \theta^*(A) + \varepsilon$, onde $\lambda(I_k)$ denota o comprimento do intervalo I_k . Para cada k , seja $I_k^1 = I_k \cap (a, \infty)$ e $I_k^2 = I_k \cap (-\infty, a]$. $\{I_k^1\}$ e $\{I_k^2\}$ são sequências de intervalos que cobrem $A \cap (a, \infty)$ e $A \cap (-\infty, a]$ e que

$$\theta^*(I_k^1) + \theta^*(I_k^2) = \lambda(I_k).$$

Logo,

$$\theta^*(A \cap (a, \infty)) + \theta^*(A \cap (-\infty, a]) \leq \sum_k \lambda(I_k) < \theta^*(A) + \varepsilon.$$

Como $\varepsilon > 0$ é arbitrário, segue o resultado.

Este resultado implica que todos os conjuntos abertos e fechados e todos os intervalos são conjuntos mensuráveis a Lebesgue.

4.5 Medida de Probabilidade

Definição 24. Um espaço de medida (Ω, σ, μ) é dito um espaço de probabilidade se $\mu(\Omega) = 1$

Definição 25. Dada uma medida de probabilidade μ sobre Ω , dizemos que uma propriedade se verifica para quase todo $x \in \Omega$ ou é **quase segura** se o conjunto de pontos onde ela não se verifica possui medida zero.

Exemplo 5. Se $f : \Omega \subset \mathbb{R} \rightarrow \mathbb{R}$ é uma função mensurável e $\lim_{x \rightarrow a} f(x) = L$ para todo $x \in \Omega$, exceto em um conjunto de medida nula, dizemos que f converge para L quase seguramente quando x tende a a e denotamos:

$$\lim_{x \rightarrow a} f(x) = L \quad (q.s.)$$

5 *A Lei de Benford*

No ano de 1881, Simon Newcomb, um matemático e astrônomo americano, percebeu que as primeiras páginas dos livros de logaritmos das bibliotecas estavam mais gastas que as últimas. Ele investigou o assunto mais profundamente e percebeu que, em amostras aleatórias de dados reais, a ocorrência dos primeiros dígitos de um número não obedeciam a distribuição mais intuitiva, de $\frac{1}{9}$, porém os números menores apareciam com maior frequência. O dígito 1 aparece quase $\frac{1}{3}$ das vezes.

Newcomb publicou neste mesmo ano um breve artigo no jornal americano “American Journal of Mathematics” [10]. Neste artigo ele mencionou que havia observado as páginas dos livros logarítmicos e especulava que a probabilidade de um número d aparecer como primeiro dígito é de $\log_{10}(1 + d^{-1})$. Seu estudo foi esquecido e, muitos anos mais tarde, em 1938, o físico Frank Benford, após uma profunda investigação, chegou à mesma conclusão que Newcomb [1]. Benford, porém, ampliou o estudo e verificou que a fórmula pode ser aplicada para uma variedade de ocorrências, inclusive naturais, de números. No ano de 1995, Theodore P. Hill publicou a demonstração rigorosa da suspeita de Benford e Newcomb [6].

Enunciamos, agora, a Lei de Benford.

5.1 A Lei dos Primeiros Dígitos

“A lei da probabilidade de ocorrência dos números é tal que todas as mantissas dos seus logaritmos são equiprováveis.”

Devemos saber que mantissa de um número real positivo x é o número $r \in [\frac{1}{10}, 1)$ com $x = r10^n$ para algum inteiro n . Por exemplo, as mantissas de 123 e 0,0123 são ambas .123.

Esta lei nos diz que a probabilidade de um número ter o primeiro dígito significativo igual a 1 é $\log_{10} 2 \approx .301$., a probabilidade de ser 2 é de $\log_{10}(\frac{3}{2}) \approx .176$. e assim por diante, monotona-mente, até a probabilidade de $\log_{10}(\frac{10}{9}) \approx .046$. para o dígito 9. Para cada número inteiro positivo x com m dígitos, considere sua representação na forma $x = d_1 d_2 \dots d_m$, onde $1 \leq d_1, d_m \leq 9$ e

$0 \leq d_2, \dots, d_{m-1} \leq 9$. Formalmente, a lei de Benford para os primeiros dígitos é da seguinte forma:

Definição 26. A lei de Benford para a distribuição dos dígitos significativos, estabelece que ao escolhermos aleatoriamente um número inteiro positivo $X = D_1 D_2 \dots D_m$, teremos para cada $1 \leq k \leq m$ e $D_1, D_m \in \{1, 2, \dots, 9\}$ e $D_j \in \{0, 1, \dots, 9\}$, $j = 2, \dots, k$

$$\text{Prob}(D_1 = d_1, \dots, D_k = d_k) = \log_{10} \left[1 + \left(\sum_{i=1}^k d_i \cdot 10^{k-i} \right)^{-1} \right]. \quad (5.1)$$

Por exemplo, $\text{Prob}(D_1 = 3, D_2 = 1, D_3 = 4) = \log_{10}(1 + (314)^{-1}) \cong .0014$. Em particular, (5.1) implica que

1. $\text{Prob}(\text{primeiro dígito significativo} = d) = \log_{10}(1 + d^{-1}) \quad d = 1, 2, \dots, 9.$

e

2. $\text{Prob}(\text{segundo dígito significativo} = d) = \sum_{k=1}^9 \log_{10}(1 + (10k + d)^{-1}) \quad d = 1, 2, \dots, 9.$

A forma geral da lei é:

3. $\text{Prob}(\text{mantissa} \leq \frac{t}{10}) = \log_{10} t, \quad t \in [1, 10)$

Podemos verificar, usando as propriedades acima, que para números inteiros obtidos em experimentos probabilísticos que sigam a lei de Benford, os dígitos significativos são *dependentes* e não independentes. Segue de (2) que a probabilidade (não condicionada) de que o segundo dígito seja 2 é aproximadamente .109., mas por (5.1) a probabilidade (condicional) de que o segundo dígito seja 2, dado que o primeiro é 1, é de aproximadamente .115.. Esta dependência entre os dígitos significativos cai rapidamente conforme a distância entre os dígitos cresce e, segue da lei geral (5.1) que, a distribuição do n-ésimo dígito significativo se aproxima da distribuição uniforme em $\{0, 1, \dots, 9\}$ exponencialmente quando $n \rightarrow \infty$.

É claro que muitas tabelas de dados numéricos não seguem esta distribuição logarítmica. Listas telefônicas, por exemplo, geralmente começam com o mesmo dígito, até mesmo dados “neutros” como tabelas de raízes quadradas de inteiros não se encaixam na lei. Mas, um número surpreendentemente grande de coleções de dados numéricos parece obedecer à lei de Benford.

Exemplo 6. Consideremos a sequência dos números de Fibonacci: $1, 1, 2, 3, 5, 8, 13, 21, \dots$, onde cada número é formado pela soma de seus 2 antecessores imediatos. Tomando o primeiro dígito dos 100 primeiros números de Fibonacci e contando a sua frequência de aparição, obtemos:

Primeiro dígito / Frequência

$$1/30$$

$$2/18$$

$$3/12$$

$$4/9$$

$$5/8$$

$$6/6$$

$$7/5$$

$$8/7$$

$$9/4$$

Se compararmos os resultados obtidos acima com a tabela dada na introdução deste trabalho, veremos que a distribuição do primeiro dígito é muito próxima da que percebeu Benford. Quanto mais números da sequência de Fibonacci tomarmos, mais a frequência de aparição dos primeiros dígitos se aproxima da lei estudada.

5.2 A σ -álgebra \mathcal{M}

Neste seção vamos dar uma estrutura formal à lei de Benford para a distribuição dos dígitos significativos. Para tal seja \mathbb{R}^+ o conjunto dos reais positivos. Sobre \mathbb{R}^+ definimos \mathcal{M} a σ -álgebra gerada pela função

$$\Phi(x) = \text{mantissa}(x), \quad \forall x \in \mathbb{R}^+$$

isto é

$$\mathcal{M} = \{\Phi^{-1}(B) : B \in \mathcal{B}\}$$

É fácil ver que \mathcal{M} é uma sub σ -álgebra de Borel e que

$$S \in \mathcal{M} \iff S = \bigcup_{n=-\infty}^{\infty} B \cdot 10^n \quad \text{para algum boreleano } B \subseteq [1, 10), \quad (5.2)$$

Teorema 7. *A σ -álgebra \mathcal{M} satisfaz as propriedades*

- i. Todo subconjunto não-vazio de \mathcal{M} é infinito e tem 0 e ∞ como pontos de acumulação.
- ii. \mathcal{M} é fechada para a multiplicação por escalar, isto é, $s > 0, S \in \mathcal{M} \implies sS \in \mathcal{M}$
- iii. \mathcal{M} é fechada para raízes inteiras, isto é, $n \in \mathbb{N}, S \in \mathcal{M} \implies S^{1/n} \in \mathcal{M}$, mas não para potências.
- iv. \mathcal{M} é auto-similar, no seguinte sentido: se $S \in \mathcal{M} \implies 10^m S \in \mathcal{M}$ para todo inteiro m .

Da propriedade i) conclui-se que intervalos finitos como $[1,2)$ não pertencem a \mathcal{M}

5.3 Invariância na base e invariância na escala

Definição 27. Uma medida de probabilidade P em $(\mathbb{R}^+, \mathcal{M})$ é invariante na escala se

$$P(S) = P(sS) \quad \text{para todo } s > 0, \text{ e todo } S \in \mathcal{M}.$$

Exemplo 7. Vamos considerar uma mudança de escala simples obtida multiplicando-se todos os números de um conjunto por 2. Se considerarmos os números com primeiro dígito significativo igual a 1, então eles são mudados para números com primeiro dígito significativo ou igual a 2 ou igual a 3. É fácil verificar que $B(1) = B(2) + B(3)$. De fato,

$$\begin{aligned} B(2) + B(3) &= \log_{10}\left(1 + \frac{1}{2}\right) + \log_{10}\left(1 + \frac{1}{3}\right) \\ &= \log_{10}\frac{3}{2} + \log_{10}\frac{4}{3} = \log_{10}\frac{3 \cdot 4}{2 \cdot 3} \\ &= \log_{10}2 = B(1), \end{aligned}$$

Analogamente $B(2) = B(4) + B(5)$, etc..

O seguinte teorema mostra que a invariância por escala caracteriza a lei geral dos dígitos significativos:

Teorema 8. [Hill] Uma medida de probabilidade P em $(\mathbb{R}^+, \mathcal{M})$ é invariante por escala se e somente se,

$$P\left(\bigcup_{n=-\infty}^{\infty} [1, t) \cdot 10^n\right) = \log_{10} t \quad \text{para todo } t \in [1, 10). \quad (5.3)$$

No lugar disso, imaginemos que uma lei universal do dígito significativo devesse satisfazer a propriedade de ser invariante na base, isto é, ser válida quando escrita em qualquer outra base além da base 10. De fato, todos os argumentos que sustentam a lei de Benford são válidos em

qualquer base[Rami, 1976]. No sentido de motivar a definição de invariância de base, considere o conjunto $\{D_1 = 1\}$ em base 10. Este conjunto pode ser escrito

$$\{D_1 = 1\} = \bigcup_{n=-\infty}^{\infty} [1,2) \cdot 100^n \cup \bigcup_{n=-\infty}^{\infty} [10,20) \cdot 100^n,$$

isto é, $\{D_1 = 1\}$ é também o conjunto dos números positivos cujo primeiro dígito significativo, em base 100, está no conjunto $\{1, 10, 11, \dots, 19\}$. De modo geral, todo conjunto de números reais S (base 10) em \mathcal{M} é o mesmo que o conjunto de reais $S^{1/2}$ (base 100). Assim, se uma probabilidade é invariante por bases, a medida de qualquer conjunto em \mathcal{M} deveria ser a mesma para todas as bases. Isto sugere a seguinte definição:

Definição 28. *Uma medida de probabilidade P em $(\mathbb{R}^+, \mathcal{M})$ é invariante por bases se $P(S) = P(S^{1/n})$ para todo inteiro positivo n e todo $S \in \mathcal{M}$.*

Seja o conjunto

$$S_1 = \{D_1 = 1, D_j = 0 \forall j > 0\} = \{\dots, .01, .1, 1, 10, 100, \dots\} = \bigcup_{n=-\infty}^{\infty} \{1\} \cdot 10^n \in \mathcal{M}$$

O conjunto S_1 não admite subconjuntos não-vazios \mathcal{M} -mensuráveis por (8), assim a medida de Dirac associada a este conjunto denotada por δ_1 está bem definida. Seja P_L a distribuição logarítmica de probabilidade em $(\mathbb{R}^+, \mathcal{M})$ dado em (8). Então:

Teorema 9. *Uma medida de probabilidade P em $(\mathbb{R}^+, \mathcal{M})$ é invariante por bases se, e somente se,*

$$P = qP_L + (1 - q)\delta_1 \quad \text{para todo } q \in [0, 1].$$

Usando os teoremas 8 e 9 pode-se mostrar que invariância por escala implica em invariância por bases mas a recíproca não é verdadeira (δ_1 é claramente invariante por bases mas não por escalas). As provas envolvem resultados além do escopo deste TCC.

5.4 Amostras aleatórias de distribuições aleatórias

Para entendermos como diferentes tipos de tabelas se aplicam à lei de Benford vamos relembrar a definição de uma medida aleatória de probabilidade:

Definição 29. *Uma medida aleatória de probabilidade (m.a.p) \mathbb{M} é um vetor aleatório (sobre um espaço de probabilidade (Ω, \mathcal{F}, P)) com valores em medidas de probabilidades boreleanas*

definidas em \mathbb{R} , regulares no sentido de que para cada conjunto de Borel $B \subset \mathbb{R}$, $\mathbb{M}(B)$ é uma variável aleatória.

Definição 30. A medida de distribuição esperada de uma m.a.p \mathbb{M} é uma medida de probabilidade

$$E\mathbb{M}(B) = E(\mathbb{M}(B)) \quad \text{para todo boreleano } B \subset \mathbb{R}, \quad (5.4)$$

(onde E denota a partir de agora a esperança com respeito a P dos valores do próprio espaço).

Definição 31. Dados uma m.a.p \mathbb{M} e um inteiro positivo k , uma sequência de k -amostras \mathbb{M} -aleatórias é uma sequência de variáveis aleatórias X_1, X_2, \dots em (Ω, \mathcal{F}, P) tal que, para alguma sequência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) $\mathbb{M}_1, \mathbb{M}_2, \dots$ de m.a.p com a mesma distribuição que \mathbb{M} , e para cada $j = 1, 2, \dots$ valem:

Dado $\mathbb{M}_j = P$, as variáveis aleatórias $X_{(j-1)k+1}, \dots, X_{jk}$ são i.i.d. com função de distribuição P ;

(5.5)

e

$$X_{(j-1)k+1}, \dots, X_{jk} \text{ são independentes de } \{\mathbb{M}_i, X_{(i-1)k+1}, \dots, X_{ik}\} \forall i \neq j. \quad (5.6)$$

Lema 1. Sejam X_1, X_2, \dots uma sequência de k -amostras \mathbb{M} -aleatória para algum k e alguma m.a.p \mathbb{M} . Então:

- i. $\{X_n\}$ são identicamente distribuídas com distribuição $E\mathbb{M}$, mas não são sempre independentes, e
- ii. Dado $\{\mathbb{M}_1, \mathbb{M}_2, \dots\}$ as $\{X_n\}$ são quase seguras, independentes, mas não são sempre identicamente distribuídas.

Prova. A primeira parte de ii) segue facilmente por (5.5) e (5.6), e a segunda parte porque sempre que $\mathbb{M}_i \neq \mathbb{M}_j$, X_{ik} não terá a mesma distribuição de X_{jk} . A primeira parte de i) segue condicionando a \mathbb{M}_j ;

$$P(X_j \in B) = E[\mathbb{M}_j(B)] = E[\mathbb{M}(B)] \quad \text{para todo boreleano } B \subset \mathbb{R},$$

onde a última igualdade segue já que \mathbb{M}_j tem a mesma distribuição que \mathbb{M} . A segunda parte de i) segue do fato que amostras independentes e identicamente distribuídas de uma distribuição podem dar informações sobre a distribuição.

O próximo lema é simplesmente a formalização do fato intuitivo que diz que distribuições empíricas de k -amostras de \mathbb{M} -aleatória convergem para a distribuição esperada de \mathbb{M} .

Lema 2. *Seja \mathbb{M} uma m.a.p, e sejam X_1, X_2, \dots uma seqüência de k -amostras \mathbb{M} -aleatória para algum k . Então*

$$\lim_{n \rightarrow \infty} \frac{\#\{i \leq n : X_i \in B\}}{n} = E[\mathbb{M}(B)] \text{ q.s. } \forall B \subset \mathbb{R}.$$

Prova. Fixemos B e $j \in \mathbb{N}$, e seja

$$Y_j = \#\{m, 1 \leq m \leq k : X_{(j-1)k+m} \in B\}.$$

Claramente

$$\lim_{n \rightarrow \infty} \frac{\#\{i \leq n : X_i \in B\}}{n} = \lim_{m \rightarrow \infty} \frac{\sum_{j=1}^m Y_j}{km} \text{ se o limite existir.} \quad (5.7)$$

Por (5.5), dado \mathbb{M}_j , Y_j 'tem uma distribuição binomial com parâmetros k e $E[\mathbb{M}_j(B)]$, assim por (5.4)

$$EY_j = E(E(Y_j|\mathbb{M}_j)) = kE[\mathbb{M}(B)] \text{ q.s. para todo } j, \quad (5.8)$$

uma vez que \mathbb{M}_j tem a mesma distribuição que \mathbb{M} .

Por (5.6), os $\{Y_j\}$ são independentes. Como eles tem médias idênticas (via (5.8)), a saber $kE[\mathbb{M}(B)]$, e são uniformemente limitadas ($\sum(\text{Var}(Y_j)/j^2) < \infty$), segue por [Loève(1977)] que

$$\lim_{m \rightarrow \infty} \frac{\sum_{j=1}^m Y_j}{m} = kE[\mathbb{M}(B)] \text{ q.s.} \quad (5.9)$$

e a conclusão segue por (5.7) e (5.9).

5.5 Lei do limite logarítmico para dígitos significativos

Definição 32. *Uma seqüência de variáveis aleatórias X_1, X_2, \dots tem uma frequência de mantissa neutra na escala se*

$$n^{-1} |\#\{i \leq n : X_i \in S\} - \#\{i \leq n : X_i \in sS\}| \rightarrow 0 \text{ q.s.}$$

para todo $s > 0$ e todo $S \in \mathbb{M}$, e tem frequência de mantissa neutra na base se:

$$n^{-1} |\#\{i \leq n : X_i \in S\} - \#\{i \leq n : S_i \in S^{\frac{1}{m}}\}| \rightarrow 0 \text{ q.s.}$$

para todo $m \in \mathbb{N}$ e todo $S \in \mathcal{M}$.

Por exemplo, se $\{X_n\}$, $\{Y_n\}$ e $\{Z_n\}$ são as seqüências das variáveis aleatórias definidas por $X_n \equiv 1, Y_n \equiv 2, Z_n \equiv 2^n$, então $\{X_n\}$ tem frequência de mantissa neutra na base mas não neutra

na escala, $\{Y_n\}$ não tem nenhum e $\{Z_n\}$ tem os dois.

Definição 33. Uma m.a.p \mathbb{M} é imparcial na escala se sua distribuição esperada $E\mathbb{M}$ for invariante na escala em $(\mathbb{R}^+, \mathbb{M})$ e é imparcial na base se $E\mathbb{M}$ é invariante na base em $(\mathbb{R}^+, \mathbb{M})$. (Lembrando que \mathbb{M} é uma sub σ -álgebra de Borel, então toda probabilidade de Borel nos \mathbb{R} (e também em $E\mathbb{M}$) leva a uma única probabilidade em $(\mathbb{R}^+, \mathbb{M})$.)

O ponto crucial aqui é que as definições de imparcial na escala e na base não requerem que as realizações individuais de \mathbb{M} sejam invariantes na escala ou na base; na verdade, é frequente o caso em que nenhuma das realizações é invariante na escala, mas apenas em que o processo de amostragem, em média, não favorece uma escala em detrimento de outro.

$\mathbb{M}(t)$ denota a variável aleatória $\mathbb{M}(D_t)$, onde $D_t = \bigcup_{n=-\infty}^{\infty} [1, t) \cdot 10^n$ é o conjunto dos números positivos com mantissa em $[\frac{1}{10}, \frac{t}{10})$.

Teorema 10. (Lei do limite logarítmico para dígitos significativos). Seja \mathbb{M} um m.a.p em $(\mathbb{R}^+, \mathbb{M})$. São equivalentes:

1. \mathbb{M} é imparcial na escala;
2. \mathbb{M} é imparcial na base e $E\mathbb{M}$ é sem átomos;
3. $E[\mathbb{M}(t)] = \log_{10} t$ para todo $t \in [1, 10)$;
4. Toda k -amostra \mathbb{M} -aleatória tem frequência de mantissa neutra na escala;
5. $E\mathbb{M}$ é sem átomos e toda k -amostra \mathbb{M} -aleatória tem frequência de mantissa neutra na base;
6. Para todas k – amostras \mathbb{M} -aleatórias X_1, X_2, \dots ,

$$n^{-1} \#\{i \leq n : \text{mantissa}(X_i) \in [\frac{1}{10}, \frac{t}{10})\} \rightarrow \log_{10} t \text{ q.s. para todo } t \in [1, 10).$$

Prova. “(i) \Leftrightarrow (iii). É imediato pelas definições 27 e 33 e pelo Teorema 8.

“(ii) \Leftrightarrow (iii)”. Segue facilmente de (5.2) que a probabilidade de Borel $E\mathbb{M}$ é sem átomos se, e somente se, ela for sem átomos em \mathcal{M} . A equivalência entre (ii) e (iii) segue facilmente pelas definições 28 e 33 e pelo teorema 9

“(iii) \Leftrightarrow (iv)” Pelo Lema 2,

$$A_n := n^{-1} \#\{i \leq n : X_i \in S\} \rightarrow E[\mathbb{M}(S)] \text{ q.s., e}$$

$$B_n := n^{-1} |\#\{i \leq n : X_i \in sS\}| \rightarrow E[\mathbb{M}(sS)] \text{ q.s.,}$$

então $|A_n - B_n| \rightarrow 0$ q.s. se, e somente se, $EM(S) = EM(sS)$, que pela definição 27 e pelo teorema 8 é equivalente a (iii).

“(iii) \Leftrightarrow (v).” Análogo, usando Lema 2, Definição 28 e teorema 9.

“(iii) \Leftrightarrow (vi).” Imediato pelo lema 2.

6 Conclusão

O objetivo geral deste trabalho foi apresentar a lei de Benford para a distribuição dos primeiros dígitos, tema que tem sido bastante estudado nos últimos anos devido a sua aplicabilidade. Essa lei nos diz que, em muitos grupos de dados, os primeiros dígitos não se distribuem uniformemente de 1 a 9. Ao invés disso, eles seguem uma distribuição logarítmica, onde $P(d) = \log_{10}(1 + d^{-1})$.

A lei de Benford tem originado um grande número de publicações (Ver [4]) e tem sido considerada uma ferramenta útil para a identificação de fraudes financeiras (Ver [3] e [2]), pois permite fazer uma comparação da frequência de primeiros dígitos observada nas amostras com a frequência esperada pela distribuição logarítmica, chamando a atenção para os casos que diferem muito do padrão.

É interessante pontuar, no entanto, que ela não é aplicável a conjuntos de números que apresentem, por exemplo, alguma das seguintes características:

- I. Conjuntos numéricos que estejam restritos a uma cota superior ou inferior, como os que representam o quociente intelectual, altura de pessoas, etc.;
- II. Conjuntos numéricos que sirvam para identificação de objetos, como números de contas bancárias, números de telefone, números de CPF, etc.;
- III. Conjuntos numéricos que tenham sido gerados aleatoriamente, como os resultados de loterias, bingos, etc.

Por outro lado, além do uso na detecção de fraudes financeiras, a lei de Benford pode ser aplicada para compreender fenômenos em diversas outras áreas tais como biologia, medicina e ciências sociais (ver [9], [7] e [11])

O fato de que dados numéricos obtidos a partir das medições dos mais variados fenômenos naturais ou sociais obedecerem à lei de Benford, provoca inegavelmente certo fascínio das pessoas pelo tema. Tal fascínio, juntamente com o grande poder de aplicabilidade dessa lei, fazem dela um tema de pesquisa aberto a muitos estudos.

Referências

- [1] F. Benford. The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78(4):551–572, Março 1938.
- [2] W. Hillison e C. Pacini C. Durtschi. The effective use of benford’s law to assist in detecting fraud in accounting data. *Journal of Forensic Accounting*, 5:17–39, 2004.
- [3] A. Diekmann e B. Jann. Benford’s law and fraud detection: Facts and legends. *German Economic Review*, 11(3):397–401, 2010.
- [4] T.P. Hill e E. Rogers. Benford online bibliography. <http://www.benfordonline.net>, Acesso em Agosto de 2014.
- [5] Pedro Jesus Fernández. *Medida e Integração*. IMPA, 2007.
- [6] T. P. Hill. A statistical derivation of the significant-digit law. *Statistical Science*, 10(4):354–363, 1995.
- [7] T. Goldman e J. P. Mercader J. L. Friar. Genome sizes and the benford distribution. *Plos One*, Maio 2012.
- [8] B. R. James. *Probabilidade: um curso em nível intermediário*. IMPA, 2004.
- [9] T.A. Mir. The law of the leading digits and the world religions. *Physica A: Statistical Mechanics and its Applications*, 391(3):792–798, Fevereiro 2012.
- [10] S. Newcomb. Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics*, 4(1):39–40, 1881.
- [11] N. Sheehy e J. Cooke S. Cournane. The novel application of benford’s second order analysis for monitoring radiation output in interventional radiology. *Physica Medica: European Journal of Medical Physics*, Novembro 2013.