

Notas de Aula do Curso  
ET581: Probabilidade 1

**Leandro Chaves Rêgo, Ph.D.**

2010.2

# Prefácio

Estas notas de aula foram feitas para compilar o conteúdo de várias referências bibliográficas tendo em vista o conteúdo programático da disciplina ET581-Probabilidade 1 do curso de graduação em Estatística da Universidade Federal de Pernambuco. Em particular, elas não contém nenhum material original e não substituem a consulta a livros textos. Seu principal objetivo é dispensar a necessidade dos alunos terem que copiar as aulas e, deste modo, poderem se concentrar em entender o conteúdo das mesmas.

*Recife, agosto de 2010.*  
*Leandro Chaves Rêgo, Ph.D.*

# Conteúdo

<b>Prefácio</b>	<b>i</b>
<b>1 Revisão Básica de Teoria dos Conjuntos</b>	<b>1</b>
1.1 Definição de Conjuntos e Exemplos . . . . .	1
1.2 Operações com Conjuntos . . . . .	4
1.3 Produto Cartesiano . . . . .	6
1.4 Conjunto das Partes . . . . .	7
1.5 Partição . . . . .	7
1.6 Função Indicadora . . . . .	8
<b>2 Técnicas de Contagem</b>	<b>10</b>
2.1 Introdução . . . . .	10
2.2 Métodos de Contagem . . . . .	10
2.2.1 Regra da Adição . . . . .	10
2.2.2 Regra da Multiplicação . . . . .	11
2.3 Aplicações em Grafos . . . . .	17
2.3.1 Grafos Não Direcionados . . . . .	17
2.3.2 Grafos Direcionados . . . . .	17
<b>3 Introdução à Probabilidade</b>	<b>19</b>
3.1 Experimento Aleatório . . . . .	19
3.2 Espaço Amostral . . . . .	19
3.3 Eventos e Coleção de Eventos . . . . .	20
3.3.1 Indução Matemática . . . . .	23
3.4 Frequências Relativas . . . . .	25
3.5 Axiomas de Kolmogorov . . . . .	26
3.5.1 Exemplos de Medidas de Probabilidade . . . . .	28
3.5.2 Propriedades de uma Medida de Probabilidade . . . . .	29
<b>4 Probabilidade Condicional</b>	<b>35</b>
4.1 Probabilidade Condicional . . . . .	35
4.2 Independência . . . . .	44

<b>5</b>	<b>Variável Aleatória Discreta</b>	<b>48</b>
5.1	Introdução . . . . .	48
5.2	Função de Distribuição Acumulada . . . . .	50
5.3	Tipos de Variável Aleatória . . . . .	53
5.4	Variável Aleatória Discreta . . . . .	53
5.4.1	Aleatória. . . . .	54
5.4.2	Bernoulli. . . . .	54
5.4.3	Binomial. . . . .	55
5.4.4	Hipergeométrica. . . . .	56
5.4.5	Geométrica. . . . .	57
5.4.6	Binomial Negativa ou Pascal. . . . .	58
5.4.7	Zeta ou Zipf. . . . .	59
5.4.8	Poisson. . . . .	59
5.5	Poisson como um Limite de Eventos Raros de Binomial . . . . .	60
<b>6</b>	<b>Esperança e Momentos de Variáveis Aleatórias Discretas</b>	<b>63</b>
6.1	O Conceito de Esperança . . . . .	63
6.1.1	Definição da Esperança . . . . .	63
6.2	Funções de Variáveis Aleatórias . . . . .	66
6.3	Esperança de Funções de Variáveis Aleatórias Discretas. . . . .	67
6.4	Propriedades da Esperança . . . . .	68
6.5	Momentos . . . . .	69
6.5.1	Momentos Centrais . . . . .	71
	<b>Referências Bibliográficas</b>	<b>73</b>

# Capítulo 1

## Revisão Básica de Teoria dos Conjuntos

### 1.1 Definição de Conjuntos e Exemplos

**Definição 1.1.1:** Um *conjunto* é uma coleção de elementos distintos<sup>1</sup> onde os elementos não são ordenados.

Esta definição intuitiva de um conjunto foi dada primeiramente por Georg Cantor (1845-1918), que criou a teoria dos conjuntos em 1895. Um conjunto pode ser especificado, listando seus elementos dentro de chaves. Por exemplo,

$$A = \{0, 1, 2, 3, 5, 8, 13\}, B = \{0, 1, 2, \dots, 1000\}.$$

Alternativamente, um conjunto pode ser especificado por uma regra que determina seus membros, como em:

$$C = \{x : x \text{ é inteiro e positivo}\} \text{ ou } D = \{x : x \text{ é par}\}.$$

Como em um conjunto a ordem dos elementos não importa, tem-se que:

$$\{1, 2, 3\} = \{2, 3, 1\}.$$

Se um dado elemento faz parte de um conjunto, diz-se que ele pertence ao conjunto e denota-se isso com símbolo  $\in$ . Por exemplo,  $2 \in D = \{x : x \text{ é par}\}$  ou  $3 \in E = \{x : x \text{ é primo}\}$ .

Por outro lado, se um dado elemento não faz parte de um conjunto, diz-se que ele não pertence ao conjunto e denota-se isso com o símbolo  $\notin$ . Por exemplo,  $3 \notin D = \{x : x \text{ é par}\}$  ou  $4 \notin E = \{x : x \text{ é primo}\}$ .

É preciso ter cuidado ao distinguir entre um elemento como 2 e o conjunto contendo somente este elemento  $\{2\}$ . Enquanto, tem-se  $2 \in F = \{2, 3, 5\}$ ,  $\{2\} \notin F = \{2, 3, 5\}$ , pois o conjunto contendo somente o elemento 2 não pertence à  $F$ .

---

<sup>1</sup>Na Estatística é comum se falar de conjuntos incluindo o caso onde seus elementos não são distintos. Por exemplo, o conjunto dos tempos de acesso a um banco de dados, o conjunto das notas de uma dada disciplina, entre outros, podem ter valores iguais, porém formalmente na teoria dos conjuntos os elementos de um conjunto devem ser distintos.

**Exemplo 1.1.2:** Seja  $G = \{2, \{3\}\}$ . Então,  $2 \in G$  e  $\{3\} \in G$ , porém  $3 \notin G$ .

O tamanho de um conjunto  $A$ ,  $\|A\|$ , é a quantidade de elementos que ele possui, a qual é chamada de sua *cardinalidade*. A cardinalidades pode ser *finita*, *infinita enumerável*, ou *infinita não-enumerável*. Um conjunto é finito quando existe uma função bijetiva cujo domínio é igual a este conjunto e a imagem é o conjunto dos inteiros não-negativos menores que um número finito; seus elementos podem ser contados, sendo possível exibir seu último elemento. Um conjunto infinito enumerável tem exatamente a mesma quantidade de elementos que os naturais, ou seja, existe uma função bijetiva cujo domínio é igual a este conjunto e a imagem é igual ao conjunto dos naturais. Um conjunto é enumerável se ele for finito ou infinito enumerável. Um conjunto é não-enumerável se ele não for enumerável. Por exemplo, os seguintes conjuntos são enumeráveis:

$$\begin{aligned} N_n &= \{0, 1, 2, \dots, n-1\}, \\ \mathcal{Z} &= \{x : x \text{ é um inteiro}\}, \\ \mathcal{Z}^+ &= \{x : x \text{ é um inteiro positivo}\}, \\ \mathcal{Q} &= \{x : x \text{ é racional}\}. \end{aligned}$$

Para notar que o conjunto dos números racionais é enumerável considere a seguinte matriz de números racionais. (Lembrando que um número  $x$  é racional se pode ser escrito sob a forma  $\frac{p}{q}$ , onde  $p$  e  $q$  são inteiros e  $q \neq 0$ .)

$$\begin{array}{cccc} 0/1 & & 0/2 & & 0/3 & \dots \\ & \swarrow & & \swarrow & & \\ 1/1 & & 1/2 & & 1/3 & \dots \\ & \swarrow & & \swarrow & & \\ 2/1 & & 2/2 & & 2/3 & \dots \\ & \swarrow & & \swarrow & & \\ 3/1 & & 3/2 & & 3/3 & \dots \\ & \swarrow & & \swarrow & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

Esta matriz contém todos os racionais não-negativos. Utilizando o método da diagonalização, os elementos da matriz são ordenados, sem repetição, da seguinte forma:

$$0/1, 1/1, 1/2, 2/1, 1/3, 3/1, \dots$$

Definindo-se uma correspondência  $f$  onde para cada racional não-negativo  $r$ ,  $f(r)$  representa a posição em que  $r$  aparece na sequência acima, tem-se que  $f$  é uma correspondência 1-1 entre os racionais não-negativos e os naturais. Por exemplo, temos que  $f(1/2) = 3$ ,  $f(3) = 6$ . Pode-se definir  $g$  no conjunto de todos os racionais tal que tal que  $g(r) = 2(f(r) - 1)$  se  $r > 0$ , e  $g(r) = 2f(|r|) - 1$  se  $r \leq 0$ . Desse modo,  $g(r)$  é um natural par se  $r$  for um racional positivo, e um natural ímpar, se  $r$  for um racional não-positivo. Portanto,  $g(r)$  é uma correspondência 1-1 entre os racionais e os naturais, o que implica que os racionais formam um conjunto enumerável.

Por outro lado, os conjuntos abaixo são não-enumeráveis:

$$\mathbb{R} = \{x : x \text{ é um número real}\},$$

$$(a, b) = \{x : a < x < b\}, \text{ onde } a < b,$$

$$[a, b] = \{x : a \leq x \leq b\}, \text{ onde } a < b.$$

Por exemplo, o matemático Georg Cantor mostrou que o intervalo  $[0, 1]$  não é enumerável com o seguinte argumento. A prova é feita por contradição. Suponha (para fins de argumentação) que o intervalo  $[0, 1]$  é infinito enumerável. Então, pode-se enumerar todos os números deste intervalo como uma sequência,  $(r_1, r_2, r_3, \dots)$  (os números não precisam estar em ordem). No caso de números com duas expansões decimais, como  $0,499\dots = 0,500\dots$ , escolhe-se aquele que acaba com nove. Suponha, por exemplo, que as expansões decimais do início da sequência são como se segue:

$r_1 = 0,5376547\dots$
$r_2 = 0,1199999\dots$
$r_3 = 0,5347824\dots$
$r_4 = 0,9870812\dots$
$r_5 = 0,3451243\dots$
$r_6 = 0,2136530\dots$
$r_7 = 0,3200985\dots$
$\dots$

Iremos agora construir um número real no intervalo  $[0, 1]$  que é diferente de todos os números na sequência anterior, o que gera uma contradição a hipótese inicial de que esta sequência contém todos os reais no intervalo  $[0, 1]$ . Constrói-se um número real  $x$  dentro do intervalo  $[0, 1]$  considerando o  $k$ -ésimo dígito depois da vírgula da expansão decimal de  $r_k$ . A partir desses dígitos nós definimos os dígitos do número  $x$  da seguinte forma:

- se o  $k$ -ésimo dígito de  $r_k$  é 5 então o  $k$ -ésimo dígito de  $x$  é 4.
- se o  $k$ -ésimo dígito de  $r_k$  não é 5 então o  $k$ -ésimo dígito de  $x$  é 5.

Para o exemplo anterior, isto resultará na seguinte expansão decimal para  $x$ :  $x = 0,4555554\dots$ . O número  $x$  construído desta forma é um número real dentro do intervalo  $[0, 1]$ . Por isso, devemos ter  $r_n = x$  para algum  $n$ , uma vez que presumimos que  $(r_1, r_2, r_3, \dots)$  enumera todos os números reais no intervalo  $[0, 1]$ . No entanto, por causa da maneira como foi definido o número  $x$ , este número difere na  $n$ -ésima posição de  $r_n$ , então  $x$  não está na sequência  $(r_1, r_2, r_3, \dots)$ .

É uma consequência direta deste resultado que o conjunto  $\mathbb{R}$  de todos os números reais é não-enumerável. Se  $\mathbb{R}$  fosse enumerável, poderíamos enumerar todos os números reais em uma sequência, e então obter uma sequência enumerando  $[0, 1]$  através da remoção de todos os números reais fora deste intervalo. Mas nós acabamos de mostrar que esta última lista não pode existir.

Existem dois conjuntos especiais que nos interessarão. Em muitos problemas nos dedicaremos a estudar um conjunto definido de objetos, e não outros. Por exemplo, em alguns

problemas podemos nos interessar pelo conjunto dos números naturais; ou em outros problemas pelo conjuntos dos números reais; ou ainda por todas as peças que saem de uma linha produção durante um período de 24h, etc. O conjunto que contém todos os elementos que queremos considerar é chamado de *conjunto universo* e é denotado por  $\Omega$ . Por outro lado, o conjunto especial que não possui elementos é chamado de *conjunto vazio* e é denotado por  $\emptyset$ . Este conjunto tem cardinalidade 0 e portanto é finito. Por exemplo,

$$\emptyset = \{\} = \{x : x \in \mathbb{R} \text{ e } x < x\} \text{ ou } \emptyset = (a, a).$$

Dois conjuntos  $A$  e  $B$  podem ser relacionados através da relação de inclusão, denotada por  $A \subseteq B$ , e lida  $A$  é um subconjunto de  $B$  ou  $B$  contém  $A$ , quando todo elemento de  $A$  é também elemento de  $B$ . Diz-se que  $A$  é um *subconjunto próprio* de  $B$  quando se tem  $A \subseteq B$ ,  $A \neq \emptyset$ , e  $B \not\subseteq A$ . Se  $A$  é subconjunto de  $B$ , então  $B$  é chamado um *superconjunto* de  $A$ . Diz-se que  $A$  e  $B$  são iguais se e somente se  $A \subseteq B$  e  $B \subseteq A$ . Se  $A \subseteq B$ , então também pode-se dizer que  $B \supseteq A$ .

A relação  $\subseteq$  possui as propriedades de (i) *reflexividade* ( $A \subseteq A$ ); (ii) *transitividade* ( $A \subseteq B, B \subseteq C \rightarrow A \subseteq C$ ); e *anti-simetria* ( $A \subseteq B, B \subseteq A \rightarrow A = B$ ). Contudo, ela não é uma relação completa, ou seja, não é verdade que, para todos os conjuntos  $A$  e  $B$ , ou  $A \subseteq B$ , ou  $B \subseteq A$ . Também é fácil verificar que  $\emptyset \subseteq A$  e  $A \subseteq \Omega$  para todo conjunto  $A$ . A relação de igualdade satisfaz as seguintes propriedades: (i) *reflexividade* ( $A = A$ ); (ii) *simetria* ( $A = B \Rightarrow B = A$ ); e (iii) *transitividade* ( $A = B, B = C \Rightarrow A = C$ ).

## 1.2 Operações com Conjuntos

Queremos estudar a importante idéia de combinar conjuntos dados, a fim de formamos um novo conjunto. Conjuntos podem ser transformados através das seguintes operações Booleanas:

1. Complementação:  $A^c = \{\omega \in \Omega : \omega \notin A\}$ . Observe que de acordo com esta definição, para todo  $\omega \in \Omega$  e todo conjunto  $A$ , não existe outra opção além de  $\omega \in A$  ou  $\omega \in A^c$ , além disso não pode ser verdade que  $\omega \in A$  e  $\omega \in A^c$  simultaneamente.
2. União:  $A \cup B = \{\omega : \omega \in A \text{ ou } \omega \in B\}$
3. Intersecção:  $A \cap B = \{\omega : \omega \in A \text{ e } \omega \in B\}$
4. Diferença:  $A - B = A \cap B^c = \{\omega : \omega \in A \text{ e } \omega \notin B\}$

Se  $A \cap B = \emptyset$ , então  $A$  e  $B$  não tem nenhum elemento em comum, e nós dizemos que  $A$  e  $B$  são *disjuntos*.

**Exemplo 1.2.1:** Seja  $\Omega = \{0, 1, 2, 3, 4, 5, 6, 7\}$ ,  $A = \{0, 1, 5\}$  e  $B = \{1, 2, 3, 4\}$ . Então segue que  $A^c = \{2, 3, 4, 6, 7\}$ ,  $A \cup B = \{0, 1, 2, 3, 4, 5\}$ ,  $A \cap B = \{1\}$ ,  $A - B = \{0, 5\}$ . ■



**Exemplo 1.2.2:** Sejam  $A, B, C$  e  $D$  subconjuntos do conjunto universo  $\Omega$  tal que  $A \cup B = \Omega$ ,  $C \cap D = \emptyset$ ,  $A \subseteq C$  e  $B \subseteq D$ . Prove que  $A = C$  e  $B = D$ .

**Solução:** Basta provar que  $C \subseteq A$  e  $D \subseteq B$ . Seja  $\omega \in C$ , então como  $C \cap D = \emptyset$ , tem-se que  $\omega \notin D$ . Logo, como  $B \subseteq D$ , segue que  $\omega \notin B$ . Mas como  $A \cup B = \Omega$ , tem-se que  $\omega \in A$ . Portanto,  $C \subseteq A$ .

Para provar que  $D \subseteq B$ , seja  $\omega \in D$ , então como  $C \cap D = \emptyset$ , tem-se que  $\omega \notin C$ . Logo, como  $A \subseteq C$ , segue que  $\omega \notin A$ . Mas como  $A \cup B = \Omega$ , tem que  $\omega \in B$ . Portanto,  $D \subseteq B$ .

Relações e propriedades das operações Booleanas incluem as seguintes:

1. Idempotência:  $(A^c)^c = A$

**Prova:** Suponha que  $\omega \in (A^c)^c$ . Então,  $\omega \notin A^c$ , o que por sua vez implica que  $\omega \in A$ , ou seja,  $(A^c)^c \subseteq A$ . Agora suponha que  $\omega \in A$ , então  $\omega \notin A^c$ , o que por sua vez implica que  $\omega \in (A^c)^c$ , ou seja,  $A \subseteq (A^c)^c$ . Portanto,  $(A^c)^c = A$ . ■

2. Comutatividade (Simetria):  $A \cup B = B \cup A$  e  $A \cap B = B \cap A$

**Prova:** Suponha que  $\omega \in A \cup B$ . Então,  $\omega \in A$ , ou  $\omega \in B$ , o que por sua vez implica que  $\omega \in B \cup A$ , ou seja,  $A \cup B \subseteq B \cup A$ . Agora suponha que  $\omega \in B \cup A$ . Então,  $\omega \in B$ , ou  $\omega \in A$ , o que por sua vez implica que  $\omega \in A \cup B$ , ou seja,  $B \cup A \subseteq A \cup B$ . Portanto,  $A \cup B = B \cup A$ .

A prova para o caso da intersecção é análoga e deixada como Exercício. ■

3. Associatividade:  $A \cup (B \cap C) = (A \cup B) \cap C$  e  $A \cap (B \cup C) = (A \cap B) \cup C$

**Prova:** Exercício. ■

4. Distributividade:  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$  e  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

**Prova:** Exercício. ■

5. Leis de De Morgan:  $(A \cup B)^c = A^c \cap B^c$  e  $(A \cap B)^c = A^c \cup B^c$ .

**Prova:** Suponha que  $\omega \in (A \cup B)^c$ . Então,  $\omega \notin (A \cup B)$ , o que por sua vez implica que  $\omega \notin A$  e  $\omega \notin B$ . Logo,  $\omega \in A^c$  e  $\omega \in B^c$ , ou seja,  $\omega \in (A^c \cap B^c)$ . Então,  $(A \cup B)^c \subseteq (A^c \cap B^c)$ . Agora suponha que  $\omega \in (A^c \cap B^c)$ . Então,  $\omega \in A^c$  e  $\omega \in B^c$ , o que por sua vez implica que  $\omega \notin A$  e  $\omega \notin B$ . Logo,  $\omega \notin (A \cup B)$ , ou seja,  $\omega \in (A \cup B)^c$ . Então,  $(A^c \cap B^c) \subseteq (A \cup B)^c$ . Portanto,  $(A^c \cap B^c) = (A \cup B)^c$ .

A prova da outra Lei de Morgan é análoga e deixada como Exercício. ■

As Leis de De Morgan permitem que se possa expressar uniões em termos de intersecções e complementos e intersecções em termos de uniões e complementos.

Uniões e intersecções podem ser estendidas para coleções arbitrárias de conjuntos. Seja  $\mathcal{I}$  um conjunto qualquer. Este conjunto  $\mathcal{I}$  será utilizado para indexar, ou seja, identificar através de um único símbolo os conjuntos na coleção arbitrária de interesse e desse modo

simplificar a notação utilizada. Por exemplo, se  $\mathcal{I} = \{1, 5, 7\}$ , então  $\cup_{i \in \mathcal{I}} A_i = A_1 \cup A_5 \cup A_7$ ; ou, se  $\mathcal{I} = \mathcal{N}$ , então  $\cap_{i \in \mathcal{N}} A_i = A_1 \cap A_2 \cap \dots \cap A_n \dots$ .

De modo análogo ao caso de dois conjuntos, define-se:

$$\cup_{i \in \mathcal{I}} A_i = \{\omega \in \Omega : \omega \text{ pertence a pelo menos um dos conjuntos } A_i, \text{ onde } i \in \mathcal{I}, \}$$

e

$$\cap_{i \in \mathcal{I}} A_i = \{\omega \in \Omega : \omega \text{ pertence a todo } A_i, \text{ onde } i \in \mathcal{I}. \}$$

Se  $\mathcal{I}$  for um conjunto enumerável, diz-se que  $\cup_{i \in \mathcal{I}} A_i$ , respectivamente,  $\cap_{i \in \mathcal{I}} A_i$ , é uma união, respectivamente intersecção, enumerável de conjuntos.

Por exemplo, se  $\Omega = 0, 1, 2, \dots$ ,  $\mathcal{I}$  é o conjunto de inteiros positivos divisíveis por 3 e  $N_\alpha = \{0, 1, 2, \dots, \alpha - 1\}$ , então

$$\cup_{\alpha \in \mathcal{I}} N_\alpha = \Omega \text{ e } \cap_{\alpha \in \mathcal{I}} N_\alpha = N_3.$$

**Exemplo 1.2.3:** Se  $A_i = [1, 2 + \frac{1}{i}]$ ,  $i \in \mathcal{N}$ , então  $\cup_{i \in \mathcal{N}} A_i = [1, 3)$  e  $\cap_{i \in \mathcal{N}} A_i = [1, 2]$ .

### 1.3 Produto Cartesiano

Uma construção diferente de um novo conjunto  $C$  dados conjuntos  $A$  e  $B$  é feita através do produto Cartesiano de conjuntos. Por um par ordenado de elementos  $(a, b)$ , onde  $a \in A$  e  $b \in B$ , descreve-se um objeto tal que  $(a_1, b_1) = (a_2, b_2) \Leftrightarrow a_1 = a_2$  e  $b_1 = b_2$ .

**Definição 1.3.1: Produto Cartesiano.** O produto Cartesiano  $A \times B$  de dois conjuntos dados  $A$  e  $B$  é o conjunto de todos os pares ordenados de elementos, onde o primeiro pertence à  $A$  e o segundo pertence à  $B$ :

$$A \times B = \{(a, b) : a \in A, b \in B\}.$$

■

Por exemplo, se  $A = \{1, 2, 3\}$  e  $B = \{c, d\}$ , então:

$$A \times B = \{(1, c), (1, d), (2, c), (2, d), (3, c), (3, d)\}, \text{ e}$$

$$B \times A = \{(c, 1), (c, 2), (c, 3), (d, 1), (d, 2), (d, 3)\}.$$

A noção de produto cartesiano pode ser estendida da seguinte maneira: Se  $A_1, \dots, A_n$  forem conjuntos, então,

$$A_1 \times A_2 \times \dots \times A_n = \{(a_1, a_2, \dots, a_n) : a_i \in A_i\},$$

ou seja, o conjunto de todas as ênuplas ordenadas.

Um caso especial importante surge quando consideramos o produto cartesiano de um conjunto por ele próprio, isto é,  $A \times A$ . Exemplos disso surgem quando tratamos do plano euclidiano,  $\mathbb{R} \times \mathbb{R}$ , onde  $\mathbb{R}$  é o conjunto de todos os números reais, e do espaço euclidiano tridimensional, representado por  $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$ .

## 1.4 Conjunto das Partes

**Definição 1.4.1:** Dado um conjunto qualquer  $A$ , pode-se definir um outro conjunto, conhecido como *conjunto das partes de  $A$* , e denotado por  $2^A$ , cujos elementos são subconjuntos de  $A$ . ■

**Exemplo 1.4.2:** Seja  $A = \{1, 2, 3\}$ , então temos que

$$2^A = \{\emptyset, A, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}\}.$$

■

O próximo Teorema conhecido como *Teorema de Cantor*, prova que a cardinalidade do conjunto das partes de qualquer conjunto dado  $A$  é maior que a cardinalidade de  $A$ .

**Teorema 1.4.3:** Se  $A$  é um conjunto e  $2^A$  é o conjunto das partes de  $A$ , não existe uma função  $f : A \rightarrow 2^A$  que seja sobrejetiva.

**Prova:** Recorde que uma função  $g : D \rightarrow I$  é sobrejetiva se para todo  $y \in I$ , existe  $x \in D$  tal que  $g(x) = y$ . Suponha por contradição, que existe uma função sobrejetiva  $f : A \rightarrow 2^A$ .

Defina o conjunto,  $B = \{x \in A : x \notin f(x)\}$ . Como  $f$  por suposição é sobrejetiva e  $B \in 2^A$ , temos que existe  $b \in A$  tal que  $f(b) = B$ . Temos que considerar dois casos:  $b \in B$  ou  $b \in B^c$ . Se  $b \in B$ , então  $b \notin f(b)$ . Mas como  $B = f(b)$ , temos que  $b \notin B$ , absurdo. Se  $b \in B^c$ , então  $b \in f(b)$ . Mas como  $B = f(b)$ , temos que  $b \in B$ , absurdo. ■

## 1.5 Partição

Intuitivamente, uma partição de um conjunto universo é uma maneira de distribuir os elementos deste conjunto em uma coleção arbitrária de subconjuntos. Formalmente, tem-se a seguinte definição:

**Definição 1.5.1:** Dado um conjunto universo  $\Omega$ , uma partição  $\Pi = \{A_\alpha, \alpha \in \mathcal{I}\}$  de  $\Omega$  é uma coleção de subconjuntos de  $\Omega$  (neste caso, indexados por  $\alpha$  que toma valores no conjunto de índices  $\mathcal{I}$ ) e satisfaz:

P1. Para todo  $\alpha \neq \beta$ ,  $A_\alpha \cap A_\beta = \emptyset$ ;

P2.  $\cup_{\alpha \in \mathcal{I}} A_\alpha = \Omega$ .

■

Deste modo os conjuntos de uma partição são disjuntos par a par e cobrem todo o conjunto universo. Portanto, cada elemento  $\omega \in \Omega$  pertence a um, e somente um, dos conjuntos  $A_\alpha$  de uma partição.

**Exemplo 1.5.2:** Se  $\Omega = \{1, 2, 3, 4\}$ , então  $\{A_1, A_2\}$ , onde  $A_1 = \{1, 2, 3\}$  e  $A_2 = \{4\}$ , é uma partição de  $\Omega$ . ■

**Exemplo 1.5.3:** A coleção de intervalos  $\{(n, n + 1] : n \in \mathbb{Z}\}$  é uma partição dos números reais  $\mathbb{R}$ . ■

## 1.6 Função Indicadora

É sempre conveniente representar um conjunto  $A$  por uma função  $I_A$  tendo domínio (conjunto dos argumentos da função)  $\Omega$  e contra-domínio (conjunto dos possíveis valores da função) binário  $\{0, 1\}$ .

**Definição 1.6.1: Função Indicadora.** A função indicadora  $I_A : \Omega \rightarrow \{0, 1\}$  de um conjunto  $A$  é dada por

$$I_A(\omega) = \begin{cases} 1 & \text{se } \omega \in A, \\ 0 & \text{se } \omega \notin A. \end{cases}$$

■

Note que podemos determinar  $A$  a partir de sua função indicadora:  $A = \{\omega : I_A(\omega) = 1\}$ .

**Exemplo 1.6.2:** Se  $I_A(\omega)$  for identicamente igual a 1, ou seja,  $I_A(\omega) = 1, \forall \omega \in \Omega$ , então  $A$  é o conjunto universo  $\Omega$ . Se  $I_A(\omega)$  for identicamente igual a 0, então  $A$  é o conjunto vazio  $\emptyset$ . Se  $I_A(\omega)$  for igual a 1 somente quando  $\omega = \omega_0$ , então  $A$  é o conjunto  $\{\omega_0\}$  que contém somente o elemento  $\omega_0$ . ■

Note que existe uma correspondência 1-1 entre conjuntos e suas funções indicadoras:

$$A = B \Leftrightarrow (\forall \omega \in \Omega) I_A(\omega) = I_B(\omega).$$

O fato que conjuntos são iguais se, e somente se, suas funções indicadoras forem idênticas nos permitem explorar a aritmética de funções indicadoras:

$$I_{A^c} = 1 - I_A,$$

$$A \subseteq B \Leftrightarrow I_A \leq I_B,$$

$$I_{A \cap B} = \min(I_A, I_B) = I_A I_B,$$

$$I_{A \cup B} = \max(I_A, I_B) = I_A + I_B - I_{A \cap B},$$

$$I_{A-B} = \max(I_A - I_B, 0) = I_A I_{B^c},$$

para construir argumentos rigorosos no que se refere a relação entre conjuntos. Ou seja, nós transformamos proposições sobre conjuntos em proposições sobre funções indicadoras e podemos então utilizar nossa familiaridade com álgebra para resolver perguntas menos familiares sobre conjuntos.

**Exemplo 1.6.3:** Utilizando funções indicadoras, verifique que  $A \subseteq B \Leftrightarrow B^c \subseteq A^c$ .

**Solução:** Temos que

$$A \subseteq B \Leftrightarrow I_A \leq I_B \Leftrightarrow 1 - I_A \geq 1 - I_B \Leftrightarrow I_{A^c} \geq I_{B^c} \Leftrightarrow B^c \subseteq A^c.$$

■

**Exemplo 1.6.4:** As seguintes questões não estão relacionadas umas com as outras.

- a. Se  $I_A I_B$  for identicamente igual a zero, o que sabemos a respeito da relação entre  $A$  e  $B$ ?
- b. Se  $A \cap B^c = B \cap A^c$ , o que sabemos a respeito da relação entre  $A$  e  $B$ ?
- c. Se  $I_A^2 + I_B^2$  for identicamente igual a 1, o que podemos concluir sobre  $A$  e  $B$ ?
- d. Se  $I_A - I_B$  for identicamente igual a 1, , o que podemos concluir sobre  $A$  e  $B$ ?
- e. Se  $A \cap B = B \cup A$ , o que podemos concluir sobre  $A$  e  $B$ ?

**Solução:** Exercício. ■

# Capítulo 2

## Técnicas de Contagem

### 2.1 Introdução

Neste capítulo estudaremos alguns métodos de contagem, também conhecidos como métodos de análise combinatória. A relevância deste estudo se deve ao fato que, como veremos adiante, em muitos casos onde o conjunto de resultados possíveis de um experimento aleatório é finito a probabilidade dos eventos é proporcional a sua cardinalidade. Deste modo, será importante ter alguma familiaridade com estes métodos. Embora conjuntos pequenos possam ser contados exaustivamente (força-bruta), mesmo conjuntos com tamanho moderado podem ser difíceis de contar sem a utilização de técnicas matemáticas.

### 2.2 Métodos de Contagem

#### 2.2.1 Regra da Adição

Suponha que um procedimento, designado por 1, possa ser realizado de  $n_1$  maneiras. Admita-se que um segundo procedimento, designado por 2, possa ser realizado de  $n_2$  maneiras. Além disso, suponha que não seja possível que ambos os procedimentos 1 e 2 sejam realizados em conjunto. Então, o número de maneiras pelas quais poderemos realizar ou 1 ou 2 será  $n_1 + n_2$ .

Esta regra também pode ser estendida da seguinte maneira: Se existirem  $k$  procedimentos e o  $i$ -ésimo procedimento puder ser realizado de  $n_i$  maneiras,  $i = 1, 2, \dots, k$ , então, o número de maneiras pelas quais poderemos realizar ou o procedimento 1, ou o procedimento 2,  $\dots$ , ou o procedimento  $k$ , é dado por  $n_1 + n_2 + \dots + n_k$ , supondo que dois quaisquer deles não possam ser realizados conjuntamente.

**Exemplo 2.2.1:** Suponha que estejamos planejando uma viagem e devamos escolher entre o transporte por ônibus ou por trem. Se existirem três rodovias e duas ferrovias, então existirão  $3 + 2 = 5$  caminhos disponíveis para a viagem. ■

### 2.2.2 Regra da Multiplicação

Suponha que um procedimento designado por 1 possa ser executado de  $n_1$  maneiras. Admita-se que um segundo procedimento, designado por 2, possa ser executado de  $n_2$  maneiras. Suponha também que cada maneira de executar 1 possa ser seguida por qualquer maneira para executar 2. Então o procedimento formado por 1 seguido de 2 poderá ser executado de  $n_1 \cdot n_2$  maneiras.

Obviamente, esta regra pode ser estendida a qualquer número finito de procedimentos. Se existirem  $k$  procedimentos e o  $i$ -ésimo procedimento puder ser executado de  $n_i$  maneiras,  $i = 1, 2, \dots, k$ , então o procedimento formado por 1, seguido por 2,  $\dots$ , seguido pelo procedimento  $k$ , poderá ser executado de  $n_1 \cdot n_2 \cdots n_k$  maneiras.

**Exemplo 2.2.2:** Uma peça manufaturada deve passar por 3 estações de controle. Em cada estação, a peça é inspecionada para determinada característica e marcada adequadamente. Na primeira estação, três classificações são possíveis, enquanto nas duas últimas, quatro classificações são possíveis. Consequentemente, existem  $3 \cdot 4 \cdot 4 = 48$  maneiras pelas quais uma peça pode ser marcada. ■

**Exemplo 2.2.3:** Quantos divisores inteiros e positivos possui o número 360? Quantos desses divisores são pares? Quantos são ímpares? Quantos são quadrados perfeitos?

**Solução:**  $360 = 2^3 \times 3^2 \times 5$ . Os divisores inteiros e positivos de 360 são os números da forma:  $2^a \times 3^b \times 5^c$ , onde  $a \in \{0, 1, 2, 3\}$ ,  $b \in \{0, 1, 2\}$ , e  $c \in \{0, 1\}$ . Portanto, existem  $4 \times 3 \times 2 = 24$  maneiras de escolher os expoentes  $a, b, c$ . Logo há 24 divisores.

Para o divisor ser par,  $a$  não pode ser zero. Então, existem  $3 \times 3 \times 2 = 18$  divisores pares. Por outro lado, para o divisor ser ímpar,  $a$  tem que ser zero. Logo, existem  $1 \times 3 \times 2 = 6$  divisores ímpares. Por fim para o divisor ser quadrado perfeito, os expoentes tem que ser pares. Logo, existem  $2 \times 2 \times 1 = 4$  divisores quadrados perfeitos. ■

**Exemplo 2.2.4:** De quantos modos o número 720 pode ser decomposto em um produto de dois inteiros positivos? Aqui consideramos, naturalmente,  $8 \times 90$  como sendo o mesmo produto que  $90 \times 8$ . E o número 144?

**Solução:**  $720 = 2^4 \times 3^2 \times 5$ . Os divisores inteiros e positivos de 720 são os números da forma:  $2^a \times 3^b \times 5^c$ , onde  $a \in \{0, 1, 2, 3, 4\}$ ,  $b \in \{0, 1, 2\}$ , e  $c \in \{0, 1\}$ . Portanto, existem  $5 \times 3 \times 2 = 30$  maneiras de escolher os expoentes  $a, b, c$ . Logo há 30 divisores. Observe que como 720 não é um quadrado perfeito, para cada divisor  $x$  de 720 existe um outro divisor  $y \neq x$  de 720 tal que  $x \times y = 720$ . Portanto, cada produto contém dois divisores diferentes de 720. Como existem 30 divisores, existem 15 produtos diferentes.

$144 = 2^4 \times 3^2$ . Seguindo o mesmo raciocínio anterior, temos  $5 \times 3 = 15$  divisores de 144. Note que  $144 = 12^2$  e este constitui um produto de inteiros positivos que é igual a 144. Para os demais produtos sempre temos que eles contém dois inteiros positivos diferentes que são divisores de 144. Como existem 14 divisores de 144 diferentes de 12, temos que existem 7 produtos envolvendo estes divisores. Logo, temos um total de 8 produtos diferentes. ■

**Exemplo 2.2.5:** O conjunto  $A$  possui 4 elementos e, o conjunto  $B$ , 7 elementos. Quantas funções  $f : A \rightarrow B$  existem? Quantas delas são injetoras?

**Solução:** Note que para cada elemento de  $A$  temos 7 opções de valores diferentes. Como  $A$  contém 4 elementos, existem  $7 \times 7 \times 7 \times 7 = 7^4$  funções diferentes. Recorde que uma função é injetora se  $f(a) \neq f(b)$  sempre que  $a \neq b$ . Portanto, não podemos repetir o mesmo elemento de  $B$  como imagem de dois elementos de  $A$ , logo existem  $7 \times 6 \times 5 \times 4 = 840$  funções injetoras. ■

**Exemplo 2.2.6:** Em uma banca há 5 exemplares iguais da “Veja”, 6 exemplares iguais da “Época” e 4 exemplares iguais da “Isto é”. Quantas coleções não-vazias de revistas dessa banca podemos formar?

**Solução:** Note que cada coleção de revistas vai ser composta por  $a$  revistas Veja,  $b$  revistas Época, e  $c$  revistas Isto é, onde  $0 \leq a \leq 5$ ,  $0 \leq b \leq 6$ ,  $0 \leq c \leq 4$ , e pelo menos 1 de  $a$ ,  $b$ , ou  $c$  é diferente de zero. Então, temos  $6 \times 7 \times 5 - 1 = 210 - 1 = 209$  diferentes coleções não-vazias destas revistas. ■

### Amostragem ou Escolhas com Reposição

Dado um conjunto com  $n$  elementos distintos, o número  $\mu_{n,r}$  de maneiras de selecionar uma sequência distinta de comprimento  $r$  escolhida desse conjunto com repetidas seleções do mesmo elemento sendo permitida (*amostragem com repetição*) é dada por  $n^r$ , já que estamos repetindo o mesmo procedimento  $r$  vezes, e cada procedimento tem  $n$  maneiras de ser executado.

Este resultado também se aplica ao número de resultados possíveis em  $r$  jogadas de uma moeda ( $n = 2$ ), ou de um dado ( $n = 6$ ), ou o número de bytes ( $r = 8, n = 2$ ) (Um byte é uma sequência ordenada de comprimento 8 de 0's e 1's).

**Exemplo 2.2.7: Número de Sequências Binárias ou Subconjuntos.** O número de sequências binárias de comprimento  $r$  é igual a  $2^r$  pois neste caso temos para cada posição  $i$  da sequência  $n_i = 2$ . O número de subconjuntos de um dado conjunto  $|A| = r$  pode ser determinado enumerando  $A = \{a_1, a_2, a_3, \dots, a_r\}$  e descrevendo cada subconjunto  $B$  de  $A$  por uma sequência binária

$$(b_1, b_2, \dots, b_r)$$

, onde  $b_i = 1$  se  $a_i \in B$  e  $b_i = 0$ , caso contrário. Como existem  $2^r$  destas sequências, então existem  $2^r$  subconjuntos de um conjunto de  $r$  elementos. Portanto, se  $|A| = r$ , o conjunto das partes de  $A$ , possui  $2^r$  elementos, o que explica a notação exponencial do conjunto das partes. ■

### Amostragem ou Escolhas sem Reposição

Dado um conjunto com  $n$  elementos distintos, o número  $(n)_r$  de maneiras de selecionar uma sequência distinta de comprimento  $r$  escolhida desse conjunto com repetidas seleções do mesmo elemento não sendo permitida (*amostragem sem repetição*) é dada por

$$(n)_r = n(n-1) \cdots (n-r+1) = \prod_{i=0}^{r-1} (n-i)$$



, já que no primeiro procedimento (escolha do primeiro elemento da sequência) temos  $n$  maneiras de executá-lo, no segundo procedimento (escolha do segundo elemento da sequência) temos  $n - 1$  maneiras de executá-lo,  $\dots$ , e no  $r$ -ésimo e último procedimento (escolha do  $r$ -ésimo elemento da sequência) temos  $n - r + 1$  maneiras de executá-lo. Este número de sequências é também chamado na literatura do número de arranjos quando temos  $n$  elementos distintos e queremos escolher  $r$  deles onde a ordem de escolha é importante.

Um caso particular de amostragem sem reposição é quando queremos saber o número de permutações de um conjunto de  $n$  elementos distintos. Neste caso temos que  $r = n$ , então o número de permutações é dado por

$$n! = (n)_n = n(n - 1) \cdots 1,$$

onde  $n!$  é conhecida como *função fatorial*. Em termos, de função fatorial, nós podemos escrever:

$$(n)_r = \frac{n!}{(n - r)!}.$$

Propriedades da função fatorial  $n!$  incluem as seguintes:

$$0! = 1! = 1 \text{ e } n! = n(n - 1)!.$$

**Exemplo 2.2.8:** Se  $A$  é um conjunto de  $n$  elementos, quantas são as funções  $f : A \rightarrow A$  bijetoras?

**Solução:** Temos que garantir que cada elemento de  $A$  tem uma imagem diferente. Como  $A$  é finito e tem  $n$  elementos, garante-se deste modo que  $f$  também é sobrejetora e, portanto, bijetora. Então, o primeiro elemento de  $A$  tem  $n$  opções, o segundo  $n - 1$  opções, até que o último elemento de  $A$  tem somente uma opção disponível. Portanto, existem  $n!$  funções bijetoras  $f : A \rightarrow A$ . ■

**Exemplo 2.2.9:** De quantos modos é possível colocar  $r$  rapazes e  $m$  moças em fila de modo que as moças permaneçam juntas?

**Solução:** Primeiro temos  $r + 1$  opções de escolher o lugar das moças. Em seguida, temos  $r!$  maneiras de escolher a posição dos rapazes entre si, e  $m!$  maneiras de escolher a posição das moças entre si. Portanto, temos  $(r + 1)r!m!$  modos diferentes de escolha. ■

**Exemplo 2.2.10:** Quantas são as permutações simples dos números  $1, 2, \dots, 10$  nas quais o elemento que ocupa o lugar de ordem  $k$ , da esquerda para a direita, é sempre maior que  $k - 3$ ?

**Solução:** Começamos escolhendo os números da direita para esquerda. Observe que o número no lugar de ordem 10, tem que ser maior que 7, portanto existem 3 opções. O número no lugar de ordem 9, tem que ser maior que 6, existem, portanto, 3 opções visto que um dos números maiores que 6 já foi utilizado na última posição. De maneira similar pode-se ver que existem 3 opções para os números que ocupam do terceiro ao oitavo lugar. O número no lugar de ordem 2, tem somente 2 opções, pois oito números já foram escolhidos anteriormente. Finalmente, resta apenas um número para o lugar de ordem 1. Portanto, existem  $2 \times 3^8$  permutações deste tipo. ■

**Enumeração de Conjuntos: Coeficientes Binomiais**

O número de conjuntos, ou coleções não ordenadas, de tamanho  $r$  escolhidas de um conjunto universo de tamanho  $n$ , onde, como apropriado para conjuntos, não é permitido a duplicação de elementos (amostragem sem repetição), é dado pelo *coeficiente binomial*:

$$\binom{n}{r} = \frac{(n)_r}{r!} = \frac{n!}{(n-r)!r!}.$$

Para verificar isto, note que o número de coleções ordenadas de tamanho  $r$  sem repetição é  $(n)_r$ . Como os elementos de cada sequência de comprimento  $r$  são distintos, o número de permutações de cada sequência é  $r!$ . Porém, utilizando a regra da multiplicação, o procedimento de escolhermos uma coleção ordenada de  $r$  termos sem repetição é igual a primeiro escolher uma coleção não-ordenada de  $r$  termos sem repetição e depois escolhermos uma ordem para esta coleção não ordenada, ou seja, temos que

$$(n)_r = \binom{n}{r} \cdot r!,$$

de onde segue o resultado.

O coeficiente binomial tem as seguintes propriedades:

$$\binom{n}{r} = \binom{n}{n-r}, \binom{n}{0} = 1, \binom{n}{1} = n, \binom{n}{r} = 0 \text{ se } n < r.$$

Note que o coeficiente binomial é também igual ao número de subconjuntos de tamanho  $r$  que pode ser formado de um conjunto de  $n$  elementos. Como já vimos que, o número total de subconjuntos de um conjunto de tamanho  $n$  é  $2^n$ , temos que

$$2^n = \sum_{r=0}^n \binom{n}{r}.$$

Os números  $\binom{n}{r}$  são chamados de coeficientes binomiais, porque eles aparecem como coeficientes na expressão binomial  $(a+b)^n$ . Se  $n$  for um inteiro positivo,  $(a+b)^n = (a+b)(a+b)\cdots(a+b)$ . Quando a multiplicação tiver sido executada, cada termo será formado de  $k$  elementos de  $a$  e de  $(n-k)$  elementos de  $b$ , para  $k = 0, 1, 2, \dots, n$ . Mas quantos termos da forma  $a^k b^{n-k}$  existirão? Simplesmente contaremos o número de maneiras possíveis de escolher  $k$  dentre os  $n$  elementos  $a$ , deixando de lado a ordem (onde o  $i$ -ésimo elemento  $a$  corresponde ao  $i$ -ésimo fator do produto acima). Mas isto é justamente dado por  $\binom{n}{k}$ . Daí obtém-se o que é conhecido como o *Teorema Binomial*:

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}.$$

**Exemplo 2.2.11:** Dentre oito pessoas, quantas comissões de três membros podem ser escolhidas, desde que duas comissões sejam a mesma comissão se forem constituídas pelas mesmas pessoas (não se levando em conta a ordem em que sejam escolhidas)? A resposta é dada por  $\binom{8}{3} = 56$  comissões possíveis. ■

**Exemplo 2.2.12:** Com oito bandeiras diferentes, quantos sinais feitos com três bandeiras diferentes se podem obter? Este problema parece-se muito com o exemplo anterior, mas neste caso a ordem acarreta diferença e por isso temos  $(8)_3 = 336$  sinais. ■

**Exemplo 2.2.13:** Um grupo de oito pessoas é formado de cinco homens e três mulheres. Quantas comissões de três pessoas podem ser constituídas, incluindo exatamente dois homens? Aqui deveremos fazer duas coisas, escolher dois homens (dentre cinco) e escolher uma mulher (dentre três). Daí obtemos como número procurado  $\binom{5}{2}\binom{3}{1} = 30$  comissões. ■

**Exemplo 2.2.14:** Quantos sequências binárias de comprimento  $n$  contém no máximo três números 1? Neste caso, temos quatro casos possíveis: todas sequências que não contém 1, todas sequências que contém apenas um número 1, todas sequências que contém dois números 1, e todas as sequências que contém três números 1. Para  $0 \leq r \leq n$ , temos que existem exatamente  $\binom{n}{r}$  sequências binárias com  $r$  números 1. Portanto, pela regra da adição temos que existem

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \binom{n}{3}$$

sequências binárias de comprimento  $n$  contendo no máximo três números 1. ■

**Exemplo 2.2.15:** Quantas sequências de cara e coroa de comprimento  $n$  contém pelo menos 1 cara? Neste caso, note que apenas uma sequência não contém nenhuma cara (a sequência que contém apenas coroa). Como o número total de sequências de cara e coroa de comprimento  $n$  é igual a  $2^n$ , temos então  $2^n - 1$  sequências de comprimento  $n$  contendo pelo menos uma cara. ■

**Exemplo 2.2.16:** Determine o coeficiente de  $x^3$  no desenvolvimento de  $(x^4 - \frac{1}{x})^7$ .

**Solução:** O termo genérico do desenvolvimento é

$$\binom{7}{k} (x^4)^k \left(-\frac{1}{x}\right)^{7-k} = (-1)^{7-k} \binom{7}{k} x^{5k-7}.$$

Portanto, temos o termo  $x^3$  se  $5k - 7 = 3$ , o que implica que  $k = 2$ . Logo, o coeficiente de  $x^3$  é  $(-1)^5 \binom{7}{2} = -21$ . ■

### Contagem Multinomial

Considere que temos  $r$  tipos de elementos e  $n_i$  cópias indistinguíveis do elemento do tipo  $i$ . Por exemplo, a palavra *probabilidade* tem duas cópias de cada uma das letras  $a, b, d, i$  e uma cópia de cada uma das letras  $l, p, r, o, e$ . O número de sequências ordenadas de comprimento  $n = \sum_{i=1}^r n_i$  é dado por:

$$\binom{n}{n_1} \binom{n-n_1}{n_2} \binom{n-n_1-n_2}{n_3} \cdots 1 = \frac{n!}{\prod_{i=1}^r n_i!}.$$

Esta quantidade é conhecida como *coeficiente multinomial* e denotada por:

$$\binom{n}{n_1 \ n_2 \ \dots \ n_r},$$

onde  $n = \sum_{i=1}^r n_i$ .

Para verificar esta contagem, note que das  $n$  posições na sequência de comprimento  $n$ , nós podemos escolher  $n_1$  posições para os  $n_1$  elementos indistinguíveis do tipo 1 de  $\binom{n}{n_1}$  maneiras. Das  $n - n_1$  posições restantes na sequência, podemos escolher  $n_2$  posições para os  $n_2$  elementos indistinguíveis do tipo 2 de  $\binom{n-n_1}{n_2}$  maneiras. Finalmente, após repetir este processo  $r - 1$  vezes, restam-nos  $n_r$  posições na sequência para os  $n_r$  elementos do tipo  $r$ , que só podem ser escolhidas de uma única maneira. Utilizando o método da multiplicação, o número total de sequências possíveis é produto do número de maneiras que podemos colocar os  $r$  tipos de elementos.

O coeficiente multinomial também calcula o número de partições de um conjunto  $n$  elementos em  $r$  subconjuntos com tamanhos dados  $n_1, n_2, \dots, n_r$ . Aplicando-se o mesmo argumento que utilizamos para demonstrar o Teorema Binomial, pode-se provar a seguinte generalização conhecida como *Teorema Multinomial*:

$$(x_1 + x_2 + \dots + x_r)^n = \sum_{i_1=0}^n \sum_{i_2=0}^{n-i_1} \dots \sum_{i_{r-1}=0}^{n-\sum_{j<r-1} i_j} \binom{n}{i_1 \ i_2 \ \dots \ i_r} \prod_{k=1}^r x_k^{i_k},$$

onde  $i_r = n - \sum_{j<r} i_j$ .

**Exemplo 2.2.17:** Um monitor tendo resolução de  $n = 1.280 \times 854$  pixels, com  $r = 3$  cores possíveis (verde, azul, e vermelho) para cada pixel, pode mostrar  $\binom{n}{i_1 \ i_2 \ i_3}$  imagens tendo  $i_1$  pixels verdes,  $i_2$  pixels azuis, e  $i_3$  pixels vermelhos. O número total de imagens que pode ser exibida por este monitor para qualquer composição de cores de verde, azul, e vermelho pode ser obtido utilizando o Teorema Multinomial fazendo  $x_1 = x_2 = x_3 = 1$ , dando o resultado de  $3^n$  possíveis imagens. ■

**Exemplo 2.2.18:** Determine o coeficiente de  $x^9 y^4$  no desenvolvimento de  $(x^3 + 2y^2 + \frac{5}{x^2})^5$ .

**Solução:** O termo genérico do desenvolvimento é

$$\begin{aligned} & \binom{5}{i_1 \ i_2 \ 5 - i_1 - i_2} (x^3)^{i_1} (2y^2)^{i_2} \left(\frac{5}{x^2}\right)^{5-i_1-i_2} = \\ & (2)^{i_2} (5)^{5-i_1-i_2} \binom{5}{i_1 \ i_2 \ 5 - i_1 - i_2} x^{3i_1-10+2i_1+2i_2} y^{2i_2}. \end{aligned} \quad (2.1)$$

Portanto, temos o termo  $x^9 y^4$  se  $5i_1 + 2i_2 - 10 = 9$  e  $2i_2 = 4$ , o que implica que  $i_2 = 2$  e  $i_1 = 3$ . Logo, o coeficiente de  $x^9 y^4$  é  $(2)^2 (5)^0 \binom{5}{3 \ 2 \ 0} = 40$ . ■

## 2.3 Aplicações em Grafos

Modelos matemáticos de conectividade em sistemas ou redes são baseados em grafos. Estes modelos permitem que possamos estudar questões como a conectividade de todos os elementos de uma rede, a robustez dessa conectividade a falhas em conexões entre pares de elementos, e o comprimento de caminhos entre pares de elementos. Nós estudaremos algumas características de grafos, utilizando as técnicas de contagem que aprendemos.

### 2.3.1 Grafos Não Direcionados

**Definição 2.3.1:** Um grafo não direcionado  $G = (V, E)$  é definido por um conjunto  $V$  de elementos chamados *nós* ou *vértices* e um conjunto  $E \subseteq \{\{u, v\} : u, v \in V\}$  de pares não ordenados de nós que são chamados de bordas ou arestas. Denotaremos por  $G_n$  um grafo não direcionado que contém  $n$  vértices. ■

A aresta  $\{u, v\}$  é vista como conectando os vértices  $u$  e  $v$ , e estes nós são chamados de adjacentes. O caso especial da aresta  $\{u, u\}$  é chamado de um laço. Note que o grafo é chamado de não direcional pois se  $u$  é adjacente a  $v$ , então  $v$  é adjacente a  $u$ .

Neste breve estudo de grafos, ao não ser que mencionemos ao contrário, assumiremos que não há laços.

**Exemplo 2.3.2: Número de Grafos não Direcionados com  $n$  Vértices** Qual o número  $\Gamma_n$  de grafos não direcionados com um conjunto  $V$  de  $n$  vértices? Note que o número possíveis de arestas é número de maneiras que podemos escolher pares de vértices de  $V$ . (Observe que a ordem dos vértices não é importante, pois o grafo é não direcionado.). Então temos  $\binom{n}{2}$  possíveis arestas em um grafo. Cada grafo, então corresponde a um subconjunto do conjunto de todas as arestas. Como existem  $2^r$  subconjuntos de um conjunto de  $r$  elementos, temos que existem

$$\Gamma_n = 2^{\binom{n}{2}}$$

grafos não direcionados com  $n$  vértices.

Quantos o número  $\Gamma_{n,m}$  de grafos não direcionados com um conjunto  $V$  de  $n$  vértices e um conjunto  $E$  de  $m$  arestas? Como existem  $\binom{n}{2}$  possíveis arestas, vemos que existem

$$\Gamma_{n,m} = \binom{\binom{n}{2}}{m}$$

grafos não direcionados com  $n$  vértices e  $m$  arestas. ■

### 2.3.2 Grafos Direcionados

Enquanto algumas conexões são simétricas, outras não são. Por exemplo, considere a relação social de  $u$  é pai de  $v$  ou que  $u$  é orientador de  $v$ . Evidentemente, essas relações não são simétricas. Para representar tais sistemas, precisamos de grafos direcionados.

**Definição 2.3.3:** Um grafo direcionado  $G = (V, E)$  é um conjunto  $V$  de vértices e um conjunto  $E \subseteq \{(u, v) : u, v \in V\} = V \times V$  de pares ordenados de vértices que definem arestas direcionadas que conectam  $u$  a  $v$ , mas não necessariamente o contrário. ■

**Exemplo 2.3.4:** Quantos grafos direcionados sem laços existem com um conjunto  $V$  de  $n$  vértices? Como existem  $n(n - 1)$  pares ordenados de vértices sem repetição, então o número total de possíveis arestas do grafo é  $n(n - 1)$ . Cada grafo, então corresponde a um subconjunto do conjunto de todas as arestas. Então, temos que existem

$$\Gamma_n = 2^{n(n-1)}$$

grafos direcionados com  $n$  vértices.

Quantos o número de grafos direcionados com um conjunto  $V$  de  $n$  vértices e um conjunto  $E$  de  $m$  arestas? Como existem  $n(n - 1)$  possíveis arestas, vemos que existem

$$\binom{n(n-1)}{m}$$

grafos direcionados com  $n$  vértices e  $m$  arestas. ■

# Capítulo 3

## Introdução à Probabilidade

### 3.1 Experimento Aleatório

Um *experimento* é qualquer processo de observação. Em muitos experimentos de interesse, existe um elemento de incerteza, ou chance, que não importa quanto nós sabemos sobre o passado de outras performances deste experimento, nós essencialmente não somos capazes de prever seu comportamento em futuras realizações. As razões para nossa falta de habilidade para prever são varias: nós podemos não saber de todas as causas envolvidas; nós podemos não ter dados suficientes sobre as condições iniciais do experimento; as causas podem ser tão complexas que o cálculo do seu efeito combinado não é possível; ou na verdade existe alguma aleatoriedade fundamental no experimento. Tais experimentos são conhecidos como *experimentos aleatórios*. Salvo mencionado em contrário, este livro restringe-se à classe de experimentos aleatórios cujo conjunto de possíveis resultados seja conhecido<sup>1</sup>.

Os resultados de um experimento aleatório são caracterizados pelos seguintes componentes:

1. o conjunto de resultados possíveis  $\Omega$ ;
2. a coleção de conjuntos de resultados de interesse  $\mathcal{A}$ ;
3. um valor numérico  $P$  da verossimilhança ou probabilidade de ocorrência de cada um dos conjuntos de resultados de interesse.

### 3.2 Espaço Amostral

O conjunto de possíveis resultados de um experimento aleatório é chamado de *espaço amostral*. Existem quatro pontos que são desejáveis da especificação de um espaço amostral:

---

<sup>1</sup>É importante ressaltar que freqüentemente são encontradas situações práticas onde não se consegue descrever todos os possíveis resultados de um experimento. Uma maneira de contornar este problema é assumir que um resultado possível do experimento é a não ocorrência de qualquer dos resultados descritos, contudo, em problemas práticos, tal suposição pode acarretar em dificuldades quando se tenta elicitar ou deduzir probabilidades.

SS1. listar os possíveis resultados do experimento;

SS2. fazê-lo sem duplicação;

SS3. fazê-lo em um nível de detalhamento suficiente para os interesses desejados;

SS4. especificar essa lista completamente em um sentido prático, embora usualmente não completa no que se refere a todos os resultados logicamente ou fisicamente possíveis.

Por exemplo, uma única jogada de uma moeda pode ter o espaço amostral tradicional  $\Omega = \{\text{cara}, \text{coroa}\}$ , ou podemos considerar que a moeda pode fisicamente ficar equilibrada na borda  $\Omega = \{\text{cara}, \text{coroa}, \text{borda}\}$  (SS1). Uma outra possibilidade seria levar em consideração as coordenadas  $(x, y)$  do centro da moeda quando ela para após ser jogada no ar. Como vemos muito mais se sabe sobre o resultado de uma jogada de uma moeda que os simples resultados binários tradicionais *cara* e *coroa*. Nós ignoramos esta informação adicional (SS3) usando uma hipótese não mencionada que existe uma aposta com pagamentos que dependem apenas de qual lado da moeda cai para cima e não em outras informações (SS4).

Podemos classificar espaços amostrais em dois tipos de acordo com o número de elementos que eles contêm. Espaços amostrais podem ser enumeráveis ou não enumeráveis; dependendo se  $\Omega$  for um conjunto enumerável ou não, respectivamente. Em um nível filosófico, pode-se argumentar que só existem espaços amostrais enumeráveis, visto que medidas não podem ser feitas com infinita precisão. Enquanto na prática isto é verdadeiro, métodos estatísticos e probabilísticos associados com espaços amostrais não enumeráveis são, em geral, menos complicados que aqueles para espaços amostrais enumeráveis, e proporcionam uma boa aproximação para a situação (enumerável) real.

### 3.3 Eventos e Coleção de Eventos

Um *evento* é um subconjunto do espaço amostral, ou seja, é um conjunto de resultados possíveis do experimento aleatório. Se ao realizarmos um experimento aleatório, o resultado pertence a um dado evento  $A$ , dizemos que  $A$  *ocorreu*. Estaremos interessados no estudo da ocorrência de combinações de eventos. Para tanto, utilizaremos as operações Booleanas de conjuntos (complementar, união, interseção, diferença) para expressar eventos combinados de interesse.

**Definição 3.3.1:** Os eventos  $A$  e  $B$  são *disjuntos* ou *mutuamente excludentes* ou *mutuamente exclusivos* se não puderem ocorrer juntos, ou, em linguagem de conjuntos,  $A \cap B = \emptyset$ .

**Exemplo 3.3.2:** Sejam  $A$ ,  $B$ , e  $C$  eventos em um mesmo espaço amostral  $\Omega$ . Expresse os seguintes eventos em função de  $A$ ,  $B$ , e  $C$  e operações Booleanas de conjuntos.

(a) Pelo menos um deles ocorre:

$$A \cup B \cup C.$$

(b) Exatamente um deles ocorre:

$$(A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C).$$



(c) Apenas  $A$  ocorre:

$$(A \cap B^c \cap C^c).$$

(d) Pelo menos dois ocorrem:

$$(A \cap B \cap C^c) \cup (A \cap B^c \cap C) \cup (A^c \cap B \cap C) \cup (A \cap B \cap C).$$

(e) No máximo dois deles ocorrem:

$$(A \cap B \cap C)^c.$$

(f) Nenhum deles ocorre:

$$(A^c \cap B^c \cap C^c).$$

(g) Ambos  $A$  e  $B$  ocorrem, mas  $C$  não ocorre:

$$(A \cap B \cap C^c).$$

Embora possa-se pensar que, dado um espaço amostral, necessariamente é de interesse analisar todos os seus subconjuntos (e isto eventualmente é verdadeiro), temos três razões para esperar que estejamos apenas interessados em alguns subconjuntos do espaço amostral. Primeiro, o espaço amostral pode conter um grau de detalhamento superior ao que estamos interessados no momento. Por exemplo, ele pode representar uma única jogada de um dado com 6 elementos, mas nós apenas estamos interessados em saber se o resultado é par ou ímpar. Segundo, nós vamos querer associar cada evento  $A$  com uma probabilidade numérica  $P(A)$ . Como essas probabilidades estão baseadas em algum conhecimento sobre a tendência de ocorrer do evento ou no grau de nossa crença que determinado evento ocorrerá, nosso conhecimento sobre  $P$  pode não estender para todos os subconjuntos de  $\Omega$ . A terceira (e técnica) razão para limitar a coleção de eventos de interesse é que condições impostas em  $P$  pelos axiomas de Kolmogorov, que estudaremos adiante, podem não permitir que  $P$  seja definida em todos os subconjuntos de  $\Omega$ , em particular isto pode ocorrer quando  $\Omega$  for não enumerável, mas não iremos demonstrar este fato que está fora do escopo deste curso.

Estaremos interessados em uma coleção especial  $\mathcal{A}$  de subconjuntos do espaço amostral  $\Omega$  (note que  $\mathcal{A}$  é um conjunto cujos elementos também são conjuntos!) que são eventos de interesse no que se refere ao experimento aleatório  $\mathcal{E}$  e os quais temos conhecimento sobre a sua verossimilhança de ocorrência.  $\mathcal{A}$  é chamado de uma *álgebra de eventos*.

**Definição 3.3.3:** Uma álgebra de eventos  $\mathcal{A}$  é uma coleção de subconjuntos do espaço amostral  $\Omega$  que satisfaz:

1. não é vazia;
2. fechada com respeito a complementos (se  $A \in \mathcal{A}$ , então  $A^c \in \mathcal{A}$ );
3. fechada com respeito a uniões finitas (se  $A, B \in \mathcal{A}$ , então  $A \cup B \in \mathcal{A}$ ).

Pelas Leis de De Morgan, vemos que  $\mathcal{A}$  é fechada com respeito a intersecções finitas também.

**Exemplo 3.3.4:**

1. A menor álgebra de eventos é  $\mathcal{A} = \{\emptyset, \Omega\}$ ;
2. A maior álgebra de eventos é o conjunto das partes de  $\Omega$ ;
3. Um exemplo intermediário, temos:

$$\Omega = \{1, 2, 3\}, \mathcal{A} = \{\Omega, \emptyset, \{1\}, \{2, 3\}\}.$$

4. A álgebra de eventos finitos e co-finitos. Seja  $\Omega = \mathbb{R}$  e

$$\mathcal{A} = \{A \subseteq \mathbb{R} : A \text{ é finito}\} \cup \{A \subseteq \mathbb{R} : A^c \text{ é finito}\},$$

ou seja,  $\mathcal{A}$  consiste dos subconjuntos de  $\mathbb{R}$  que ou são finitos ou têm complementos finitos.  $\mathcal{A}$  é uma álgebra de eventos.

**Lema 3.3.5:** *Se  $\mathcal{A}$  é uma álgebra, então  $\Omega \in \mathcal{A}$*

**Prova:** Como  $\mathcal{A}$  é não vazio, seja  $A$  um elemento qualquer seu. Pela segunda propriedade de álgebras, temos que  $A^c \in \mathcal{A}$ , e pela terceira propriedade temos que  $\Omega = A \cup A^c \in \mathcal{A}$ . ■

**Teorema 3.3.6:** *Sejam  $\mathcal{A}_1$  e  $\mathcal{A}_2$  álgebras de subconjuntos de  $\Omega$  e seja  $\mathcal{A} = \mathcal{A}_1 \cap \mathcal{A}_2$  a coleção de subconjuntos comuns as duas álgebras. Então,  $\mathcal{A}$  é uma álgebra.*

**Prova:** Como  $\mathcal{A}_1$  e  $\mathcal{A}_2$  são álgebras, ambos contêm  $\Omega$ . Então,  $\Omega \in \mathcal{A}$ . Se  $A \in \mathcal{A}$ , então  $A$  está em ambos  $\mathcal{A}_1$  e  $\mathcal{A}_2$ . Logo,  $A^c$  está em ambos  $\mathcal{A}_1$  e  $\mathcal{A}_2$ , e portanto na sua intersecção  $\mathcal{A}$ . Se  $A, B \in \mathcal{A}$ , então eles estão em ambos  $\mathcal{A}_1$  e  $\mathcal{A}_2$ . Consequentemente,  $A \cup B$  está em ambos  $\mathcal{A}_1$  e  $\mathcal{A}_2$  e, portanto, em  $\mathcal{A}$ . Como  $\mathcal{A}$  satisfaz as três condições da definição de álgebra de eventos,  $\mathcal{A}$  é uma álgebra de eventos. ■

É fácil ver que a prova do Teorema 3.3.6 pode ser estendida para o caso de uma intersecção de um número arbitrário de álgebras. O seguinte corolário usa este fato para provar que sempre existe uma menor álgebra contendo uma família qualquer de eventos.

**Corolário 3.3.7:** *Existe uma menor (no sentido de inclusão) álgebra contendo qualquer família dada de subconjuntos de  $\Omega$ .*

**Prova:** Seja  $\mathcal{C}$  uma coleção qualquer de subconjuntos de  $\Omega$ , defina  $\mathcal{A}(\mathcal{C})$  como sendo o conjunto que é igual a intersecção de todas as álgebras de eventos que contêm  $\mathcal{C}$ , isto é:

$$\mathcal{A}(\mathcal{C}) = \bigcap_{\mathcal{A} \supseteq \mathcal{C}: \mathcal{A} \text{ é uma álgebra de eventos}} \mathcal{A}.$$

Pelo Teorema 3.3.6,  $\mathcal{A}(\mathcal{C})$  é uma álgebra de eventos, e consequentemente é a menor álgebra de eventos contendo  $\mathcal{C}$ .  $\mathcal{A}(\mathcal{C})$  é conhecida como a *álgebra de eventos gerada por  $\mathcal{C}$* . ■

**Observação 3.3.8:** Uma maneira de construir uma álgebra de eventos, é primeiro particionar  $\Omega$  em um número finito subconjuntos e depois considerar álgebra que consiste dos eventos que são uniões finitas dos subconjuntos da partição. ■

**Exemplo 3.3.9:** Por exemplo,  $\Omega = \{a, b, c, d\}$ . Considere a partição,  $\{\{a, c\}, \{b, d\}\}$ , então considere a coleção de eventos que consiste de uniões finitas dos eventos desta partição:  $\mathcal{A} = \{\emptyset, \Omega, \{a, c\}, \{b, d\}\}$ . É fácil ver que  $\mathcal{A}$  é uma álgebra de eventos.

Dada uma coleção finita eventos  $\mathcal{C} = \{A_1, A_2, \dots, A_n\}$ , define-se um átomo de  $\mathcal{C}$  como sendo qualquer evento  $B$  da seguinte forma:  $B = B_1 \cap B_2 \cap \dots \cap B_n$ , onde  $B_i = A_i$  ou  $B_i = A_i^c$  para  $i = 1, 2, \dots, n$ . Note que existem no máximo  $2^{|\mathcal{C}|}$  átomos diferentes e que eles formam uma partição de  $\mathcal{C}$  (verifique!). Quando  $\mathcal{C}$  for uma coleção finita de eventos, um evento pertencerá a  $\mathcal{A}(\mathcal{C})$ , se e somente se, for igual a uma união finita de átomos de  $\mathcal{C}$ . Note que  $\mathcal{A}(\mathcal{C})$  terá no máximo  $2^{2^{|\mathcal{C}|}}$  elementos (verifique!).

**Exemplo 3.3.10:** Se  $\Omega = \{a, b, c, d, e, f\}$ , encontre a álgebra gerada por  $\mathcal{C} = \{\{a, b, d\}, \{b, d, f\}\}$ . Os átomos de  $\mathcal{C}$  são  $\{\{a\}, \{f\}, \{c, e\}, \{b, d\}\}$ . Logo,

$$\begin{aligned} \mathcal{A}(\mathcal{C}) = & \{\emptyset, \Omega, \{a\}, \{f\}, \{c, e\}, \{b, d\}, \{a, f\}, \{a, c, e\}, \\ & \{a, b, d\}, \{c, e, f\}, \{b, d, f\}, \{b, c, d, e\}, \{a, f, c, e\}, \\ & \{a, f, b, d\}, \{a, b, c, d, e\}, \{b, c, e, d, f\}\}. \end{aligned}$$

### 3.3.1 Indução Matemática

Abriremos um parênteses na nossa discussão sobre experimentos aleatórios para discutirmos um método de prova matemática que será bastante utilizado ao longo do curso: a *Indução Matemática*.

Indução Matemática é um método de prova matemática usado para demonstrar a verdade de um número infinito de proposições. A forma mais simples e mais comum de indução matemática prova que um enunciado vale para todos os números naturais  $n$  e consiste de dois passos:

- A base: mostrar que o enunciado vale para  $n = 1$ .
- O passo indutivo: mostrar que, se o enunciado vale para  $n = k$ , então o mesmo enunciado vale para  $n = k + 1$ .

Esse método funciona provando que o enunciado é verdadeiro para um valor inicial, e então provando que o processo usado para ir de um valor para o próximo é válido. Se ambas as coisas são provadas, então qualquer valor pode ser obtido através da repetição desse processo. Para entender por que os dois passos são suficientes, é útil pensar no efeito dominó: se você tem uma longa fila de dominós em pé e você puder assegurar que:

- O primeiro dominó cairá.

- Sempre que um dominó cair, seu próximo vizinho também cairá. então você pode concluir que todos os dominós cairão.

**Exemplo 3.3.11:** Suponha que desejemos provar o seguinte enunciado: para todos os números naturais  $n$ ,

$$\sum_{i=1}^n i = \frac{n(n+1)}{2}.$$

Esta é uma fórmula simples para a soma dos números naturais de 1 a  $n$ . A prova de que o enunciado é verdadeiro para todos os números naturais  $n$  é dada a seguir.

**Prova:** Verificar se o enunciado é verdadeiro para  $n = 1$  (base). Claramente, do lado esquerdo da equação fica 1 e do lado direito  $1(1+1)/2$ , resolvendo dá  $1 = 1$ . Então o enunciado é verdadeiro para  $n = 1$ . Podemos definir este enunciado como  $P(n)$  e portanto temos que  $P(1)$  é verdadeiro.

Agora precisamos mostrar que se o enunciado vale quando  $n = k$ , então ele também vale quando  $n = k + 1$  (passo indutivo). Isto pode ser feito da seguinte maneira:

Assuma que o enunciado é válido para  $n = k$ , ou seja:

$$\sum_{i=1}^k i = \frac{k(k+1)}{2}.$$

Adicionando  $k + 1$  a ambos os lados:

$$\sum_{i=1}^{k+1} i = \frac{k(k+1)}{2} + k + 1 = \frac{k(k+1)}{2} + \frac{2(k+1)}{2} = \frac{(k+2)(k+1)}{2}.$$

Este último é o enunciado para  $n = k + 1$ . Note que ele não foi provado como verdadeiro: nós assumimos que  $P(k)$  é verdadeiro, e desta suposição concluímos que  $P(k+1)$  é verdadeiro. Simbolicamente, mostramos que:

$$P(k) \Rightarrow P(k+1)$$

Por indução, no entanto, podemos concluir que o enunciado  $P(n)$  vale para todos os números naturais  $n$ :

1.  $P(1)$  é verdadeiro, logo  $P(2)$  é verdadeiro (usando o passo indutivo)
2. Como  $P(2)$  é verdadeiro, então  $P(3)$  também é
3. Então usando-se o passo indutivo  $P(N)$  será verdadeiro e o  $P(N+1)$  também

Agora podemos provar que toda álgebra é fechada com respeito a um número finito de uniões.

**Teorema 3.3.12:** Se  $\mathcal{A}$  é uma álgebra de eventos, então

$$A_i \in \mathcal{A}, i = 1, 2, \dots, n \Rightarrow \cup_{i=1}^n A_i \in \mathcal{A}$$

**Prova:** Para  $n = 1$ , o resultado é óbvio. Para  $n = 2$ , o resultado segue diretamente da terceira propriedade na definição de álgebra de eventos. Vamos agora provar o passo indutivo, suponha que

$$A_i \in \mathcal{A}, i = 1, 2, \dots, k \Rightarrow \cup_{i=1}^k A_i \in \mathcal{A}.$$

Vamos agora provar que o caso  $n = k + 1$  é verdadeiro. Suponha que  $A_i, i = 1, 2, \dots, k + 1 \in \mathcal{A}$ , então como

$$\cup_{i=1}^{k+1} A_i = (\cup_{i=1}^k A_i) \cup A_{k+1},$$

temos que utilizando o caso  $n = k$ ,  $\cup_{i=1}^k A_i \in \mathcal{A}$ . Como  $\cup_{i=1}^k A_i \in \mathcal{A}$  e  $A_{k+1} \in \mathcal{A}$ , temos que utilizando o caso  $n = 2$ ,  $(\cup_{i=1}^k A_i) \cup A_{k+1} \in \mathcal{A}$ . ■

### 3.4 Frequências Relativas

Resta-nos discutir o terceiro elemento para modelagem do raciocínio probabilístico, a associação de uma medida numérica a eventos que representam a probabilidade com que eles ocorrem. As propriedades desta associação são motivadas em grande parte pelas propriedades de frequência relativas. Considere uma coleção de experimentos aleatórios  $\mathcal{E}_i$  que possuem a mesma  $\sigma$ -álgebra de eventos  $\mathcal{A}$  e tem resultados individuais não necessariamente numéricos  $\{\omega_i\}$ . Fixando uma dada seqüência de resultados  $\{\omega_i\}$ , se estamos interessados na ocorrência de um dado evento  $A$ , a frequência relativa de  $A$  nada mas é que uma média aritmética da função indicadora de  $A$  calculada em cada um dos termos da seqüência  $\{\omega_i\}$ , ou seja,

**Definição 3.4.1:** A frequência relativa de um evento  $A$ , determinada pelos resultados  $\{\omega_1, \dots, \omega_n\}$  de  $n$  experimentos aleatórios, é

$$r_n(A) = \frac{1}{n} \sum_{i=1}^n I_A(\omega_i) = \frac{N_n(A)}{n}.$$

Propriedades chaves da frequência relativa são:

FR0.  $r_n : \mathcal{A} \rightarrow \mathbb{R}$ .

FR1.  $r_n(A) \geq 0$ .

FR2.  $r_n(\Omega) = 1$ .

FR3. Se  $A$  e  $B$  são disjuntos, então  $r_n(A \cup B) = r_n(A) + r_n(B)$ .

FR4. Se  $A_1, A_2, \dots, A_n, \dots$  é uma seqüência de eventos disjuntos dois a dois, então  $r_n(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} r_n(A_i)$ .

Nós prosseguiremos como se existisse alguma base empírica ou metafísica que garanta que  $r_n(A) \rightarrow P(A)$ , embora que o sentido de convergência quando  $n$  cresce só será explicado pela Lei dos Grandes Números, que não será discutida neste curso. Esta tendência da frequência relativa de estabilizar em um certo valor é conhecida como *regularidade estatística*. Deste modo,  $P$  herdará propriedades da frequência relativa  $r_n$ .

### 3.5 Axiomas de Kolmogorov

Primeiro por razões técnicas, fora do escopo deste curso, temos que o domínio da medida formal de probabilidade é uma álgebra de eventos que também é fechada com relação a um número enumerável de uniões.

**Definição 3.5.1:** Uma  $\sigma$ -álgebra  $\mathcal{A}$  é uma álgebra de eventos que também é fechada com relação a uma união enumerável de eventos,

$$(\forall i \in \mathbb{Z}) A_i \in \mathcal{A} \Rightarrow \cup_{i \in \mathbb{Z}} A_i \in \mathcal{A}.$$

Pelas Leis de De Morgan, tem-se que  $\mathcal{A}$  também é fechada com respeito a intersecções enumeráveis.

**Exemplo 3.5.2:** A coleção de conjuntos de números reais finitos e co-finitos é uma álgebra que não é uma  $\sigma$ -álgebra.

**Exemplo 3.5.3:** A  $\sigma$ -álgebra de Borel  $\mathcal{B}$  de subconjuntos reais é, por definição, a menor  $\sigma$ -álgebra contendo todos os intervalos e é a  $\sigma$ -álgebra usual quando lidamos com quantidades reais ou vetoriais. Em particular, temos que uniões enumeráveis de intervalos (por exemplo, o conjunto dos números racionais), seus complementos (por exemplo, o conjunto dos números irracionais), e muito mais está em  $\mathcal{B}$ .

Os axiomas que descreveremos a seguir não descrevem um único modelo probabilístico, eles apenas determinam uma família de modelos probabilísticos, com os quais poderemos utilizar métodos matemáticos para descobrir propriedades que serão verdadeiras em qualquer modelo probabilístico. A escolha de um modelo específico satisfazendo os axiomas é feito pelo analista/estatístico familiar com o fenômeno aleatório sendo modelado.

Motivados pelas propriedades de frequência relativa, impõe-se os primeiros quatro axiomas de Kolmogorov:

**K0. Inicial.** O experimento aleatório é descrito pelo espaço de probabilidade  $(\Omega, \mathcal{A}, P)$  que consiste do espaço amostral  $\Omega$ , de uma  $\sigma$ -álgebra  $\mathcal{A}$ , e de uma função de valores reais  $P : \mathcal{A} \rightarrow \mathbb{R}$ .

**K1. Não-negatividade.**  $\forall A \in \mathcal{A}, P(A) \geq 0$ .

**K2. Normalização Unitária.**  $P(\Omega) = 1$ .

**K3. Aditividade Finita.** Se  $A, B$  são disjuntos, então  $P(A \cup B) = P(A) + P(B)$ .

É fácil provar (tente!) utilizando indução matemática que K3 é válida para qualquer coleção finita de eventos disjuntos par a par, ou seja, se  $A_i, i = 1, 2, \dots, n$ , são eventos disjuntos par a par, então  $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$ .

Um quinto axioma, embora não tenha significado em espaços amostrais finitos, foi proposto por Kolmogorov para garantir um certo grau de continuidade da medida de probabilidade.

**K4. Continuidade Monotônica.** Se para todo  $i > 0$ ,  $A_{i+1} \subseteq A_i$  e  $\bigcap_i A_i = \emptyset$ , então

$$\lim_{i \rightarrow \infty} P(A_i) = 0.^2$$

Um forma equivalente de K4 é a seguinte:

**K4'.  $\sigma$ -aditividade.** Se  $\{A_i\}$  é uma coleção enumerável de eventos disjuntos dois a dois, então

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

**Teorema 3.5.4:** *Se  $P$  satisfaz K0—K3, então  $P$  satisfaz K4' se, e somente se, ela satisfaz K4.*

**Prova:** Primeiro, vamos provar que K0—K4 implicam o axioma da  $\sigma$ -aditividade K4'. Seja  $\{A_i\}$  qualquer sequência enumerável de eventos disjuntos par a par, e defina para todo  $n$

$$B_n = \bigcup_{i>n} A_i,$$

$$\bigcup_{i=1}^{\infty} A_i = B_n \cup (\bigcup_{i=1}^n A_i).$$

Claramente, para todo  $i \leq n$ , temos que  $A_i$  e  $B_n$  são disjuntos. Por K3, temos

$$P(\bigcup_{i=1}^{\infty} A_i) = P(B_n) + \sum_{i=1}^n P(A_i).$$

Por definição de série numérica,

$$\lim_n \sum_{i=1}^n P(A_i) = \sum_{i=1}^{\infty} P(A_i).$$

K4' segue se conseguirmos mostrar que  $\lim_n P(B_n) = 0$ . Note que  $B_{n+1} \subseteq B_n$ , e que  $\bigcap_{n=1}^{\infty} B_n = \emptyset$ . Então por K4, temos que o limite acima é zero e K4' é verdadeiro.

Agora, vamos provar que K0—K3, K4' implicam o axioma da continuidade monotônica K4. Seja  $\{B_n\}$  qualquer coleção enumerável de eventos satisfazendo as hipóteses do axioma K4:  $B_{n+1} \subseteq B_n$  e  $\bigcap_{n=1}^{\infty} B_n = \emptyset$ . Defina,  $A_n = B_n - B_{n+1}$  e observe que  $\{A_n\}$  é uma coleção enumerável de eventos disjuntos par a par. Note que

$$B_n = \bigcup_{j \geq n} A_j.$$

Então, por K4' temos que

$$P(B_n) = P(\bigcup_{j \geq n} A_j) = \sum_{j \geq n} P(A_j).$$

---

<sup>2</sup>K4 é uma idealização que não é aceita por alguns tratamentos subjetivistas de probabilidade.

$\Omega$	conjunto universo	espaço amostral, evento certo
$\omega$	elemento	resultado do experimento
$A$	conjunto $A$	evento $A$
$\emptyset$	conjunto vazio	evento impossível
$A^c$ ou $\bar{A}$	complemento de $A$	não ocorreu o evento $A$
$A \cap B$	$A$ intersecção $B$	os eventos $A$ e $B$ ocorreram
$A \cup B$	$A$ união $B$	os eventos $A$ ou $B$ ocorreram
$\bigcap_n A_n$	intersecção dos conjuntos $A_n$	todos os eventos $A_n$ ocorreram
$\bigcup_n A_n$	união dos conjuntos $A_n$	ao menos um dos eventos $A_n$ ocorreu

Tabela 3.1: Interpretações interessantes

Como por K4',

$$\sum_{j=1}^{\infty} P(A_j) = P(\bigcup_{j=1}^{\infty} A_j) \leq 1,$$

temos que

$$\lim_n P(B_n) = \lim_n \sum_{j \geq n} P(A_j) = 0,$$

logo K4 é verdadeiro. ■

**Definição 3.5.5:** Uma função que satisfaz K0—K4 é chamada de uma medida de probabilidade.

A terna  $(\Omega, \mathcal{A}, P)$  é chamada de **espaço de probabilidade**. Intuitivamente quando se modela uma problema através de probabilidade, basicamente, o que se faz é especificar cada uma das componentes da terna acima.

Eventos são os elementos de  $\mathcal{A}$ , aos quais se pode atribuir probabilidade. Probabilidade é uma função cujo argumento é um conjunto. Portanto, não somente conjuntos, como também as operações sobre eles, têm uma importância fundamental em teoria da probabilidade. Entretanto, é preciso que a linguagem de conjuntos seja traduzida para a linguagem de probabilidade. A Tabela 3.1, a seguir, exhibe algumas dessas traduções. A idéia subjacente é que um experimento aleatório foi realizado e aconteceu algum evento.

### 3.5.1 Exemplos de Medidas de Probabilidade

**Exemplo 3.5.6:** Se  $\Omega$  for um conjunto finito, então temos que a probabilidade clássica que assume que todos os resultados são igualmente prováveis, é um exemplo de uma medida de probabilidade. Neste caso, temos que

$$P(A) = \frac{||A||}{||\Omega||}$$

definido para qualquer subconjunto  $A$  de  $\Omega$ . O fato que  $0 \leq ||A|| \leq ||\Omega||$  e que

$$||A \cup B|| = ||A|| + ||B|| - ||A \cap B||,$$



permitted que verifiquemos que  $P$  satisfaz os axiomas de Kolmogorov.

**Exemplo 3.5.7:** Seja  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$  um conjunto finito, e seja  $P(\{\omega_i\}) = p_i$ , onde  $p_i \geq 0, i \geq 1$  e  $\sum_{i=1}^n p_i = 1$ , e  $P(A) = \sum_{\omega_i \in A} P(\{\omega_i\})$ . Neste caso, também é fácil verificar que  $P$  é uma medida de probabilidade verificando os axiomas.

### 3.5.2 Propriedades de uma Medida de Probabilidade

**Teorema 3.5.8:** *Se  $P$  é uma medida de probabilidade, então*

1.  $P(A^c) = 1 - P(A)$ .
2.  $P(\emptyset) = 0$ .
3.  $P(A) \leq 1$ .

**Prova:** Parte 1, segue do fato que  $\Omega = A \cup A^c$ , K2, e K3, pois

$$1 = P(\Omega) = P(A) + P(A^c).$$

Parte 2, segue da Parte 1, do fato que  $\Omega^c = \emptyset$ , e K2, K3, pois

$$P(\emptyset) = 1 - P(\Omega) = 0.$$

Parte 3, segue do fato que  $1 = P(\Omega) = P(A) + P(A^c) \geq P(A)$ , já que  $P(A^c) \geq 0$  por K1. ■

**Teorema 3.5.9: Monotonicidade.** *Se  $A \subseteq B$ , então  $P(A) \leq P(B)$ .*

**Prova:** Note que  $B = A \cup (B - A)$ , onde  $A$  e  $B - A$  são disjuntos. Então K3 implica que  $P(B) = P(A) + P(B - A)$ . O resultado segue do fato que  $P(B - A) \geq 0$ . ■

**Corolário 3.5.10:**  $P(A \cup B) \geq \max(P(A), P(B)) \geq \min(P(A), P(B)) \geq P(A \cap B)$ .

**Teorema 3.5.11:** *Uma expressão exata para a probabilidade de uma união não-disjunta é dada por*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

**Prova:** Como  $A \cup B = A \cup (B - A)$ , e  $A$  e  $B - A$  são disjuntos, K3 implica que  $P(A \cup B) = P(A) + P(B - A)$ . E como  $B = (A \cap B) \cup (B - A)$ ,  $A \cap B$  e  $B - A$  são disjuntos, K3 implica que  $P(B) = P(A \cap B) + P(B - A)$ . Logo,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

■

**Teorema 3.5.12: Probabilidade de Partições.** Se  $\{A_i\}$  é uma partição enumerável de  $\Omega$  feita de conjuntos em  $\mathcal{A}$ , então para todo  $B \in \mathcal{A}$

$$P(B) = \sum_i P(B \cap A_i).$$

**Prova:** Como  $\{A_i\}$  é uma partição, segue que

$$B = B \cap \Omega = B \cap (\cup_i A_i) = \cup_i (B \cap A_i).$$

O resultado segue então por K4'. ■

**Teorema 3.5.13: Desigualdade de Boole.** Para  $n$  eventos arbitrários  $\{A_1, \dots, A_n\}$ , a desigualdade de Boole é

$$P(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i).$$

**Prova:** Provaremos por indução matemática em  $n$ . A desigualdade é trivialmente verdadeira para  $n = 1$  e verdadeira para  $n = 2$ , pois é uma consequência imediata do Teorema 3.5.11. Assuma que a desigualdade é válida para  $n = k$  e vamos provar que ela é válida para  $n = k+1$ . Para ver isto, escrevemos  $\cup_{i=1}^{k+1} A_i = A_{k+1} \cup \cup_{i=1}^k A_i$ .

Pela desigualdade para  $n = 2$ ,

$$P(\cup_{i=1}^{k+1} A_i) \leq P(A_{k+1}) + P(\cup_{i=1}^k A_i).$$

Pela hipótese do passo indutivo, para  $n = k$ ,

$$P(\cup_{i=1}^k A_i) \leq \sum_{i=1}^k P(A_i),$$

portanto, a desigualdade de Boole é verdadeira. ■

**Corolário 3.5.14:** Para  $n$  eventos arbitrários  $\{A_1, \dots, A_n\}$ ,

$$P(\cap A_i) \geq \sum_{i=1}^n P(A_i) - (n-1).$$

**Prova:** Utilizando a Lei de De Morgan e a desigualdade de Boole para os eventos  $\{A_1^c, \dots, A_n^c\}$ , temos

$$P(\cup_{i=1}^n A_i^c) = 1 - P(\cap A_i) \leq \sum_{i=1}^n P(A_i^c) = \sum_{i=1}^n (1 - P(A_i)).$$

Logo,

$$P(\cap A_i) \geq \sum_{i=1}^n P(A_i) - (n-1).$$

■

O próximo teorema permite que possamos calcular de maneira exata a probabilidade  $P(\cup_{i=1}^n A_i)$  para  $n$  eventos arbitrários.

**Teorema 3.5.15: Princípio da Inclusão-Exclusão.** *Seja  $I$  um conjunto genérico de índices que é um subconjunto não-vazio qualquer de  $\{1, 2, \dots, n\}$ . Para eventos arbitrários  $\{A_1, \dots, A_n\}$ ,*

$$P(\cup_{i=1}^n A_i) = \sum_{\emptyset \neq I \subseteq \{1, \dots, n\}} (-1)^{|I|+1} P(\cap_{i \in I} A_i),$$

onde o somatório é sobre todos os  $2^n - 1$  conjuntos de índices excluindo apenas o conjunto vazio.

No caso particular de  $n = 3$ , o princípio de inclusão-exclusão afirma que

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3).$$

**Prova:** A prova é por indução matemática em  $n$ . O resultado é trivialmente verdadeiro para  $n = 1$  e já foi provado para  $n = 2$  no Teorema 3.5.11. Assuma que o resultado vale para  $n = k$  e vamos provar que ele é verdadeiro para  $n = k + 1$ . Como na prova da desigualdade de Boole,  $\cup_{i=1}^{k+1} A_i = A_{k+1} \cup \cup_{i=1}^k A_i$ . Usando o resultado para  $n = 2$ , temos

$$P(\cup_{i=1}^{k+1} A_i) = P(A_{k+1}) + P(\cup_{i=1}^k A_i) - P(A_{k+1} \cap \cup_{i=1}^k A_i).$$

Reescrevendo o último termo como  $P(\cup_{i=1}^k (A_{k+1} \cap A_i))$ , nos dá uma expressão que contém uma união de exatamente  $k$  conjuntos. Então, usando a hipótese do passo indutivo para os dois últimos termos

$$P(\cup_{i=1}^{k+1} A_i) = P(A_{k+1}) + \sum_{\emptyset \neq I \subseteq \{1, \dots, k\}} (-1)^{|I|+1} P(\cap_{i \in I} A_i) - \sum_{\emptyset \neq I \subseteq \{1, \dots, k\}} (-1)^{|I|+1} P(\cap_{i \in I} (A_{k+1} \cap A_i)).$$

O resultado segue ao rearranjarmos os termos destes somatórios. ■

**Exemplo 3.5.16:** Professor Leônidas está tentando calcular a probabilidade  $p = P(A)$  do evento  $A$ , e determinou que ela é uma raiz do seguinte polinômio de grau cinco:

$$(p - 3)(p - 3\sqrt{-1})(p + 3\sqrt{-1})(p + 0.3)(p - 0.3) = 0.$$

Baseado nesta fato, qual é o valor de  $p$ ?

**Exemplo 3.5.17:** Se  $\Omega = \{a, b, c\}$ , e a álgebra  $\mathcal{A}$  é o conjunto das partes de  $\Omega$ , e a medida de probabilidade  $P$  é parcialmente definida por

$$P(\{a, b\}) = 0.5, P(\{b, c\}) = 0.8, P(\{a, c\}) = 0.7,$$

então complete a especificação de  $P$  para todos os eventos em  $\mathcal{A}$ .

**Exemplo 3.5.18:** Se  $\{A_i\}$  for uma partição enumerável de  $\Omega$  e  $P(A_i) = ab^i$ ,  $i \geq 1$ , então quais as condições que  $a$  e  $b$  devem satisfazer para que  $P$  seja uma medida de probabilidade?

**Exemplo 3.5.19:** Em um grupo de  $r$  pessoas qual a probabilidade de haver pelo menos duas pessoas que façam aniversário no mesmo dia, assumindo que a distribuição de aniversários é uniforme ao longo do ano e desprezando a existência de anos bissextos?

**Solução:** Para determinar esta probabilidade, vamos utilizar a probabilidade clássica. O número de resultados possíveis para os aniversários de  $r$  pessoas é  $365^r$ . O número de casos possíveis onde todas as pessoas fazem aniversário em dias diferentes é dado por  $365 \times 364 \times \dots \times (365 - (r - 1))$ . Portanto, o número de casos possíveis onde pelo menos duas pessoas fazem aniversário no mesmo dia é a diferença entre o número total de aniversários possíveis e o número de casos onde as pessoas têm aniversários em datas diferentes, ou seja, é igual a

$$365^r - 365 \times 364 \times \dots \times (365 - (r - 1)).$$

Logo, a probabilidade deste evento é:

$$1 - \frac{365 \times 364 \times \dots \times (365 - (r - 1))}{365^r}.$$

Para  $r = 23$ , temos que essa probabilidade é aproximadamente igual a 0,51. E para  $r = 50$ , essa probabilidade é igual a 0,97.

**Exemplo 3.5.20:** Em uma loteria de  $N$  números há um só prêmio. Salvador compra  $n$  ( $1 < n < N$ ) bilhetes para uma só extração e Sílvio compra  $n$  bilhetes, um para cada uma de  $n$  extrações. Qual dos dois jogadores têm mais chances de ganhar algum prêmio?

**Solução:** A probabilidade de Salvador ganhar algum prêmio é  $\frac{n}{N}$ . O número total de  $n$  extrações possíveis é  $N^n$ . O número de casos onde Sílvio não ganha nenhum prêmio é  $(N - 1)^n$ , logo o número de casos onde Sílvio ganha algum prêmio é igual a  $N^n - (N - 1)^n$ . Logo, a probabilidade de Sílvio ganhar algum prêmio é  $1 - \frac{(N-1)^n}{N^n}$ .

Vamos provar por indução que Salvador tem mais chance de ganhar, ou seja,  $\frac{n}{N} > 1 - \frac{(N-1)^n}{N^n}$ , que equivale a

$$\frac{(N - 1)^n}{N^n} > 1 - \frac{n}{N}.$$

Para  $n = 2$ , temos:

$$\frac{(N - 1)^2}{N^2} = 1 - \frac{2}{N} + \frac{1}{N^2} > 1 - \frac{2}{N}.$$

Suponha que para  $n = k$ , temos que

$$\frac{(N - 1)^k}{N^k} > 1 - \frac{k}{N}.$$

Multiplicando esta expressão por  $\frac{N-1}{N}$ , obtemos:

$$\frac{(N - 1)^{k+1}}{N^{k+1}} > \left(\frac{N - 1}{N}\right)\left(1 - \frac{k}{N}\right) = 1 - \frac{1}{N} - \frac{k}{N} + \frac{k}{N^2} > 1 - \frac{k + 1}{N}.$$

**Exemplo 3.5.21:** Doze pessoas são divididas em três grupos de 4. Qual é a probabilidade de duas determinadas dessas pessoas ficarem no mesmo grupo?

**Solução:** O número total de divisões de doze pessoas em 3 grupos de 4 é igual a  $\binom{12}{4}\binom{8}{4}\binom{4}{4}$ . Vamos agora contar o número de casos favoráveis ao nosso evento. Existem 3 opções de escolhermos em qual grupo as duas pessoas determinadas podem ficar. Das 10 pessoas restantes, temos que escolher mais duas para estarem neste grupo, o que podemos fazer de  $\binom{10}{2}$  maneiras diferentes. E temos  $\binom{8}{4}\binom{4}{4}$  maneiras diferentes de dividir as outras 8 pessoas nos dois grupos restantes. Portanto, a probabilidade de duas determinadas pessoas ficarem no mesmo grupo é:

$$\frac{3\binom{10}{2}\binom{8}{4}\binom{4}{4}}{\binom{12}{4}\binom{8}{4}\binom{4}{4}} = \frac{3}{11}.$$

**Exemplo 3.5.22:** Suponha que temos em uma sala  $n$  mães cada uma com um filho. Suponha formemos duplas aleatoriamente, onde cada dupla contém uma mãe e um filho, qual a probabilidade de que pelo menos uma mãe forme uma dupla com seu próprio filho?

**Solução:** Seja  $A_i$  o evento que a  $i$ -ésima mãe forma dupla com seu filho. Queremos determinar

$$P(\cup_{i=1}^n A_i).$$

Vamos calcular esta probabilidade utilizando a fórmula da inclusão exclusão. Note que:

$$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n} \text{ para todo } i \in \{1, 2, \dots, n\}$$

$$P(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)} \text{ para } i \neq j$$

e em geral, para um grupo  $I \in \{1, 2, \dots, n\}$  de mães temos que

$$P(\cap_{i \in I} A_i) = \frac{(n - ||I||)!}{n!}.$$

Como existem  $\binom{n}{||I||}$  grupos de mães com cardinalidade  $||I||$ , temos que

$$P(\cup_{i=1}^n A_i) = \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} \frac{(n-i)!}{n!}$$

$$= \sum_{i=1}^n (-1)^{i+1} \frac{1}{i!}$$

Note que quando  $n \rightarrow \infty$ , temos que esta probabilidade tende a  $1 - \frac{1}{e}$ .

**Exemplo 3.5.23:** Demonstre que se  $P(A_i) = 1$  para  $i = 1, 2, \dots$ , então  $P(\cap_{i=1}^{\infty} A_i) = 1$ .

**Solução:** Como  $P(A_i) = 1$ , temos que  $P(A_i^c) = 1 - P(A_i) = 0$ . Logo, pela desigualdade de Boole, temos  $P(\cup_{i=1}^{\infty} A_i^c) \leq \sum_{i=1}^{\infty} P(A_i^c) = 0$ . Logo,  $P(\cup_{i=1}^{\infty} A_i^c) = 0$ . Portanto, como pela Lei de De'Morgan,  $\cap_{i=1}^{\infty} A_i = (\cup_{i=1}^{\infty} A_i^c)^c$ , temos que  $P(\cap_{i=1}^{\infty} A_i) = 1 - P(\cup_{i=1}^{\infty} A_i^c) = 1$ .

**Exemplo 3.5.24:** Demonstre: se  $A_1, A_2, \dots$  e  $B_1, B_2, \dots$  são eventos aleatórios do mesmo espaço de probabilidade tais que  $P(A_n) \rightarrow 1$  e  $P(B_n) \rightarrow p$ , então  $P(A_n \cap B_n) \rightarrow p$ .

**Solução:** Note que

$$\begin{aligned} P(A_n \cap B_n) &= 1 - P((A_n \cap B_n)^c) = 1 - P(A_n^c \cup B_n^c) \\ &\geq 1 - P(A_n^c) - P(B_n^c) = P(A_n) + P(B_n) - 1. \end{aligned} \tag{3.1}$$

Como  $P(B_n) \leq P(A_n \cap B_n) \leq P(A_n) + P(B_n) - 1$ ,  $P(A_n) + P(B_n) - 1 \rightarrow p$  e  $P(B_n) \rightarrow p$ , pelo teorema do confronto (ou sanduíche), temos que  $P(A_n \cap B_n) \rightarrow p$ .

# Capítulo 4

## Probabilidade Condicional

### 4.1 Probabilidade Condicional

Existem várias possíveis interpretações de probabilidade. Por exemplo, pode-se interpretar probabilidade de um evento  $A$  como um limite das freqüências relativas de ocorrência do evento  $A$  em realizações independentes de um experimento. Por outro lado, a interpretação subjetiva de probabilidade associa a probabilidade de um evento  $A$  com o grau de crença pessoal que o evento  $A$  ocorrerá. Em ambos os casos, probabilidade é baseada em informação e conhecimento. Revisão desta base de informação ou conhecimento pode levar a revisão do valor da probabilidade. Em particular, conhecimento que determinado evento ocorreu pode influenciar na probabilidade dos demais eventos.

Considerando-se a interpretação freqüentista de probabilidade, suponha que estejamos interessados em saber qual a probabilidade de um dado evento  $A$ , visto que sabe-se que um dado evento  $B$  ocorreu. Suponha que realizasse um experimento  $n$  vezes das quais o evento  $A$  (resp.,  $B$  e  $A \cap B$ ) ocorre  $N_A$  (resp.,  $N_B > 0$  e  $N_{A \cap B}$ ) vezes. Seja  $r_A = N_A/n$  a freqüência relativa do evento  $A$  nestas  $n$  realizações do experimento. A probabilidade condicional de  $A$  dado que sabe-se que  $B$  ocorreu segundo esta interpretação freqüentista, sugere que ela deve ser igual ao limite das freqüências relativas condicionais do evento  $A$  dado o evento  $B$ , isto é, ela deve ser o limite da razão  $N_{A \cap B}/N_B$  quando  $n$  tende ao infinito. É fácil provar que esta razão é igual a  $r_{A \cap B}/r_B$ , que por sua vez segundo a interpretação freqüentista de probabilidade é aproximadamente igual a  $P(A \cap B)/P(B)$  para valores grandes de  $n$ .

Considerando-se uma interpretação mais subjetiva suponha que a incerteza de um agente é descrita por uma probabilidade  $P$  em  $(\Omega, \mathcal{A})$  e que o agente observa ou fica sabendo que o evento  $B$  ocorreu. Como o agente deve atualizar sua probabilidade  $P(\cdot|B)$  de modo a incorporar esta nova informação? Claramente, se o agente *acredita* que  $B$  é verdadeiro, então parece razoável requerer que

$$P(B^c|B) = 0 \tag{4.1}$$

Em relação aos eventos contidos em  $B$ , é razoável assumir que sua chance relativa permanece inalterada se tudo que o agente descobriu foi que o evento  $B$  ocorreu, ou seja, se

$A_1, A_2 \subseteq B$  com  $P(A_2) > 0$ , então

$$\frac{P(A_1)}{P(A_2)} = \frac{P(A_1|B)}{P(A_2|B)} \quad (4.2)$$

Segue que (4.1) e (4.2) determinam completamente  $P(\cdot|B)$  se  $P(B) > 0$ .

**Teorema 4.1.1:** Se  $P(B) > 0$  e  $P(\cdot|B)$  é uma medida de probabilidade em  $\Omega$  que satisfaz (4.1) e (4.2), então

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

**Prova:** Como  $P(\cdot|B)$  é uma medida de probabilidade e satisfaz  $P(B^c|B) = 0$ , nós temos que  $P(B|B) = 1 - P(B^c|B) = 1$ . Considerando  $A_1 = A$  e  $A_2 = B$  em (4.2), temos então  $P(A|B) = \frac{P(A)}{P(B)}$  para  $A \subseteq B$ . Se  $A$  não é um subconjunto de  $B$ , temos que  $A = (A \cap B) \cup (A \cap B^c)$ . Como  $(A \cap B)$  e  $(A \cap B^c)$  são eventos disjuntos, temos  $P(A|B) = P(A \cap B|B) + P(A \cap B^c|B)$ . Como  $A \cap B^c \subseteq B^c$  e  $P(B^c|B) = 0$ , temos que  $P(A \cap B^c|B) = 0$ . Como  $A \cap B \subseteq B$ , usando o caso anterior

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

■

Deste modo as interpretações frequentista e subjetivista de probabilidade justificam a seguinte definição.

**Definição 4.1.2:** Seja  $(\Omega, \mathcal{A}, P)$  um espaço de probabilidade. Se  $A, B \in \mathcal{A}$  e  $P(B) > 0$  a *probabilidade condicional de A dado B* é definida por

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

Vamos provar que para um evento fixo  $B$  que satisfaz  $P(B) > 0$ ,  $P(\cdot|B)$  satisfaz os axiomas K1-K4 acima e realmente é uma medida de probabilidade. Para provar K1, note que para todo  $A \in \mathcal{A}$ , como  $P(A \cap B) \geq 0$ , nós temos

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0.$$

Para provar K2, note que  $\Omega \cap B = B$ , então

$$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Finalmente, para provar K4' (que implica K3), note que se  $A_1, A_2, \dots$  são mutuamente exclusivos  $A_1 \cap B, A_2 \cap B, \dots$  também o são, então

$$\begin{aligned} P(\cup_i A_i|B) &= \frac{P((\cup_i A_i) \cap B)}{P(B)} = \frac{P(\cup_i (A_i \cap B))}{P(B)} \\ &= \frac{\sum_i P(A_i \cap B)}{P(B)} = \sum_i P(A_i|B). \end{aligned}$$

A probabilidade condicional também satisfaz as seguintes propriedades:



1.  $P(B|B) = 1$ ;
2.  $P(A|B) = P(A \cap B|B)$ ;
3. se  $A \supseteq B$ , então  $P(A|B) = 1$ ;
4.  $P(A \cap B|C) = P(A|B \cap C)P(B|C)$ .

Fazendo  $C = \Omega$  na propriedade 4 acima, temos que:

$$P(A \cap B) = P(A|B)P(B).$$

Utilizando indução matemática, pode-se facilmente provar que

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1) \dots P(A_n|A_1 \cap \dots \cap A_{n-1}).$$

Um método de se obter uma probabilidade (incondicional) de uma probabilidade condicional é utilizando o Teorema da Probabilidade Total. Antes de enunciar este teorema precisamos recordar o que é uma *partição* do espaço amostral. Uma seqüência de eventos  $A_1, A_2, A_3, \dots$  é uma partição do espaço amostral  $\Omega$  se estes eventos são mutuamente exclusivos e contém todos os elementos de  $\Omega$  ( $\cup_i A_i = \Omega$ ).

**Teorema 4.1.3:** *Seja a seqüência de eventos  $B_1, B_2, \dots$  uma partição de  $\Omega$ , então para todo  $A \in \mathcal{A}$*

$$P(A) = \sum_{i:P(B_i) \neq 0} P(A|B_i)P(B_i)$$

**Prova:**

Como  $B_1, B_2, \dots$  é uma partição de  $\Omega$ , temos que

$$A = A \cap \Omega = A \cap (\cup_i B_i) = \cup_i (A \cap B_i).$$

Como os eventos  $B_i$ 's são mutuamente exclusivos, os eventos  $(A \cap B_i)$ 's também são mutuamente exclusivos. Então axioma *K3* implica que

$$\begin{aligned} P(A) &= P(\cup_i (A \cap B_i)) = \sum_i P(A \cap B_i) \\ &= \sum_{i:P(B_i) \neq 0} P(A \cap B_i) = \sum_{i:P(B_i) \neq 0} P(A|B_i)P(B_i). \end{aligned}$$

■

Se nós interpretarmos a partição  $B_1, B_2, \dots$  como possíveis causas e o evento  $A$  corresponda a um efeito particular associado a uma causa,  $P(A|B_i)$  especifica a relação estocástica entre a causa  $B_i$  e o efeito  $A$ .

Por exemplo, seja  $\{D, D^c\}$  uma partição do espaço amostral, onde o evento  $D$  significa que um dado indivíduo possui uma certa doença. Seja  $A$  o evento que determinado teste para

o diagnóstico da doença deu positivo. Então,  $P(A|D^c)$  descreve a probabilidade do exame dá positivo mesmo que o paciente esteja saudável, é a chamada probabilidade de *falso positivo*.  $P(A^c|D)$  é a probabilidade do exame dá negativo mesmo que o paciente esteja doente, é a chamada probabilidade de *falso negativo*. Estas probabilidades determinam a qualidade do teste, quanto menores as probabilidades de falso negativo e falso positivo melhor a qualidade do teste. Caso as probabilidades  $P(D)$ ,  $P(A|D)$ ,  $P(A|D^c)$  sejam conhecidas pode-se usando o Teorema da Probabilidade Total obter a probabilidade incondicional de determinado exame dar positivo  $P(A)$ . Porém geralmente, o que se busca é saber que dado que o resultado de um exame deu positivo qual a probabilidade de que o indivíduo esteja doente. Pode-se obter esta probabilidade utilizando a famosa *fórmula de Bayes*:

$$P(D|A) = \frac{P(A \cap D)}{P(A \cap D) + P(A \cap D^c)} = \frac{P(A|D)P(D)}{P(A|D)P(D) + P(A|D^c)P(D^c)}.$$

Mais geralmente, quando temos uma partição  $B_1, B_2, \dots$ , temos que a fórmula de Bayes é dada por:

$$\begin{aligned} P(B_i|A) &= \frac{P(A \cap B_i)}{\sum_j P(A \cap B_j)} = \frac{P(A \cap B_i)}{\sum_{j:P(B_j) \neq 0} P(A \cap B_j)} \\ &= \frac{P(A|B_i)P(B_i)}{\sum_{j:P(B_j) \neq 0} P(A|B_j)P(B_j)}. \end{aligned}$$

É fácil de provar esta fórmula usando o Teorema da Probabilidade Total. As probabilidades  $P(B_i)$  são usualmente chamadas de probabilidades *a priori* e as probabilidades condicionais  $P(B_i|A)$  são chamadas de probabilidades *a posteriori*. O seguinte exemplo ilustra uma aplicação da fórmula de Bayes.

**Exemplo 4.1.4:** Considere uma imagem formada por  $n \times m$  pixels com a  $k$ -ésima linha contendo  $d_k (\leq m)$  pixels defeituosos. No primeiro estágio do experimento uma linha é escolhida ao acaso e nós não sabemos qual foi a escolha. Nós então examinamos um pixel selecionada ao acaso nesta linha e descobrimos que o pixel é defeituoso (chamamos este evento de  $D$ ). Qual a probabilidade de que este pixel defeituoso esteja na linha  $k$ ? Seja  $R = k$  o evento que este pixel pertencia a  $k$ -ésima linha da imagem. A fórmula de Bayes nos permite determinar que dado que

$$P(R = k) = \frac{1}{n} \quad \text{e} \quad P(D|R = k) = \frac{d_k}{m},$$

nós temos que

$$P(R = k|D) = \frac{\frac{1}{n} \frac{d_k}{m}}{\sum_{i=1}^n \frac{1}{n} \frac{d_i}{m}} = \frac{d_k}{\sum_{i=1}^n d_i}.$$

Então, mesmo que a linha tenha inicialmente sido escolhida ao acaso, dado o evento que encontramos ao acaso um pixel defeituoso nesta linha, agora é mais provável que seja uma linha contendo um número grande de pixels defeituosos  $d_k$ .

**Exemplo 4.1.5:** Um canal de comunicação binário envia um dentre dois tipos de sinais, denotados por 0 e 1. Devido ao ruído, um 0 transmitido é alguma vezes recebido como um 1 e um 1 transmitido é alguma vezes recebido como um 0. Para um dado canal, assuma uma probabilidade de 0.94 que um 0 transmitido seja corretamente recebido como um 0 e uma probabilidade de 0.91 que um 1 transmitido seja corretamente recebido como um 1. Adicionalmente, assuma uma probabilidade de 0.45 de se transmitir um 0. Se um sinal é enviado, determine,

- (a) A probabilidade de que um 1 seja recebido.
- (b) A probabilidade de que um 0 seja recebido.
- (c) A probabilidade de que um 1 foi transmitido, dado que um 1 foi recebido.
- (d) A probabilidade de que um 0 foi transmitido, dado que um zero foi recebido.
- (e) A probabilidade de um erro.

Sejam os eventos

$$T_0 = \{\text{um 0 é transmitido}\},$$

$$T_1 = \{\text{um 1 é transmitido}\},$$

$$R_0 = \{\text{um 0 é recebido}\},$$

$$R_1 = \{\text{um 1 é recebido}\}.$$

Logo,

$$P(R_0 | T_0) = 0.94 \Rightarrow P(R_1 | T_0) = 0.06,$$

$$P(R_1 | T_1) = 0.91 \Rightarrow P(R_0 | T_1) = 0.09,$$

$$P(T_0) = 0.45,$$

$$P(T_1) = 0.55.$$

(a)

$$R_1 = (R_1 \cap T_1) \cup (R_1 \cap T_0),$$

logo,

$$P(R_1) = P(R_1 | T_1)P(T_1) + P(R_1 | T_0)P(T_0) = 0.91 \times 0.55 + 0.06 \times 0.45 = 0.5275.$$

(b)

$$R_0 = (R_0 \cap T_0) \cup (R_0 \cap T_1),$$

logo,

$$P(R_0) = P(R_0 | T_0)P(T_0) + P(R_0 | T_1)P(T_1) = 0.94 \times 0.45 + 0.09 \times 0.55 = 0.4725,$$

ou,

$$P(R_0) = 1 - P(R_1) = 1 - 0.5275 = 0.4725.$$

(c)

$$\begin{aligned}
 P(T_1 | R_1) &= \frac{P(T_1 \cap R_1)}{P(R_1)} \\
 &= \frac{P(R_1 | T_1)P(T_1)}{P(R_1)} \\
 &= \frac{0.91 \times 0.55}{0.5275} = 0.9488.
 \end{aligned}$$

(d)

$$\begin{aligned}
 P(T_0 | R_0) &= \frac{P(T_0 \cap R_0)}{P(R_0)} \\
 &= \frac{P(R_0 | T_0)P(T_0)}{P(R_0)} \\
 &= \frac{0.94 \times 0.45}{0.4725} = 0.8952.
 \end{aligned}$$

(e)

$$E = \{\text{acontece um erro}\}.$$

Logo,

$$E = (T_1 \cap R_0) \cup (T_0 \cap R_1),$$

$$P(E) = P(R_0 | T_1)P(T_1) + P(R_1 | T_0)P(T_0) = 0.09 \times 0.55 + 0.06 \times 0.45 = 0.0765.$$

**Exemplo 4.1.6:** Uma urna contém 4 bolas brancas e 6 bolas pretas. Sacam-se, sucessivamente e sem reposição, duas bolas dessa urna. Determine a probabilidade da primeira bola ser branca sabendo que a segunda bola é branca.

**Solução:** Sejam  $B_1$  e  $B_2$  os eventos a primeira bola é branca e a segunda bola é branca, respectivamente. Queremos calcular  $P(B_1|B_2)$ . Utilizando a fórmula de Bayes, temos

$$P(B_1|B_2) = \frac{P(B_2|B_1)P(B_1)}{P(B_2|B_1)P(B_1) + P(B_2|B_1^c)P(B_1^c)}.$$

Mas  $P(B_2|B_1) = \frac{3}{9}$ ,  $P(B_2|B_1^c) = \frac{4}{9}$ ,  $P(B_1) = \frac{4}{10}$  e  $P(B_1^c) = \frac{6}{10}$ . Logo,

$$P(B_1|B_2) = \frac{\frac{3}{9} \cdot \frac{4}{10}}{\frac{3}{9} \cdot \frac{4}{10} + \frac{4}{9} \cdot \frac{6}{10}} = \frac{\frac{2}{15}}{\frac{2}{5}} = \frac{1}{3}.$$

Embora probabilidade condicional seja bastante útil, ela sofre de alguns problemas, em particular quando se quer tratar de eventos de probabilidade zero. Tradicionalmente, se  $P(B) = 0$ , então  $P(A|B)$  não é definida. Isto leva a um número de dificuldades filosóficas em relação a eventos com probabilidade zero. São eles realmente impossíveis? Caso contrário, quão improvável um evento precisa ser antes de ele ser atribuído probabilidade zero? Deve um evento em algum caso ser atribuído probabilidade zero? Se existem eventos com

probabilidade zero que não são realmente impossíveis, então o que significa condicionar em eventos de probabilidade zero? Por exemplo, considere o espaço de probabilidade  $([0, 1], \mathcal{B}, \mu)$  onde  $\mathcal{B}$  é a  $\sigma$ -álgebra de Borel restrita a eventos contidos em  $[0, 1]$  e  $\mu$  é uma medida de probabilidade na qual todo intervalo em  $[0, 1]$  possui probabilidade igual ao seu comprimento. Seja  $B = \{1/4, 3/4\}$  e  $A = \{1/4\}$ . Como  $P(B) = 0$ ,  $P(A|B)$  não é definida. Porém parece razoável assumir que neste caso  $P(A|B) = 1/2$  já que  $\mu$  intuitivamente implica que todos os estados são equiprováveis, mas a definição formal de probabilidade condicional não nos permite obter esta conclusão.

Alguns dos problemas mencionados no parágrafo anterior podem ser tratados considerando-se probabilidades condicionais (e não probabilidade incondicionais) como a noção fundamental, porém a discussão destes modelos está fora do escopo deste curso.

**Exemplo 4.1.7:** Se  $P(C|D) = 0,4$  e  $P(D|C) = 0,5$ , que evento é mais provável  $C$  ou  $D$ ?

**Solução:**

**Exemplo 4.1.8:** Se  $P(E) = 0,4$  e  $P(F) = 0,7$ , o que pode-se concluir sobre  $P(E|F)$ ?

**Solução:** Por definição, temos que:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Porém, sabemos que  $\max(P(E) + P(F) - 1, 0) \leq P(E \cap F) \leq \min(P(E), P(F))$ . Logo,  $0,1 \leq P(E \cap F) \leq 0,4$ , portanto

$$\frac{0,1}{0,7} \leq P(E|F) \leq \frac{0,4}{0,7}.$$

**Exemplo 4.1.9: (Paradoxo de Monty Hall)** Monty Hall foi um popular apresentador de programa de jogos em TV cujo jogo começava mostrando ao participante 3 portas fechadas  $d_1, d_2, d_3$ , e atrás de apenas uma delas havia um prêmio valioso. O participante selecionava uma porta, por exemplo,  $d_1$ , mas antes que a porta fosse aberta, Monty Hall, que sabia em que porta estava o prêmio, por exemplo,  $d_2$ , abria a porta restante  $d_3$ , que não continha o prêmio. O participante tinha então permissão para ficar com sua porta original,  $d_1$ , ou escolher a outra porta fechada. A pergunta é se é melhor ficar com a porta original ou trocar de porta. Vamos agora utilizar a fórmula de Bayes para analisar este problema. Seja  $G$  uma porta escolhida aleatoriamente para conter o prêmio;  $Y$  a porta que o participante escolhe primeiro; e  $M$  a porta que Monty Hall abre. O participante não tem nenhum conhecimento a priori sobre a localização do prêmio, ou seja ele considera todas as portas equiprováveis, e isto pode ser modelado por:

$$P(G = d_i | Y = d_j) = \frac{1}{3};$$

todas as portas tem a mesma probabilidade de conter o prêmio não importa qual porta o participante escolhe. Se o participante escolher uma porta que não contém o prêmio, Monty Hall necessariamente terá de abrir a porta que não contém o prêmio, isto pode ser modelado por:

$$P(M = d_{i_1} | Y = d_{i_2}, G = d_{i_3}) = 1,$$

onde  $i_1, i_2, i_3 \in \{1, 2, 3\}$  e são distintos. Se o participante escolher corretamente, por exemplo,  $Y = G = d_{i_2}$ , então assumimos que Monty Hall escolhe aleatoriamente entre as outras duas portas:

$$P(M = d_{i_1} | Y = G = d_{i_2}) = \frac{1}{2}, \text{ para } d_{i_1} \neq d_{i_2}.^1$$

Para determinar se o participante deve trocar de porta, devemos calcular

$$\begin{aligned} P(G = d_1 | Y = d_2, M = d_3) &= \frac{P(G = d_1, Y = d_2, M = d_3)}{P(Y = d_2, M = d_3)} \\ &= \frac{P(M = d_3 | G = d_1, Y = d_2)P(G = d_1 | Y = d_2)P(Y = d_2)}{P(M = d_3 | Y = d_2)P(Y = d_2)} \\ &= \frac{P(M = d_3 | G = d_1, Y = d_2)P(G = d_1 | Y = d_2)}{P(M = d_3 | Y = d_2)} \\ &= \frac{1/3}{1/3} \end{aligned}$$

Para determinar o valor de  $P(M = d_3 | Y = d_2)$  utilizamos o Teorema da Probabilidade Total e a definição de probabilidade condicional:

$$\begin{aligned} P(M = d_3 | Y = d_2) &= \frac{P(Y = d_2, M = d_3)}{P(Y = d_2)} \\ &= \frac{P(Y = d_2, M = d_3, G = d_1) + P(Y = d_2, M = d_3, G = d_2) + P(Y = d_2, M = d_3, G = d_3)}{P(Y = d_2)} \\ &= \frac{P(M = d_3 | Y = d_2, G = d_1)P(G = d_1 | Y = d_2)P(Y = d_2)}{P(Y = d_2)} \\ &\quad + \frac{P(M = d_3 | Y = d_2, G = d_2)P(G = d_2 | Y = d_2)P(Y = d_2)}{P(Y = d_2)} \\ &\quad + \frac{P(M = d_3 | Y = d_2, G = d_3)P(G = d_3 | Y = d_2)P(Y = d_2)}{P(Y = d_2)} \\ &= P(M = d_3 | Y = d_2, G = d_1)P(G = d_1 | Y = d_2) \\ &\quad + P(M = d_3 | Y = d_2, G = d_2)P(G = d_2 | Y = d_2) \\ &\quad + P(M = d_3 | Y = d_2, G = d_3)P(G = d_3 | Y = d_2) \\ &= 1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 = \frac{1}{2}. \end{aligned}$$

Logo,  $P(G = d_1 | Y = d_2, M = d_3) = \frac{2}{3}$ , e o participante deve trocar de porta de sua escolha original  $d_2$  para  $d_1$ !

**Exemplo 4.1.10:** Seja  $D$  o evento que um indivíduo selecionado ao acaso de uma população tem uma doença particular,  $D^c$  seu complemento. A probabilidade que um indivíduo selecionado ao acaso nesta população tenha determinada doença é  $p_d$ . Existe um teste para diagnóstico desta doença que sempre acusa presença da doença quando o indivíduo tem a

<sup>1</sup>A solução depende como resolvemos este caso.

doença. Contudo, quando o indivíduo não tem a doença, o teste reporta falsamente que o indivíduo tem a doença com probabilidade  $p_t$ . Seja  $TP$  o evento que o teste reporta positivamente que o indivíduo tem a doença. Formalmente, temos:

$$P(D) = p_d, P(TP|D) = 1, P(TP|D^c) = p_t.$$

Um indivíduo deve estar interessado em saber a probabilidade  $P(D|TP)$  que ele tenha a doença dado que o teste deu positivo. Se, por exemplo, a doença for rara e  $p_d = 0,001$ , e o teste reportar falsamente com probabilidade pequena  $p_t = 0,05$ , veremos que apesar desta pequena probabilidade do teste de um resultado errado, a probabilidade do indivíduo ter a doença é pequena. Pela fórmula de Bayes

$$P(D|TP) = \frac{P(TP|D)P(D)}{P(TP|D)P(D) + P(TP|D^c)P(D^c)} = \frac{p_d}{p_d + p_t(1 - p_d)} = 0,02.$$

**Exemplo 4.1.11:** Sabemos que os eventos  $\{B_1, B_2, B_3\}$  são disjuntos par a par e que sua união é igual ao espaço amostral. Estes eventos tem as seguintes probabilidades  $P(B_1) = 0,2$  e  $P(B_2) = 0,3$ . Existe um outro evento  $A$  que sabemos que  $P(A|B_1) = 0,3$ ;  $P(A|B_2) = 0,4$ ; e  $P(A|B_3) = 0,1$ . Calcule:

- (a)  $P(A)$
- (b)  $P(B_2|A)$

**Exemplo 4.1.12:** Suponha que todos os bytes tenham a mesma probabilidade. Seja  $W$  o número de 1's em um byte. Considere os seguintes eventos:

$$A = \{\text{O primeiro e o segundo bit são iguais a 1, e}\}$$

$$B = \{W \text{ é um número ímpar.}\}$$

Calcule:

- (a)  $P(A)$
- (b)  $P(B)$
- (c)  $P(B|A)$
- (d)  $P(A|B)$

**Solução:**

$$P(A) = \frac{||A||}{||\Omega||} = \frac{2^6}{2^8} = \frac{1}{4}.$$

$$P(B) = \frac{||B||}{||\Omega||} = \frac{\binom{8}{1} + \binom{8}{3} + \binom{8}{5} + \binom{8}{7}}{2^8} = \frac{1}{2}.$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)},$$

onde  $P(A \cap B) = \frac{|A \cap B|}{\Omega} = \frac{\binom{6}{1} + \binom{6}{3} + \binom{6}{5}}{2^8} = \frac{1}{8}$ . Portanto,

$$P(B|A) = \frac{\frac{1}{8}}{\frac{1}{4}} = \frac{1}{2}.$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}.$$

**Exemplo 4.1.13:** Se jogarmos dois dados um após o outro e observamos o evento que a soma dos dois dados é igual a 9, então qual a probabilidade do primeiro dado ter dado resultado 4?

**Solução:**

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{36}}{\frac{4}{36}} = \frac{1}{4}.$$

**Exemplo 4.1.14:** Em um teste de múltipla escolha, a probabilidade do aluno saber a resposta da questão é  $p$ . Havendo  $m$  escolhas, se ele sabe a resposta ele responde corretamente com probabilidade 1; se não sabe ele responde corretamente com probabilidade  $\frac{1}{m}$ .

- (a) Qual a probabilidade que a pergunta foi respondida corretamente?
- (b) Qual a probabilidade que o aluno sabia a resposta dado que a pergunta foi respondida corretamente?

**Solução:** Para a parte (a), usamos o Teorema da Probabilidade Total:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c) = 1 \cdot p + \frac{1}{m}(1 - p).$$

Para a parte (b), usamos a fórmula de Bayes

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)} = \frac{1 \cdot p}{1 \cdot p + \frac{1}{m}(1 - p)}$$

## 4.2 Independência

O que exatamente significa que dois eventos são independentes? Intuitivamente, isto significa que eles não têm nada haver um com o outro, eles são totalmente não relacionados; a ocorrência de um não tem nenhuma influência sobre o outro. Por exemplo, suponha que duas diferentes moedas são lançadas. A maioria das pessoas viria os resultados desses lançamentos como independentes. Portanto, a intuição por trás da frase “o evento  $A$  é independente do evento  $B$ ” é que nosso conhecimento sobre a tendência para  $A$  ocorrer dado que sabemos que  $B$  ocorreu não é alterada quando ficamos sabendo que  $B$  ocorreu. Então, usando probabilidades condicionais podemos formalizar esta intuição da seguinte forma,  $A$  é independente de  $B$  se  $P(A|B) = P(A)$ . Mas usando a definição de probabilidade condicional, chega-se a seguinte conclusão  $A$  é independente de  $B$  se  $P(A \cap B) = P(A)P(B)$ . Como esta última expressão é definida inclusive para o caso de  $P(B) = 0$ , ela é a expressão adotada como a definição de independência entre eventos.



**Definição 4.2.1:** O evento  $A$  é independente do evento  $B$  se  $P(A \cap B) = P(A)P(B)$ .

Note que esta definição de independência implica que independência é um conceito simétrico em teoria da probabilidade, isto é,  $A$  é independente de  $B$  se e somente se  $B$  é independente de  $A$ . Note que esta definição também implica que eventos  $A$  e  $B$  são independentes se  $P(A) = 0$  ou  $P(B) = 0$ , o que pode gerar algumas conclusões não intuitivas se de fato  $P(A) = 0$  ou  $P(B) = 0$ . Por exemplo, se  $P(A) = 0$ , então  $A$  é independente dele mesmo, porém  $A$  certamente não é não relacionado consigo mesmo. Similarmente, é fácil provar que se  $P(A) = 1$ ,  $A$  é independente dele mesmo. O seguinte teorema prova que estes são os únicos casos em que um evento é independente dele mesmo.

**Teorema 4.2.2:**  $A$  é independente dele mesmo se e somente se  $P(A) = 0$  ou  $P(A) = 1$ .

**Prova:**

$$P(A \cap A) = P(A) = P(A)P(A) \Leftrightarrow P(A) = 0 \text{ ou } P(A) = 1.$$

■

Intuitivamente, se  $A$  é independente de  $B$  o fato que  $B$  não ocorreu, ou seja que  $B^c$  ocorreu, não deve alterar a probabilidade de  $A$ . Portanto, é de se esperar que se  $A$  e  $B$  são independentes, então  $A$  e  $B^c$  também são. O seguinte teorema prova que esta intuição é verdadeira.

**Teorema 4.2.3:** Se  $A$  e  $B$  são eventos independentes,  $A$  e  $B^c$  (resp.,  $A^c$  e  $B$ ,  $A^c$  e  $B^c$ ) também o são.

**Prova:** Note que

$$A = A \cap \Omega = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c).$$

Então, como  $A \cap B$  e  $A \cap B^c$  são mutuamente exclusivos, axioma K3 implica que

$$P(A) = P(A \cap B) + P(A \cap B^c).$$

Como  $A$  e  $B$  são independentes, nós temos

$$P(A) = P(A)P(B) + P(A \cap B^c).$$

Rearrajando os termos e utilizando o fato que  $P(B^c) = 1 - P(B)$ , temos  $P(A \cap B^c) = P(A)P(B^c)$ , como queríamos demonstrar. ■

O conceito de independência também se aplica a uma coleção arbitrária de eventos  $\{A_i\}_{i \in \mathcal{I}}$ , onde  $\mathcal{I}$  é um conjunto de índices. Neste caso, têm-se duas definições.

**Definição 4.2.4:** Uma coleção de eventos  $\{A_i\}_{i \in \mathcal{I}}$  é independente par a par se para todo  $i \neq j \in \mathcal{I}$ ,  $A_i$  e  $A_j$  são eventos independentes.

**Definição 4.2.5:** Uma seqüência finita de eventos  $A_1, A_2, \dots, A_n$ ,  $n \geq 1$ , é *mutuamente independente* se para todo  $I \subseteq \{1, \dots, n\}$ ,

$$P(\cap_{i \in I} A_i) = \prod_{i \in I} P(A_i)$$

E uma coleção de eventos  $\{A_i\}_{i \in \mathcal{I}}$  é mutuamente independente se para todo  $J \subseteq \mathcal{I}$  finito,  $\{A_i\}_{i \in J}$  é mutuamente independente.

Considere os seguintes exemplos que ilustram o conceito de independência.

**Exemplo 4.2.6:** Se  $\Omega = \{1, 2, 3, 4\}$  e  $P(\{w\}) = 1/4$ , então  $A = \{1, 2\}$ ,  $B = \{1, 3\}$ , e  $C = \{2, 3\}$  são eventos independentes par a par. Pode-se verificar isto pelo fato que

$$P(A \cap B) = P(\{1\}) = \frac{1}{4} = \frac{1}{2} \frac{1}{2} = P(A)P(B).$$

Similarmente, pode-se provar o mesmo resultado para os outros pares. Contudo, a probabilidade

$$P(A \cap B \cap C) = P(\emptyset) = 0 \neq P(A)P(B)P(C) = \frac{1}{8}.$$

Então,  $A$ ,  $B$ , e  $C$  não são mutuamente independentes.

**Exemplo 4.2.7:** Certo experimento consiste em lançar um dado equilibrado duas vezes, independentemente. Dado que os dois números sejam diferentes, qual é a probabilidade condicional de

- (a) pelo menos um dos números ser 6,
- (b) a soma dos números ser 8?

**Solução:** Para parte (a), note que existem 30 resultados possíveis para os lançamentos do dado de modo que o mesmo número não se repita, dos quais 10 o número 6 ocorre. Portanto, esta probabilidade é igual a  $1/3$ .

Para parte (b), note que existem 4 resultados possíveis que somam 8 dado que os números são diferentes, logo esta probabilidade é igual a  $4/30$ .

**Exemplo 4.2.8:** O evento  $F$  de um determinado sistema falhar ocorre se os eventos  $A_1$  ou  $A_2$  ocorrerem, mas o evento  $A_3$  não ocorrer. Se  $A_1, A_2, A_3$  são mutuamente independentes e  $P(A_1) = 0,4$ ,  $P(A_2) = 0,35$ , e  $P(A_3) = 0,1$ , então calcule  $P(F)$ .

**Solução:**

**Exemplo 4.2.9:** Assuma que  $A_1, \dots, A_n$  são eventos mutuamente independentes e que  $P(A_i) = p_i$ . Nós calculamos as probabilidades dos seguintes eventos:

- O evento  $A$  é o evento que todos estes eventos ocorrem, então

$$P(A) = P(\cap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i) = \prod_{i=1}^n p_i$$

- O evento  $B$  é o evento que nenhum desses eventos ocorre, então

$$P(B) = P(\cap_{i=1}^n A_i^c) = \prod_{i=1}^n P(A_i^c) = \prod_{i=1}^n (1 - p_i)$$

- O evento  $C$  é o evento que pelo menos um desses eventos ocorre, então  $C = B^c$

$$P(C) = P(B^c) = 1 - P(B) = 1 - \prod_{i=1}^n (1 - p_i)$$

# Capítulo 5

## Variável Aleatória Discreta

### 5.1 Introdução

Suponha que uma moeda é lançada cinco vezes. Qual é o número de caras? Esta quantidade é o que tradicionalmente tem sido chamada de *variável aleatória*. Intuitivamente, é uma variável porque seus valores variam, dependendo da sequência de lançamentos da moeda realizada; o adjetivo “aleatória” é usado para enfatizar que o seu valor é de certo modo incerto. Formalmente, contudo, uma variável aleatória não é nem “aleatória” nem é uma variável.

**Definição 5.1.1:** Seja  $(\Omega, \mathcal{A}, P)$  um espaço de probabilidade. Uma função  $X : \Omega \rightarrow R$  é chamada de variável aleatória se para todo número real  $\lambda$ ,  $\{\omega \in \Omega : X(\omega) \leq \lambda\} \in \mathcal{A}$ .

**Exemplo 5.1.2:** Considere três lançamentos de uma moeda honesta. O espaço amostral para este experimento aleatório consiste de todas as possíveis sequências de tamanho 3 de caras e coroas, isto é:

$$\begin{aligned} \Omega = \{ & (cara, cara, cara), (cara, cara, coroa), (cara, coroa, cara), \\ & (cara, coroa, coroa), (coroa, cara, cara), (coroa, cara, coroa), \\ & (coroa, coroa, cara), (coroa, coroa, coroa) \}. \end{aligned}$$

Seja  $\mathcal{A}$  o conjunto de todos os subconjuntos de  $\Omega$ . Neste caso qualquer função real de  $\Omega$  é uma variável aleatória. Por exemplo, seja  $X$  a diferença entre o número de caras e o número de coroas obtidos nos três lançamentos. Então,  $X$  pode assumir quatro valores, 3, 1, -1, ou -3. Nosso objetivo é estudar a probabilidade de  $X$  assumir cada um desses possíveis valores. Como a moeda é honesta cada um dos possíveis resultados em  $\Omega$  tem a mesma probabilidade 1/8. Então, por exemplo, como poderemos obter a probabilidade de  $X$  ser negativo?

Recorde que a  $\sigma$ -álgebra de Borel,  $\mathcal{B}$ , é a menor  $\sigma$ -álgebra de eventos reais que contém todos os intervalos. Pode-se provar que se  $X$  é uma variável aleatória em  $(\Omega, \mathcal{A}, P)$ , então para todo evento  $B \in \mathcal{B}$ , chamado evento Boreliano, temos que  $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}$ .

Deste modo, dada uma variável aleatória  $X$  em  $(\Omega, \mathcal{A}, P)$ , pode-se definir uma probabilidade induzida  $P_X$  no espaço mensurável  $(R, \mathcal{B})$  da seguinte maneira: para todo  $A \in \mathcal{B}$ , definimos  $P_X(A) = P(X^{-1}(A))$ . Como  $X$  é uma variável aleatória, tem-se que  $X^{-1}(A) \in \mathcal{A}$ , então  $P_X$  está bem definida. Resta provar que  $P_X$  satisfaz os axiomas K1, K2, e K4' de probabilidade:

$$\text{K1. } P_X(A) = P(X^{-1}(A)) \geq 0.$$

$$\text{K2. } P_X(R) = P(X^{-1}(R)) = P(\Omega) = 1.$$

K4'. Suponha que  $A_1, A_2, \dots$  são eventos Borelianos disjuntos. Então,

$$P_X(\cup_i A_i) = P(X^{-1}(\cup_i A_i)) = P(\cup_i X^{-1}(A_i)) = \sum_i P(X^{-1}(A_i)) = \sum_i P_X(A_i).$$

**Exemplo 5.1.3:** No exemplo anterior, temos que se o evento de interesse  $A$  são todos os reais negativos, então  $X^{-1}(A)$  são todos os resultados do experimento que nos dão valores negativos para  $X$ , ou seja, são os resultados que contém menos caras que coroas:  $(cara, coroa, coroa)$ ,  $(coroa, cara, coroa)$ ,  $(coroa, coroa, cara)$  e  $(coroa, coroa, coroa)$ . Portanto,  $P_X(A) = 4 \times 1/8 = 1/2$ .

Vale a pena salientar que em muitos problemas, já teremos a informação sobre a distribuição induzida  $P_X$  definida em  $(R, \mathcal{B})$ . Nestes casos, estaremos “esquecendo” a natureza funcional de  $X$  e nos preocupando apenas com os valores assumidos por  $X$ . Estes casos podem ser pensados como se o experimento aleatório fosse descrito por  $(R, \mathcal{B}, P_X)$  e  $X(w) = w, \forall w \in R$ , ou seja, os resultados do experimento aleatório já são numéricos e descrevem a característica de interesse que queremos analisar.

É importante enfatizar que é usual se referir a variáveis aleatórias por letras maiúsculas  $X, Y, Z, \dots$  e aos valores que tais variáveis podem assumir por letras minúsculas  $x, y, z, \dots$ . Muitas vezes escreve-se  $P(X \in A)$  para representar  $P(\{w \in \Omega : X(w) \in A\})$ . Por exemplo,  $P(X \leq 5) = P(\{w \in \Omega : X(w) \leq 5\})$ .

**Exemplo 5.1.4:** Considere que lançamos 3 vezes uma moeda que tem probabilidade de cair cara igual  $2/3$ . Seja  $X$  o número de coroas obtido. Determine:

(a)  $P(X < 3)$ .

(b)  $P(1 < X < 3)$ .

(c)  $P(X > 1 | X < 3)$ .

■

## 5.2 Função de Distribuição Acumulada

Para uma variável aleatória  $X$ , uma maneira simples e básica de descrever a probabilidade induzida  $P_X$  é utilizando sua *função de distribuição acumulada*.

**Definição 5.2.1:** A função de distribuição acumulada de uma variável aleatória  $X$ , representada por  $F_X$ , é definida por

$$F_X(x) = P_X((-\infty, x]), \forall x \in \mathbb{R}.$$

A função de distribuição acumulada  $F_X$  satisfaz as seguintes propriedades:

F1. Se  $x \leq y$ , então  $F_X(x) \leq F_X(y)$ .

$$x \leq y \Rightarrow (-\infty, x] \subseteq (-\infty, y] \Rightarrow P_X((-\infty, x]) \leq P_X((-\infty, y]) \Rightarrow F_X(x) \leq F_X(y).$$

F2. Se  $x_n \downarrow x$ , então  $F_X(x_n) \downarrow F_X(x)$ .

Se  $x_n \downarrow x$ , então os eventos  $(-\infty, x_n]$  são decrescentes e  $\cap_n (-\infty, x_n] = (-\infty, x]$ . Logo, pela continuidade da medida de probabilidade, tem-se que  $P_X((-\infty, x_n]) \downarrow P((-\infty, x])$ , ou seja,  $F_X(x_n) \downarrow F_X(x)$ .

F3. Se  $x_n \downarrow -\infty$ , então  $F_X(x_n) \downarrow 0$ , e se  $x_n \uparrow \infty$ , então  $F_X(x_n) \uparrow 1$ .

Se  $x_n \downarrow -\infty$ , então os eventos  $(-\infty, x_n]$  são decrescentes e  $\cap_n (-\infty, x_n] = \emptyset$ . Logo, pela continuidade da medida de probabilidade, tem-se que  $P_X((-\infty, x_n]) \downarrow P(\emptyset)$ , ou seja,  $F_X(x_n) \downarrow 0$ . Similarmente, se  $x_n \uparrow \infty$ , então os eventos  $(-\infty, x_n]$  são crescentes e  $\cup_n (-\infty, x_n] = \mathbb{R}$ . Logo, pela continuidade da medida de probabilidade, tem-se que  $P_X((-\infty, x_n]) \uparrow P(\Omega)$ , ou seja,  $F_X(x_n) \uparrow 1$ .

**Teorema 5.2.2:** Uma função real  $G$  satisfaz F1-F3 se e somente se  $G$  é uma distribuição de probabilidade acumulada.

**Prova:** A prova de que se  $G$  for uma distribuição de probabilidade acumulada, então  $G$  satisfaz F1-F3 foi dada acima. A prova de que toda função real que satisfaz F1-F3 é uma função de probabilidade acumulada é complexa envolvendo o Teorema da Extensão de Carathéodory, e está fora do escopo deste curso. ■

Condição F2 significa que toda função distribuição de probabilidade acumulada  $F_X$  é contínua à direita. Ainda mais, como  $F_X$  é não-decrescente e possui valores entre 0 e 1, pode-se provar que ela tem um número enumerável de descontinuidades do tipo salto. Pela continuidade à direita, o salto no ponto  $x$  é igual a

$$\begin{aligned} F_X(x) - F_X(x^-) &= F_X(x) - \lim_{n \rightarrow \infty} F(x - \frac{1}{n}) \\ &= P_X((-\infty, x]) - \lim_{n \rightarrow \infty} P_X((-\infty, x - \frac{1}{n}]) \\ &= \lim_{n \rightarrow \infty} P_X((x - \frac{1}{n}, x]). \end{aligned}$$

Como a seqüência de eventos  $(x - \frac{1}{n}, x]$  é decrescente e  $\cap_n(x - \frac{1}{n}, x] = \{x\}$ . Temos que  $\{x\}$  é Boreliano e

$$P_X(x) = F_X(x) - F_X(x^-).^1 \quad (5.1)$$

Ou seja, a probabilidade da variável aleatória  $X$  assumir o valor  $x$  é igual ao salto da função de distribuição acumulada  $F_X$  no ponto  $x$ .

**Exemplo 5.2.3:** Determine quais das seguintes funções são funções de distribuição acumuladas, especificando a propriedade que não for satisfeita caso a função não seja uma distribuição acumulada.

- (a)  $\frac{e^x}{1+e^x}$
- (b)  $I_{[0, \text{infy})}(x) + [1 - I_{[0, \text{infy})}(x)](1 + e^x)/2$
- (c)  $e^{-|x|}$
- (d)  $I_{[0, \text{infy})}(x)$
- (e)  $I_{(0, \text{infy})}(x)$

**Exemplo 5.2.4:** Considere a seguinte função  $G(x)$ .

$$G(x) = \begin{cases} a - 2b & \text{se } x < 0, \\ ax & \text{se } 0 \leq x < 1, \\ a + b(x - 1) & \text{se } 1 \leq x < 2, \\ 1 & \text{se } x \geq 2. \end{cases}$$

- (a) Determine as restrições que as constantes  $a$  e  $b$  devem satisfazer para que a função  $G(x)$  seja função de distribuição acumulada de alguma variável aleatória  $X$ .
- (b) Determine o valor de  $P(1/2 \leq X \leq 3/2)$  em função de  $a$  e  $b$ .

**Exemplo 5.2.5:** Seja  $K$  o número de íons emitidos por uma fonte em um tempo  $T$ . Se  $F_K(1) - F_K(1/2) = 0,1$ , qual o valor de  $P(K = 1)$ ?

**Exemplo 5.2.6:** Uma seqüência de 10 bytes independentes foi recebida. É sabido que a probabilidade é igual a 0,3 que o primeiro símbolo de um byte sea igual a 0. Seja  $K$  o número de bytes recebidos tendo 0 como primeiro símbolo.

- (a) Calcule  $P(K = 2)$
- (b) Calcule  $F_K(1)$

---

<sup>1</sup> $F_X(a^-) = \lim_{x \rightarrow a^-} F_X(x)$  é o limite de  $F_X(x)$  quando  $x$  tende a  $a$  por valores menores que  $a$ , ou seja, o limite a esquerda  $F_X(x)$  quando  $x$  tende a  $a$ .

**Exemplo 5.2.7:** Este exemplo mostra como usar a função de distribuição acumulada para calcular probabilidades.

Lembrando que

$$F_X(x) = P(X \leq x) = P_X((-\infty, x]).$$

$$\begin{aligned} \text{(a)} \quad (-\infty, b] &= (-\infty, a] \cup (a, b], \quad a \leq b \Rightarrow \\ P_X((-\infty, b]) &= P_X((-\infty, a]) + P_X((a, b]) \Rightarrow \\ P_X((a, b]) &= P_X((-\infty, b]) - P_X((-\infty, a]) = F_X(b) - F_X(a) \Rightarrow \end{aligned}$$

$$P(a < X \leq b) = F_X(b) - F_X(a). \quad (5.2)$$

$$\begin{aligned} \text{(b)} \quad (a, b) \cup \{b\} &= (a, b] \Rightarrow \\ P_X((a, b)) + P_X(\{b\}) &= P_X((a, b]) \Rightarrow \\ P_X((a, b)) &= P_X((a, b]) - P_X(b) = F_X(b) - F_X(a) - P(X = b) \Rightarrow \end{aligned}$$

$$P(a < X < b) = F_X(b^-) - F_X(a). \quad (5.3)$$

O resultado em 5.3 foi obtido usando 5.1 e 5.2.

$$\begin{aligned} \text{(c)} \quad (a, b] \cup \{a\} &= [a, b] \Rightarrow \\ P_X((a, b]) + P_X(a) &= P_X([a, b]) \Rightarrow \\ P_X([a, b]) &= P_X((a, b]) + P(X = a) = F_X(b) - F_X(a) + P(X = a) \Rightarrow \end{aligned}$$

$$P(a \leq X \leq b) = F_X(b) - F_X(a^-). \quad (5.4)$$

O resultado em 5.4 foi obtido usando 5.1 e 5.2.

$$\begin{aligned} \text{(d)} \quad [a, b) &= (a, b) \cup \{a\} \Rightarrow \\ P_X([a, b)) &= P_X((a, b)) + P_X(a) = F_X(b) - F_X(a) - P(X = b) + P(X = a) \Rightarrow \end{aligned}$$

$$P(a \leq X < b) = F_X(b^-) - F_X(a^-). \quad (5.5)$$

5.5 foi obtida a partir de 5.1 e 5.2.

$$\begin{aligned} \text{(e)} \quad (-\infty, b) &= (-\infty, b) \cup \{b\} \Rightarrow \\ P_X((-\infty, b)) &= P_X((-\infty, b)) + P(X = b) \Rightarrow \\ P_X((-\infty, b)) &= P_X((-\infty, b]) - P(X = b) \Rightarrow \end{aligned}$$

$$P(-\infty < X < b) = F_X(b^-). \quad (5.6)$$

5.6 foi obtida a partir de 5.1 e 5.2.



### 5.3 Tipos de Variável Aleatória

**Definição 5.3.1:** Existem três tipos de variáveis aleatórias:

- **Discreta.** Uma variável aleatória  $X$  é *discreta* se assume um número enumerável de valores, ou seja, se existe um conjunto enumerável  $\{x_1, x_2, \dots\} \subseteq R$  tal que  $X(w) \in \{x_1, x_2, \dots\}, \forall w \in \Omega$ . A função  $p(x_i)$  definida por  $p(x_i) = P_X(\{x_i\}), i = 1, 2, \dots$  e  $p(x) = 0$  para  $x \notin \{x_1, x_2, \dots\}$ , é chamada de *função probabilidade* de  $X$ .

- **Contínua.** Uma variável aleatória  $X$  é *contínua* se existe uma função  $f_X(x) \geq 0$  tal que

$$F_X(x) = \int_{-\infty}^x f_X(t) dt, \forall x \in R.$$

Neste caso, a função  $f_X$  é chamada de *função densidade de probabilidade* de  $X$ .

- **Singular.** Uma variável aleatória  $X$  é *singular* se  $F_X$  é uma função contínua cujos pontos de crescimento formam um conjunto de comprimento (medida de Lebesgue) nulo.

Pode-se provar que toda função de distribuição de probabilidade acumulada  $F_X$  pode ser decomposta na soma de no máximo três funções de distribuição de probabilidade acumuladas, sendo uma discreta, uma contínua e outra singular. Neste curso, nós estudaremos apenas as variáveis aleatórias discretas.

### 5.4 Variável Aleatória Discreta

Vamos considerar agora o caso das variáveis aleatórias discretas. Nós vimos na seção anterior que se uma variável aleatória é discreta, então nós podemos definir uma função de probabilidade  $p$  de modo que  $p(x_i) = P_X(\{x_i\}), i = 1, 2, \dots$ , onde  $X \subseteq \{x_1, x_2, \dots\}$  e  $p(x) = 0$  para  $x \notin \{x_1, x_2, \dots\}$ . Note que toda função de probabilidade é uma função dos reais  $R$  e assume valores entre 0 e 1, sendo positiva para um número enumerável de pontos e satisfaz a seguinte propriedade  $\sum_i p(x_i) = 1$ .

Por outro lado, dada uma função  $p : R \rightarrow [0, 1]$ , onde  $p$  é positiva para um número enumerável de pontos  $\{x_1, x_2, \dots\}$  e satisfaz  $\sum_i p(x_i) = 1$ , uma função  $P$  definida nos eventos Borelianos de modo que  $P(A) = \sum_{x_i \in A} p(x_i), \forall A \in \mathcal{B}$  é uma medida de probabilidade em  $(R, \mathcal{B})$  (é fácil verificar que  $P$  satisfaz os axiomas de Kolmogorov e portanto é uma medida de probabilidade). Logo, a distribuição de uma variável aleatória discreta  $X$  pode ser determinada tanto pela função de distribuição acumulada  $F_X$  ou pela sua função de probabilidade  $p$ . Em particular, temos que a função de distribuição acumulada de uma variável aleatória discreta  $X$ ,  $F_X(x) = P(X \leq x) = \sum_{i: x_i \leq x} p(x_i)$ , é sempre uma função degrau com saltos nos pontos de  $x_i$  de altura  $p(x_i)$ .

**Exemplo 5.4.1:** Assuma que  $X$  é uma variável aleatória discreta que assume os valores 2, 5, e 7 com probabilidades  $1/2, 1/3$ , e  $1/6$ , então sua função de distribuição acumulada é:

$$F_X(x) = \begin{cases} 0 & \text{se } x < 2, \\ 1/2 & \text{se } 2 \leq x < 5, \\ 5/6 & \text{se } 5 \leq x < 7, \\ 1 & \text{se } x \geq 7. \end{cases}$$

**Exemplo 5.4.2:** A função de distribuição acumulada de uma variável aleatória  $X$  é dada por

$$F(x) = \begin{cases} 0 & \text{se } x < 0, \\ 1/4 & \text{se } 0 \leq x < 1, \\ 3/8 & \text{se } 1 \leq x < 3, \\ 1/2 & \text{se } 3 \leq x < 6, \\ 3/4 & \text{se } 6 \leq x < 10, \\ 1 & \text{se } x \geq 10, \end{cases}$$

- (a) Determine a função probabilidade de massa de  $X$ .
- (b) Determine  $P(1 \leq X < 6 | X < 9)$ .

■

Vamos agora explorar alguns exemplos importantes de variáveis aleatórias discretas.

### 5.4.1 Aleatória.

Dizemos que  $X$  tem uma distribuição *aleatória* com parâmetro  $n$ , onde  $n$  é um número inteiro, se  $X(w) \in \{x_1, x_2, \dots, x_n\}$  e  $p(x_i) = \frac{1}{n}$ , para  $i \in \{1, \dots, n\}$ .

A função de probabilidade aleatória pode ser utilizada para modelar mecanismos de jogos (por exemplo, dados e moedas balanceados, cartas bem embaralhadas). Utilizando a propriedade de aditividade da probabilidade, é fácil ver que para qualquer evento  $A \subseteq \{x_1, x_2, \dots, x_n\}$ , temos que  $P(X \in A) = \frac{|A|}{n}$ .

### 5.4.2 Bernoulli.

Dizemos que  $X$  tem uma distribuição *Bernoulli* com parâmetro  $p$ , onde  $0 \leq p \leq 1$ , se  $X(w) \in \{x_0, x_1\}$  e  $p(x_1) = p = 1 - p(x_0)$ .

A função de probabilidade Bernoulli pode ser utilizada para modelar a probabilidade de sucesso em uma única realização de um experimento. Em geral, qualquer variável aleatória dicotômica, ou seja que assume somente dois valores, pode ser modelada por uma distribuição Bernoulli. Denomina-se de *ensaio de Bernoulli*, qualquer experimento que tem uma resposta dicotômica. Um exemplo clássico de um ensaio Bernoulli é o lançamento de uma moeda não necessariamente balanceada.

### 5.4.3 Binomial.

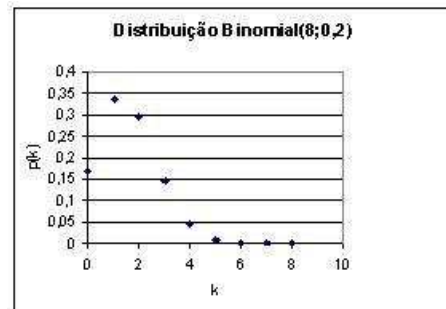
Dizemos que  $X$  tem uma distribuição *Binomial* com parâmetros  $n$  e  $p$ , onde  $n$  é um número inteiro e  $0 \leq p \leq 1$ , se  $X(w) \in \{0, 1, \dots, n\}$  e  $p(k) = \binom{n}{k} p^k (1-p)^{n-k}$ , para  $k \in \{0, 1, \dots, n\}$ .

Note que utilizando o Teorema Binomial, temos que

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

Logo, esta é uma legítima função probabilidade de massa.

Uma distribuição binomial pode ser obtida quando se considera  $n$  repetições independentes de ensaios Bernoulli, e estamos interessados no total de vezes que nesses ensaios obtivemos valor  $x_1$  para a variável. A função de probabilidade binomial pode ser utilizada para modelar a quantidade de erros em um texto de  $n$  símbolos quando os erros entre símbolos são assumidos independentes e a probabilidade de erro em um símbolo do texto é igual a  $p$ . Também pode ser utilizada para modelar o número de caras em  $n$  lançamentos de uma moeda que possui probabilidade  $p$  de cair cara em cada lançamento. Se  $p = 1/2$ , temos um modelo para o número de 1's em uma seqüência binária de comprimento  $n$  escolhida aleatoriamente ou o número de caras em  $n$  lançamentos de uma moeda justa. A Figura 5.4.3 nos mostra a função probabilidade de massa da Binomial(8; 0,2).



Podemos examinar a função probabilidade de massa binomial analiticamente para encontrarmos seu valor mais provável. Note que a razão entre as probabilidades de dois valores consecutivos da binomial

$$\frac{p(k)}{p(k-1)} = \frac{\frac{n!}{(k)!(n-k)!} p^k (1-p)^{n-k}}{\frac{n!}{(k-1)!(n-k+1)!} p^{k-1} (1-p)^{n-k+1}} = \frac{n-k+1}{k} \frac{p}{1-p}$$

é estritamente decrescente em  $k$ . Portanto, se

$$\frac{p(1)}{p(0)} = \frac{np}{1-p} < 1,$$

então as probabilidades são sempre decrescentes em  $k$ , e o valor mais provável é 0. No outro extremo, se

$$\frac{p(n)}{p(n-1)} = \frac{p}{n(1-p)} > 1,$$

então as probabilidades são estritamente crescentes em  $k$ , e o valor mais provável é  $n$ . Se  $\frac{1}{n} < \frac{p}{1-p} < n$ , então a função começa crescendo em  $k$ , enquanto  $\frac{n-k+1}{k} \frac{p}{1-p} > 1$ , e depois decresce em  $k$ . Portanto, se  $\frac{p(k)}{p(k-1)} = \frac{n-k+1}{k} \frac{p}{1-p} = 1$  para algum valor de  $k$ , temos que  $k$  e  $k-1$  são os valores mais prováveis. Caso contrário, o valor mais provável será o maior valor de  $k$  para o qual  $\frac{p(k)}{p(k-1)} = \frac{n-k+1}{k} \frac{p}{1-p} > 1$ , isto é, o valor mais provável será o maior valor de  $k$  tal que  $k < (n+1)p$ . No exemplo da Figura 5.4.3, observe que o valor mais provável é para  $k = 1$ , pois  $(n+1)p = 1,8$ .

**Exemplo 5.4.3:** Uma moeda com probabilidade 0,4 de cair cara é jogada 5 vezes, qual a probabilidade de se obter exatamente 2 coroas?

**Solução:** Seja  $X$  o número de caras obtidos. Como jogamos a moeda 5 vezes, o evento obter exatamente 2 coroas é igual ao evento obter exatamente 3 caras. Portanto,  $P(X = 3) = \binom{5}{3}(0,4)^3(0,6)^2$ .

**Exemplo 5.4.4:** A taxa de sucesso de um bit em uma transmissão digital é 90%. Se 20 bits forem transmitidos, qual a probabilidade de que exatamente 15 deles tenha sido transmitidos com sucesso? Qual a probabilidade de que no máximo 18 deles tenham sido transmitidos com sucesso?

**Exemplo 5.4.5:** Suponha que para uma dada moeda viciada a probabilidade de que ocorram 3 caras seja igual a probabilidade que ocorram 4 caras se esta moeda for jogada 8 vezes de forma independente. Determine a probabilidade de ocorrerem 3 caras em 8 lançamentos independentes desta moeda.

#### 5.4.4 Hipergeométrica.

A distribuição hipergeométrica descreve o número de sucessos em uma seqüência de  $n$  amostras de uma população finita sem reposição.

Por exemplo, considere que tem-se uma carga com  $N$  objetos dos quais  $D$  têm defeito. A distribuição hipergeométrica descreve a probabilidade de que em uma amostra de  $n$  objetos distintos escolhidos da carga aleatoriamente exatamente  $k$  objetos sejam defeituosos.

Em geral, se uma variável aleatória  $X$  segue uma distribuição hipergeométrica com parâmetros  $N$ ,  $D$ , e  $n$ , então a probabilidade de termos exatamente  $k$  sucessos é dada por

$$p(k) = \frac{\binom{D}{k} \binom{N-D}{n-k}}{\binom{N}{n}}.$$

Esta probabilidade é positiva se:  $N - D \geq n - k$ , ou seja  $k \geq \max(0, D + n - N)$ , e  $k \leq \min(n, D)$ .

Esta fórmula pode ser entendida assim: existem  $\binom{N}{n}$  possíveis amostras sem reposição. Existem  $\binom{D}{k}$  maneiras de escolher  $k$  objetos defeituosos e existem  $\binom{N-D}{n-k}$  maneiras de preencher o resto da amostra com objetos sem defeito.

Quando a população é grande quando comparada ao tamanho da amostra (ou seja,  $N$  for muito maior que  $n$ ) a distribuição hipergeométrica é aproximada razoavelmente bem por uma distribuição binomial com parâmetros  $n$  (tamanho da amostra) e  $p = D/N$  (probabilidade de sucesso em um único ensaio).

**Exemplo 5.4.6:** Suponha que uma urna contém 20 bolas brancas e 10 bolas pretas. Se 4 bolas são retiradas da urna. Determine:

- (a) A probabilidade de pelo menos uma bola ser branca, se as bolas são retiradas com reposição.
- (b) A probabilidade de pelo menos uma bola ser branca, se as bolas são retiradas sem reposição.

**Exemplo 5.4.7:** Por engano 3 peças defeituosas foram misturadas com boas formando um lote com 12 peças no total. Escolhendo ao acaso 4 dessas peças, determine a probabilidade de encontrar:

- (a) Pelo menos 2 defeituosas.
- (b) No máximo 1 defeituosa.
- (c) No mínimo 1 boa.

### 5.4.5 Geométrica.

Dizemos que  $X$  tem uma distribuição *Geométrica* com parâmetro  $\beta$ , onde  $0 \leq \beta < 1$ , se  $X(w) \in \{0, 1, \dots\}$  e  $p(k) = (1 - \beta)\beta^k$ , para  $k \in \{0, 1, \dots\}$ .

Utilizando o resultado de uma soma infinita de uma Progressão Geométrica, temos que

$$\sum_{k=0}^{\infty} p(k) = \sum_{k=0}^{\infty} (1 - \beta)\beta^k = (1 - \beta) \sum_{k=0}^{\infty} \beta^k = 1.$$

Logo, esta é uma legítima função probabilidade de massa.

A função de probabilidade Geométrica pode ser utilizada para modelar o tempo de espera medido em unidades de tempo inteira até a chegada do próximo consumidor em uma fila, até a próxima emissão de um fóton, ou até a primeira ocorrência de cara numa seqüência de lançamentos de uma moeda.

**Exemplo 5.4.8:** Suponha que joga-se uma moeda com probabilidade de cara igual a  $0 < p < 1$  independentemente até que uma coroa ocorra. Seja  $X$  o tempo de espera até que coroa apareça nesta seqüência, de modo que se o primeiro lançamento for coroa temos que  $X = 0$ . Qual a probabilidade do evento  $X = k$  para  $k \in \{0, 1, 2, \dots\}$ ? Note que para que  $X = k$  é necessário que os primeiros  $k$  lançamentos sejam caras e o  $(k + 1)$ -ésimo lançamento seja coroa, logo pela independência dos lançamentos, temos que  $P(X = k) = p^k(1 - p)$ . Ou seja  $X$  é uma variável geométrica com parâmetro  $p$ .

**Exemplo 5.4.9:** Suponha que  $X$  tenha uma distribuição geométrica com parâmetro  $\beta$ . Mostre que para quaisquer dois inteiros positivos  $s$  e  $t$ ,

$$P(X > s + t | X > s) = P(X > t).$$

**Solução:** Note que

$$P(X > s + t | X > s) = \frac{P(X > s + t, X > s)}{P(X > s)} = \frac{P(X > s + t)}{P(X > s)}.$$

Mas

$$P(X > s + t) = \sum_{k=s+t+1}^{\infty} (1 - \beta)\beta^{k-1} = \beta^{s+t}.$$

Similarmente, temos que  $P(X > s) = \beta^s$ . Portanto,

$$P(X > s + t | X > s) = \beta^t = P(X > t).$$

Esta propriedade da distribuição geométrica é conhecida como *falta de memória*.

### 5.4.6 Binomial Negativa ou Pascal.

Esta distribuição é uma generalização óbvia da distribuição geométrica. Suponha que ao invés de estarmos interessados no tempo de espera até a primeira ocorrência de um evento, estejamos interessados em calcular o tempo de espera até a  $r$ -ésima ocorrência de um evento. Seja  $Y$  o tempo de espera necessário a fim de que um evento  $A$  possa ocorrer exatamente  $r$  vezes. Temos que  $Y = k$  se, e somente se,  $A$  ocorrer na  $(k + 1)$ -ésima repetição e  $A$  tiver ocorrido  $r - 1$  vezes nas  $k$  repetições anteriores. Assumindo independência entre os experimentos, esta probabilidade é igual  $p \binom{k}{r-1} p^{r-1} (1-p)^{k-r+1}$ . Portanto,

$$P(Y = k) = \binom{k}{r-1} p^r (1-p)^{k-r+1}, \text{ onde } k \geq r-1.$$

Note que se  $r = 1$ , temos que  $Y$  tem uma distribuição geométrica com parâmetro  $\beta = 1 - p$ . No caso geral, dizemos que  $Y$  tem uma distribuição *Binomial Negativa ou Pascal*.

#### Relação entre as Distribuições Binomial e Binomial Negativa.

Suponhamos que  $X$  tenha distribuição binomial com parâmetros  $n$  e  $p$ , ou seja,  $X$  é igual ao número de sucessos em  $n$  ensaios repetidos de Bernoulli com probabilidade de sucesso  $p$ . Suponhamos que  $Y$  tenha uma distribuição Binomial Negativa com parâmetros  $r$  e  $p$ , ou seja,  $Y + 1$  é o número de ensaios de Bernoulli necessários para se obter  $r$  sucessos com probabilidade de sucesso  $p$ . Então, temos que  $\{X \geq r\} = Y + 1 \leq n$ , ou seja, o número de sucessos em  $n$  ensaios é maior ou igual a  $r$  se, e somente se, o tempo de espera para o  $r$ -ésimo sucesso for menor ou igual a  $n - 1$ . Portanto,

$$P(X \geq r) = P(Y \leq n - 1).$$

Observe que estas duas distribuições tratam de ensaios de Bernoulli repetidos. A distribuição binomial surge quando lidamos com um número fixo de ensaios e estamos interessados no número de sucessos que venham a ocorrer. A distribuição binomial negativa é encontrada quando fixamos o número de sucessos e então registramos o tempo de espera necessário.

### 5.4.7 Zeta ou Zipf.

Dizemos que  $X$  tem uma distribuição Zeta ou Zipf com parâmetro  $\alpha$ , onde  $\alpha > 1$ , se  $X(w) \in \{1, 2, \dots\}$  e

$$p(k) = \frac{k^{-\alpha}}{\zeta(\alpha)}, k = 1, 2, \dots,$$

onde  $\zeta(\alpha) = \sum_{k=1}^{\infty} k^{-\alpha}$  é conhecida como a função Zeta de Riemann.

A função de probabilidade Zeta ou Zipf é um exemplo de uma distribuição de cauda pesada cuja importância cresceu bastante desde meados dos anos 1990. As aplicações desta função de probabilidade incluem: número de consumidores afetados por um blackout, tamanhos de arquivos solicitados em transferência via Web e atraso de pacotes na internet.

**Exemplo 5.4.10:** Os tamanhos de arquivos armazenados em um grande sistema de arquivos Unix segue uma distribuição Zeta com parâmetro  $\alpha$  quando estes tamanhos são medidos em unidades de kilobytes.

- Se os tamanhos dos arquivos de 1KB são 10.000 vezes mais prováveis que tamanhos de arquivos de 1MB, então qual o valor do parâmetro  $\alpha$ ?
- Quanto mais provável são tamanhos de arquivos de 1MB em comparação com tamanhos de arquivos de 1GB?

### 5.4.8 Poisson.

Dizemos que  $X$  tem uma distribuição *Poisson* com parâmetro  $\lambda$ , onde  $\lambda \geq 0$ , se  $X(w) \in \{0, 1, \dots\}$  e  $p(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ , para  $k \in \{0, 1, \dots\}$ .

Usando o resultado da expansão em série de Taylor da função exponencial, temos que para todo  $x$  real,

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Utilizando este fato, temos que

$$\sum_{k=0}^{\infty} p(k) = \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

Logo, esta é uma legítima função probabilidade de massa.

A função de probabilidade Poisson é utilizada para modelar a contagem do número de ocorrências de eventos aleatórios em um certo tempo  $T$ : número de fótons emitidos por uma fonte de luz de intensidade  $I$  fótons/seg em  $T$  segundos ( $\lambda = IT$ ), número de clientes chegando em uma fila no tempo  $T$  ( $\lambda = CT$ ), número de ocorrências de eventos raros no tempo  $T$  ( $\lambda = CT$ ).

**Exemplo 5.4.11:** Se a probabilidade de 0 fótons serem emitidos no tempo  $T$  é igual a 0,1, então qual a probabilidade de pelo menos 2 fótons serem emitidos no tempo  $T$ ?

Podemos analisar o valor mais provável de uma distribuição de Poisson, através da razão de dois valores sucessivos da função probabilidade de massa:

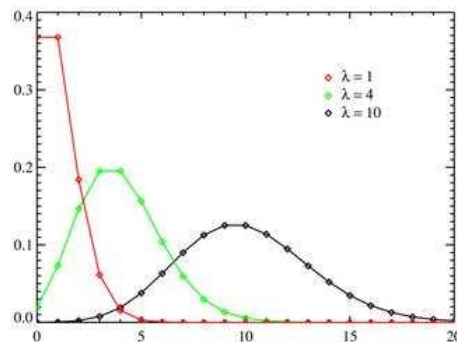
$$\frac{p_{k+1}}{p_k} = \frac{\lambda}{k+1}.$$

Note que esta razão é estritamente decrescente em  $k$ . Logo,  $\{p_k\}$  é sempre decrescente se  $\lambda < 1$ , decresce após  $p_0 = p_1$  se  $\lambda = 1$ , e cresce inicialmente se  $\lambda > 1$  e eventualmente decresce qualquer que seja o valor de  $\lambda$ . Formalmente, um valor mais provável de uma distribuição de Poisson é definido como  $k^*$  se  $p_{k^*+1} \leq p_{k^*}$  e  $p_{k^*-1} \leq p_{k^*}$ . (Note que podem existir valores adjacentes que possuam o mesmo valor.) Mas esta condição é equivalente a,

$$k^* \leq \lambda \leq k^* + 1, \text{ ou}$$

$$\lambda - 1 \leq k^* \leq \lambda.$$

Note que se tomarmos  $k^*$  como sendo o maior inteiro menor ou igual a  $\lambda$  esta restrição é satisfeita, e portanto este é um valor mais provável desta distribuição. A Figura 5.4.8 nos mostra a função probabilidade de massa da Poisson para 3 valores de parâmetros 1, 4, e 10.



**Exemplo 5.4.12:** Suponha que o número de clientes que chegam em um banco segue uma distribuição de Poisson. Se a probabilidade de chegarem 3 clientes for o triplo da de chegarem 4 clientes em um dado período de 10 minutos. Determine qual o número mais provável de clientes que chegam em um período de 1 hora neste banco.

## 5.5 Poisson como um Limite de Eventos Raros de Binomial

Suponhamos que chamadas telefônicas cheguem em uma grande central, e que em um período particular de três horas (180 minutos), um total de 270 chamadas tenham sido recebidas, ou seja, 1,5 chamadas por minuto. Suponhamos que queiramos calcular a probabilidade de serem recebidas  $k$  chamadas durante os próximos três minutos.

Ao considerar o fenômeno da chegada de chamadas, poderemos chegar à conclusão de que, a qualquer instante, uma chamada telefônica é tão provável de ocorrer como em qualquer



outro instante. Como em qualquer intervalo de tempo, temos um número infinito de pontos, vamos fazer uma série de aproximações para este cálculo.

Para começar, pode-se dividir o intervalo de 3 minutos em nove intervalos de 20 segundos cada um. Poderemos então tratar cada um desses nove intervalos como um ensaio de Bernoulli, durante o qual observaremos uma chamada (sucesso) ou nenhuma chamada (falha), com probabilidade de sucesso igual a  $p = 1,5 \times \frac{20}{60} = 0,5$ . Desse modo, poderemos ser tentados a afirmar que a probabilidade de 2 chamadas é igual a  $\binom{9}{2}(0,5)^9 = \frac{9}{128}$ . Porém, este cálculo ignora a possibilidade de que mais de uma chamada possa ocorrer em um único intervalo. Então, queremos aumentar o número  $n$  de subintervalos de tempo de modo que cada subintervalo corresponde a  $\frac{180}{n}$  segundos e então a probabilidade de ocorrência de uma chamada em um subintervalo é igual a  $p = 1,5 \times \frac{180}{60n}$ . Desta maneira temos que  $np = 4,5$  permanece constante ao crescermos o número de subintervalos. Utilizando novamente o modelo binomial, temos que a probabilidade de ocorrerem  $k$  chamadas é dada por:  $\binom{n}{k}(\frac{4,5}{n})^k(1 - \frac{4,5}{n})^{n-k}$ . Queremos saber então o que acontece com esta probabilidade quando  $n \rightarrow \infty$ . A resposta como veremos a seguir é que esta distribuição tende a distribuição de Poisson e este resultado é conhecido como *limite de eventos raros*.

Consideremos a expressão geral da probabilidade binomial,

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k}.$$

Como queremos estudar o caso em que  $np$  é constante, façamos  $np = \alpha$ , ou seja,  $p = \alpha/n$  e  $1-p = \frac{n-\alpha}{n}$ . Então,

$$\begin{aligned} p(k) &= \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\alpha}{n}\right)^k \left(\frac{n-\alpha}{n}\right)^{n-k} \\ &= \frac{\alpha^k}{k!} \left[\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right] \left[1-\frac{\alpha}{n}\right]^{n-k} \end{aligned}$$

Fazendo  $n \rightarrow \infty$ , temos que os termos da forma  $(1 - \frac{j}{n})$ , para  $1 \leq j \leq k-1$ , tendem para 1 e como existe um número fixo  $k-1$  deles, o seu produto também tende a 1. O mesmo ocorre com  $(1 - \frac{\alpha}{n})^{n-k}$ . Finalmente, por definição do número  $e$ , temos que  $(1 - \frac{\alpha}{n})^n \rightarrow e^{-\alpha}$  quando  $n \rightarrow \infty$ . Portanto,

$$\lim_n p(k) = e^{-\alpha} \frac{\alpha^k}{k!},$$

ou seja obtemos a expressão de Poisson.

Então, provamos o seguinte teorema:

**Teorema 5.5.1:** *Se  $\lim_{n \rightarrow \infty} np_n = \alpha > 0$ , então*

$$\lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\alpha} \frac{\alpha^k}{k!}.$$

Este resultado é importante tanto para motivar a forma da distribuição de Poisson, como também fornece um método para aproximar o cálculo de uma probabilidade que pode ser

feito a partir da distribuição binomial pela distribuição Poisson, ou seja, em geral, se  $np = \alpha$  e  $n$  for um número grande, então é válida a seguinte aproximação:

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{e^{-\alpha} \alpha^k}{k!}.$$

**Exemplo 5.5.2:** Ao formar números binários com  $n$  dígitos, a probabilidade de que um dígito incorreto possa aparecer é 0,002. Se os erros forem independentes, qual é a probabilidade de encontrar  $k$  dígitos incorretos em um número binário de 25 dígitos? Se um computador forma  $10^6$  desses números de 25 dígitos por segundo, qual é a probabilidade de que pelo menos um número incorreto seja formado durante qualquer período de 1 segundo?

**Solução:** A probabilidade de que  $k$  dígitos sejam incorretos em um número binários de 25 dígitos é igual a  $\binom{25}{k} (0,002)^k (0,998)^{25-k}$ . Em particular, a probabilidade de que pelo menos um dígito seja incorreto é igual a  $1 - (0,998)^{25} \approx 0,049$ . Se tivéssemos usado a aproximação pela Poisson então teríamos uma Poisson com parâmetro  $25 \times 0,002 = 0,05$ , logo a probabilidade de pelos menos um dígito incorreto neste número de 25 dígitos seria  $1 - e^{-0,05} \approx 0,049$ .

A probabilidade de que pelo menos um número incorreto seja formado durante o período de 1 segundo é igual a  $1 - (0,049)^{10^6} \approx 1 - e^{-49000} \approx 1$ .

# Capítulo 6

## Esperança e Momentos de Variáveis Aleatórias Discretas

### 6.1 O Conceito de Esperança

O conceito de Esperança ou Valor Esperado de uma variável aleatória  $X$ , ou a “média” é tão antigo quanto o próprio conceito de probabilidade. Na verdade, é até possível definir probabilidade em termos de esperança, mas esta não é uma maneira comum de se apresentar a teoria. Existem quatro tipos de interpretações da Esperança:

1. Parâmetro  $m$  de uma medida de probabilidade, função de distribuição, ou função probabilidade de massa, também conhecido como média.
2. Um operador linear em um conjunto de variáveis aleatórias que retorna um valor típico da variável aleatória interpretado como uma medida de localização da variável aleatória.
3. média do resultado de repetidos experimentos independentes no longo prazo.
4. preço justo de um jogo com pagamentos descritos por  $X$ .

#### 6.1.1 Definição da Esperança

Vamos motivar a definição de esperança considerando o cálculo do resultado médio de 1000 lançamentos de um dado. Uma maneira de calcular este resultado médio seria somar todos os resultados e dividir por 1000. Uma maneira alternativa seria calcular a fração  $p(k)$  de todos os lançamentos que tiveram resultado igual a  $k$  e calcular o resultado médio através da soma ponderada:

$$1p(1) + 2p(2) + 3p(3) + 4p(4) + 5p(5) + 6p(6).$$

Quando o número de lançamentos se torna grande as frações de ocorrência dos resultados tendem a probabilidade de cada resultado. Portanto, em geral definimos a esperança de uma variável discreta como uma soma ponderada onde as probabilidades são os pesos de ponderação.

**Definição 6.1.1:** Se  $X$  é uma variável aleatória discreta assumindo valores  $\{x_1, x_2, x_3, \dots\}$  com probabilidade  $\{p_1, p_2, p_3, \dots\}$ , respectivamente, então sua esperança é dada pela fórmula

$$EX = \sum_{i: x_i < 0} x_i p_i + \sum_{i: x_i \geq 0} x_i p_i,$$

desde que pelo menos um dos somatórios seja finito. Em caso os dois somatórios não sejam finitos, a esperança não existe.

**Exemplo 6.1.2:** Considere uma variável aleatória  $X$  tal que:  $P(X = -1) = 0.25$ ,  $P(X = 0) = 0.5$  e  $P(X = 2) = 0.25$ . Então,

$$EX = -1(0.25) + 0(0.5) + 2(0.25) = 0.25.$$

**Exemplo 6.1.3:** Considere uma variável aleatória  $X$  tal que:  $P(X = -a) = P(X = a) = 1/2$ . Então,

$$EX = -a(0.5) + a(0.5) = 0.$$

Note então que muitas variáveis aleatórias diferentes podem ter o mesmo valor esperado ou esperança. (É só variar o valor de  $a$  no exemplo anterior.)

**Exemplo 6.1.4: Aleatória.** Se  $X \in \{1, 2, \dots, n\}$  for uma variável aleatória com distribuição de probabilidade aleatória com parâmetro  $n$ , temos que sua esperança é dada por:

$$EX = \sum_{k=1}^n k p(k) = \sum_{k=1}^n k \frac{1}{n} = \frac{1}{n} \sum_{k=1}^n k = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}.$$

Onde utilizamos a fórmula da soma dos primeiros  $n$  termos de uma progressão aritmética.

**Exemplo 6.1.5: Bernoulli.** Se  $X \in \{0, 1\}$  for uma variável aleatória com distribuição de probabilidade Bernoulli com parâmetro  $p$ , temos que sua esperança é dada por:

$$EX = 0(1-p) + 1(p) = p.$$

**Exemplo 6.1.6: Binomial.** Se  $X$  for uma variável aleatória com distribuição de probabilidade Binomial com parâmetros  $n$  e  $p$ , temos que sua esperança é dada por:

$$\begin{aligned} EX &= \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n n \frac{(n-1)!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} = np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} = np. \end{aligned}$$

Onde utilizamos o Teorema Binomial na última igualdade.

**Exemplo 6.1.7: Geométrica.** Se  $X$  for uma variável aleatória com distribuição de probabilidade Geométrica com parâmetro  $\beta$ , temos que sua esperança é dada por:

$$\begin{aligned} EX &= \sum_{k=0}^{\infty} k(1-\beta)\beta^k = \sum_{k=1}^{\infty} k(1-\beta)\beta^k = \sum_{k=1}^{\infty} \sum_{j=1}^k (1-\beta)\beta^k \\ &= (1-\beta) \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} \beta^k = \sum_{j=1}^{\infty} \beta^j = \frac{\beta}{1-\beta} \end{aligned}$$

Onde utilizamos a fórmula da soma infinita de uma progressão geométrica com razão  $\beta$ .

**Exemplo 6.1.8: Binomial Negativa.** Se  $X$  for uma variável aleatória com distribuição de probabilidade Binomial Negativa com parâmetros  $r$  e  $p$ , temos que sua esperança é dada por:

$$\begin{aligned} EX &= \sum_{k=r-1}^{\infty} k \binom{k}{r-1} p^r (1-p)^{k-r+1} = \left( \sum_{k=r-1}^{\infty} (k+1) \binom{k}{r-1} p^r (1-p)^{k-r+1} \right) - 1 \\ &= \left( \sum_{k=r-1}^{\infty} \frac{(k+1)k!}{(r-1)!(k-r+1)!} p^r (1-p)^{k-r+1} \right) - 1 \\ &= \frac{r}{p} \left( \sum_{k=r-1}^{\infty} \frac{(k+1)!}{r!(k+1-r)!} p^{r+1} (1-p)^{k+1-r} \right) - 1 \end{aligned}$$

Substituindo  $j = k + 1$  e  $s = r + 1$  no somatório, temos

$$EX = \frac{r}{p} \left( \sum_{j=s-1}^{\infty} \frac{(j)!}{(s-1)!(j-s+1)!} p^s (1-p)^{j-s+1} \right) - 1 = \frac{r}{p} - 1$$

Onde utilizamos o fato que o somatório é igual soma da função probabilidade de massa de uma variável aleatória Binomial Negativa para todos os valores que tem probabilidade positiva, e portanto, é igual a 1.

**Exemplo 6.1.9: Poisson.** Se  $X$  for uma variável aleatória com distribuição de probabilidade Poisson com parâmetros  $\lambda$ , temos que sua esperança é dada por:

$$EX = \sum_{k=0}^{\infty} k \frac{e^{-\lambda} \lambda^k}{k!} = \sum_{k=1}^{\infty} k \frac{e^{-\lambda} \lambda^k}{k!} = \lambda \sum_{k=1}^{\infty} \frac{e^{-\lambda} \lambda^{k-1}}{(k-1)!} = \lambda.$$

**Exemplo 6.1.10: Zeta.** Se  $X$  for uma variável aleatória com distribuição de probabilidade Zeta com parâmetro  $\alpha > 2$ , temos que sua esperança é dada por:

$$EX = \sum_{k=1}^{\infty} k \frac{k^{-\alpha}}{\zeta(\alpha)} = \frac{1}{\zeta(\alpha)} \sum_{k=1}^{\infty} k^{-(\alpha-1)} = \frac{\zeta(\alpha-1)}{\zeta(\alpha)},$$

onde  $\zeta(\alpha) = \sum_{k=1}^{\infty} k^{-\alpha}$ .

**Exemplo 6.1.11: Hipergeométrica.** Se  $X$  for uma variável aleatória com distribuição de probabilidade Hipergeométrica com parâmetro  $N, D, n$ , temos que sua esperança é dada por:

$$\begin{aligned} EX &= \sum_{k=0}^n k \frac{\binom{D}{k} \binom{N-D}{n-k}}{\binom{N}{n}} = \sum_{k=1}^n \frac{D!(N-D)!(N-n)!n!}{k!(D-k)!(n-k)!(N-D-n+k)!N!} \\ &= \frac{nD}{N} \sum_{k=1}^n \frac{(D-1)!(N-D)!(N-n)!(n-1)!}{(k-1)!(D-k)!(n-k)!(N-D-n+k)!(N-1)!} = \frac{nD}{N} \sum_{k=1}^n \frac{\binom{D-1}{k-1} \binom{N-D}{n-k}}{\binom{N-1}{n-1}} \end{aligned}$$

Substituindo no somatório  $D^* = D - 1, k^* = k - 1, n^* = n - 1$  e  $N^* = N - 1$ , temos

$$EX = \frac{nD}{N} \sum_{k^*=0}^{n^*} \frac{\binom{D^*}{k^*} \binom{N^*-D^*}{n^*-k^*}}{\binom{N^*}{n^*}} = \frac{nD}{N}.$$

Onde utilizamos o fato que o somatório é igual soma da função probabilidade de massa de uma variável aleatória Hipergeométrica para todos os valores que tem probabilidade positiva, e portanto, é igual a 1.

## 6.2 Funções de Variáveis Aleatórias

Muitas vezes sabemos a distribuição de probabilidade que descreve o comportamento de uma variável aleatória  $X$  definida no espaço mensurável  $(\Omega, \mathcal{A})$ , mas estamos interessados na descrição de uma função  $Y = H(X)$ . Por exemplo,  $X$  pode ser uma mensagem enviada em um canal de telecomunicações e  $Y$  ser a mensagem recebida. Nosso problema é determinar  $P(Y \in A)$ , onde  $A$  é um evento Boreliano, dado  $P_X$ . Para determinarmos esta probabilidade, estaremos interessados na imagem inversa a função  $H$ , ou seja, a probabilidade do evento  $\{Y \in A\}$  será por definição igual a probabilidade do evento  $\{X \in H^{-1}(A)\}$ , onde  $H^{-1}(A) = \{x \in \mathbb{R} : H(x) \in A\}$ . Para que esta probabilidade esteja bem definida, precisamos restringir  $H$  tal que  $H^{-1}(A)$  seja um evento boreliano para todo  $A$  boreliano, caso contrário não poderemos determinar  $P(\{X \in H^{-1}(A)\})$ . Note que  $Y$  também pode ser

vista como uma função do espaço amostral  $\Omega$ ,  $Y(\omega) = H(X(\omega))$  para todo  $\omega \in \Omega$ . Visto dessa maneira  $Y$  é uma variável aleatória definida em  $(\Omega, \mathcal{A})$ , pois para todo boreliano  $A$   $Y^{-1}(A) = X^{-1}(H^{-1}(A))$  e como por suposição  $H^{-1}(A)$  é boreliano e  $X$  é uma variável aleatória, temos que  $X^{-1}(H^{-1}(A)) \in \mathcal{A}$  e portanto satisfaz a definição de uma variável aleatória. Trataremos esse problema apenas no caso de variáveis aleatórias discretas. Neste caso para qualquer função  $H$ , temos que  $Y = H(X)$  é uma variável aleatória discreta.

Suponha que  $X$  assuma os valores  $x_1, x_2, \dots$  e seja  $H$  uma função real tal que  $Y = H(X)$  assumam os valores  $y_1, y_2, \dots$ . Vamos agrupar os valores que  $X$  assume de acordo com os valores de suas imagens quando se aplica a função  $H$ , ou seja, denotemos por  $x_{i1}, x_{i2}, x_{i3}, \dots$  os valores de  $X$  tal que  $H(x_{ij}) = y_i$  para todo  $j$ . Então, temos que

$$P(Y = y_i) = P(X \in \{x_{i1}, x_{i2}, x_{i3}, \dots\}) = \sum_{j=1}^{\infty} P(X = x_{ij}) = \sum_{j=1}^{\infty} p_X(x_{ij}),$$

ou seja, para calcular a probabilidade do evento  $\{Y = y_i\}$ , acha-se o evento equivalente em termos de  $X$ , isto é, todos os valores  $x_{ij}$  de  $X$  tal que  $H(x_{ij}) = y_i$  e somam-se as probabilidades de  $X$  assumir cada um desses valores.

**Exemplo 6.2.1:** Admita-se que  $X$  tenha os valores possíveis  $1, 2, 3, \dots$  e suponha que  $P(X = n) = (1/2)^n$ . Seja  $Y = 1$  se  $X$  for par e  $Y = -1$  se  $X$  for ímpar. Então, temos que

$$P(Y = 1) = \sum_{n=1}^{\infty} (1/2)^{2n} = \sum_{n=1}^{\infty} (1/4)^n = \frac{1/4}{1 - 1/4} = 1/3.$$

Conseqüentemente,

$$P(Y = -1) = 1 - P(Y = 1) = 2/3.$$

### 6.3 Esperança de Funções de Variáveis Aleatórias Discretas.

Como vimos anteriormente, se  $X$  for uma variável aleatória discreta e se  $Y = H(X)$ , então  $Y$  também será uma variável aleatória discreta. Conseqüentemente, pode-se calcular  $EY$ . Existem duas maneiras de calcular  $EY$  que são equivalentes.

**Definição 6.3.1:** Seja  $X$  uma variável aleatória discreta e seja  $Y = H(X)$ . Se  $Y$  assumir os seguintes valores  $y_1, y_2, \dots$  e se  $p(y_i) = P(Y = y_i)$ , definimos:

$$EY = \sum_{i=1}^{\infty} y_i p(y_i).$$

Conforme vimos na seção anterior podemos determinar as probabilidades  $p(y_i)$  dado que sabemos a distribuição de  $X$ . No entanto, podemos encontrar  $EY$  sem preliminarmente encontrarmos a distribuição de probabilidade de  $Y$ , partindo-se apenas do conhecimento da distribuição de probabilidade de  $X$ , conforme mostra o seguinte teorema.

**Teorema 6.3.2:** *Seja  $X$  uma variável aleatória discreta assumindo os valores  $x_1, x_2, \dots$  e seja  $Y = H(X)$ . Se  $p(x_i) = P(X = x_i)$ , temos*

$$EY = E(H(X)) = \sum_{i=1}^{\infty} H(x_i)p(x_i).$$

**Prova:** Vamos re-ordenar o somatório  $\sum_{i=1}^{\infty} H(x_i)p(x_i)$ , agrupando os termos onde  $x_i$  tem a mesma imagem de acordo com a função  $H$ , ou seja, sejam  $x_{i1}, x_{i2}, \dots$ , todos os valores  $x_i$  tal que  $H(x_{ij}) = y_i$  para  $j \geq 1$ , onde  $y_1, y_2, \dots$  são os possíveis valores de  $Y$ . Desse modo podemos reescrever

$$\sum_{i=1}^{\infty} H(x_i)p(x_i) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} H(x_{ij})p(x_{ij}) = \sum_{i=1}^{\infty} y_i \sum_{j=1}^{\infty} p(x_{ij}) = \sum_{i=1}^{\infty} y_i p(y_i) = EY.$$

■

**Exemplo 6.3.3:** Suponha que  $X$  é uma variável aleatória Poisson com parâmetro  $\lambda$ . Seja  $Y = X^2$ , vamos calcular  $EY$ . Utilizando o Teorema 6.3.2, temos

$$\begin{aligned} EY &= \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} + \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} \\ &= \lambda^2 \sum_{k=2}^{\infty} e^{-\lambda} \frac{\lambda^{k-2}}{(k-2)!} + \lambda = \lambda^2 + \lambda. \end{aligned}$$

## 6.4 Propriedades da Esperança

As seguintes propriedades são aplicações imediatas da definição de esperança:

1.  $P(X = c) = 1 \Rightarrow EX = c$ .
2.  $P(X \geq 0) = 1 \Rightarrow EX \geq 0$ .
3.  $E(aX) = aEX$ , onde  $a$  um número real qualquer. Esta propriedade segue facilmente da expressão da esperança de uma função de variável aleatória.



4.  $E(X + Y) = EX + EY$ . A prova desta propriedade envolve variáveis aleatórias bidimensionais, que está fora do escopo deste curso.
5.  $P(X \geq Y) = 1 \Rightarrow EX \geq EY$ . Propriedade 5 segue das propriedades 2, 3, e 4, pois

$$P(X \geq Y) = P(X - Y \geq 0),$$

o que, pela propriedade 2, implica que  $E(X - Y) \geq 0$ . Pela propriedade 4, temos que  $E(X - Y) = EX + E(-Y)$ . Finalmente, pela propriedade 3, temos que  $E(X - Y) = EX - EY$ , ou seja podemos concluir que  $EX - EY \geq 0$ .

6. Se  $X$  tem uma distribuição simétrica em torno de  $a$ , ou seja,  $P(X - a \geq x) = P(X - a \leq -x)$ , e se a esperança de  $X$  tiver bem definida, então  $EX = a$ . Para provar esta expressão, primeiro note que se  $X$  é simétrica em relação a  $a$  então  $Y = X - a$  é simétrica em relação a zero. Se provarmos que  $EY = 0$ , então segue da linearidade da esperança que  $EX = a$ . No caso discreto, como  $Y$  é simétrica em torno de 0, temos que  $p(x_i) = p(-x_i)$  para todo  $x_i$ , portanto segue que  $EY = \sum_i x_i p(x_i) = 0$ .

Pode-se definir outras medidas de posição de uma variável aleatória, tais como: *mediana* e *moda*. A *mediana de uma v.a.*  $X$  é qualquer número  $m$  tal que  $P(X \geq m) \geq 0,5$  e  $P(X \leq m) \geq 0,5$ . Por exemplo, se  $X$  assume os valores  $-1, 0, 1$  com probabilidades  $1/4, 1/4, 1/2$ , respectivamente, então qualquer número no intervalo fechado de 0 a 1. A *moda* de uma variável aleatória discreta é o seu valor mais provável, não necessariamente único.

Em distribuições unimodais simétricas, isto é, distribuições tal que existe um número  $m$  tal que  $P(X - m \geq x) = P(X - m \leq -x)$  para todo  $x \in \mathbb{R}$ , a esperança (se bem definida), mediana, e moda coincidem e são iguais a  $m$ .

## 6.5 Momentos

Momentos dão informações parciais sobre a medida de probabilidade  $P$ , a função de distribuição acumulada, ou a função probabilidade de massa de uma variável aleatória discreta  $X$ . Momentos de  $X$  são esperanças de potências de  $X$ .

**Definição 6.5.1:** Para qualquer inteiro não-negativo  $n$ , o  $n$ -ésimo momento da variável aleatória  $X$  é  $EX^n$ , se esta esperança existe.

Na seção anterior, vimos que o segundo momento de uma variável aleatória Poisson com parâmetro  $\lambda$  é dado por:  $\lambda^2 + \lambda$ . Vamos agora calcular o segundo momento de uma variável

aleatória  $X$  Binomial com parâmetros  $n$  e  $p$ :

$$\begin{aligned} EX^2 &= \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \\ &= \sum_{k=1}^n k(k-1) \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} + \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= n(n-1)p^2 \sum_{k=2}^n \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} (1-p)^{n-k} + np \\ &= n(n-1)p^2 \sum_{j=0}^m \frac{(m)!}{(j)!(m-j)!} p^j (1-p)^{m-j} + np = n(n-1)p^2 + np. \end{aligned}$$

**Teorema 6.5.2:** *Se  $j < k$ , então*

$$|EX^k| < \infty \Rightarrow |EX^j| < \infty$$

**Prova:** Vamos provar apenas para o caso em que  $X$  é uma variável aleatória discreta. Por hipótese, temos que existem constantes finitas  $c_1$  e  $c_2$  tais que:

$$c_1 > \sum_{i:x_i \geq 0} x_i^k p(x_i), \text{ e } c_2 > \left| \sum_{i:x_i < 0} x_i^k p(x_i) \right|.$$

Então, para  $j < k$

$$\begin{aligned} c_1 &> \sum_{i:x_i \geq 0} x_i^k p(x_i) \geq \sum_{i:x_i \geq 1} x_i^k p(x_i) \geq \sum_{i:x_i \geq 1} x_i^j p(x_i) \\ &\geq \sum_{i:x_i \geq 0} x_i^j p(x_i) - \sum_{i:0 \leq x_i < 1} x_i^j p(x_i) \geq \sum_{i:x_i \geq 0} x_i^j p(x_i) - \sum_{i:0 \leq x_i < 1} p(x_i) \geq \sum_{i:x_i \geq 0} x_i^j p(x_i) - 1 \end{aligned}$$

Portanto,

$$0 \leq \sum_{i:x_i \geq 0} x_i^j p(x_i) < c_1 + 1 < \infty.$$

Um cálculo similar prova que  $\sum_{i:x_i < 0} x_i^j p(x_i)$  é finito. ■

Portanto, temos que *momentos de ordem superiores finitos implicam momentos de ordem inferiores finitos.*

### 6.5.1 Momentos Centrais

**Definição 6.5.3:** Se  $X$  é uma variável aleatória seu  $n$ -ésimo momento central é:  $E(X - EX)^n$ , se esta esperança existir.

Note que o primeiro momento central é zero, pois  $E(X - EX) = EX - EEX = EX - EX = 0$ . O segundo momento central é conhecido como *variância* e denota-se por  $VarX$ . A variância pode ser também calculada por:

$$\begin{aligned} VarX &= E(X - EX)^2 = E(X^2 - 2XEX + (EX)^2) = EX^2 - 2E(XEX) + E((EX)^2) \\ &= EX^2 - 2(EX)^2 + (EX)^2 = EX^2 - (EX)^2. \end{aligned} \quad (6.1)$$

Do Teorema Binomial e da linearidade da esperança, temos

$$E(X - EX)^n = \sum_{k=0}^n \binom{n}{k} (-EX)^{n-k} EX^k$$

e

$$EX^n = E(X - EX + EX)^n = \sum_{k=0}^n \binom{n}{k} (EX)^{n-k} E(X - EX)^k.$$

Como um corolário, temos que o  $n$ -ésimo momento central existe se, e somente se, o  $n$ -ésimo momento existe.

**Exemplo 6.5.4:** Considere uma variável aleatória  $X$  tal que

$$P(X = m - a) = P(X = m + a) = \frac{1}{2} \Rightarrow EX^k = \frac{1}{2}[(m - a)^k + (m + a)^k].$$

$$EX = m, EX^2 = \frac{1}{2}[2m^2 + 2a^2] = m^2 + a^2, VarX = a^2.$$

Este exemplo, mostra que podemos encontrar uma variável aleatória bem simples possuindo qualquer esperança e variância predeterminadas.

**Exemplo 6.5.5: (Aleatória ou Uniforme Discreta.)** Se  $X$  tem uma distribuição uniforme discreta assumindo os valores  $\{x_1, x_2, \dots, x_n\}$  com mesma probabilidade, então:

$$VarX = \frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{1}{n^2} \left( \sum_{i=1}^n x_i \right)^2.$$

**Exemplo 6.5.6: (Binomial.)** Já demonstramos que se  $X$  tem uma distribuição binomial, então  $EX = np$  e  $E(X^2) = n(n - 1)p^2 + np$ . Portanto,  $VarX = n(n - 1)p^2 + np - n^2p^2 = np(1 - p)$ .

O *desvio-padrão*  $\sigma$  de uma variável aleatória  $X$  é definido como a raiz quadrada da variância,  $\sigma(X) = \sqrt{VarX}$ .

### Propriedades da Variância

As seguintes propriedades da variância são conseqüências imediatas de sua definição.

1.  $Var X \geq 0$ .

2. Se  $X = c$ ,  $Var(X) = 0$ .

**Prova:** Temos que  $EX = c$ , logo  $Var(X) = E(X - c)^2 = E(0) = 0$ . ■

3.  $Var(X + a) = Var X$ , onde  $a$  é uma constante real.

**Prova:**

$$\begin{aligned} Var(X + a) &= E(X + a)^2 - (E(X + a))^2 \\ &= EX^2 + 2aEX + a^2 - (EX)^2 - 2aEX - a^2 = EX^2 - (EX)^2 = Var X. \end{aligned}$$

■

4.  $Var(aX) = a^2 Var X$

**Prova:**

$$Var(aX) = E(aX)^2 - (E(aX))^2 = a^2 EX^2 - a^2 (EX)^2 = a^2 Var X.$$

■

A tabela a seguir contém a esperança e a variância das distribuições de probabilidade discretas mais comuns.

Distribuição	Esperança	Variância
Bernoulli( $p$ )	$p$	$p(1-p)$
Binomial( $n, p$ )	$np$	$np(1 - p)$
Geométrica( $p$ )	$\frac{(1-p)}{p}$	$\frac{(1-p)}{p^2}$
Poisson( $\lambda$ )	$\lambda$	$\lambda$
Hipergeométrica( $N, D, n$ )	$\frac{nD}{N}$	$\frac{nD}{N} \frac{(N-D)(N-n)}{N(N-1)}$

# Referências Bibliográficas

1. Meyer, P. (1983), "Probabilidade - Aplicações à Estatística", 2a. edição, Livros Técnicos e Científicos Editora, Rio de Janeiro.
2. Davenport Jr., W. (1987), "Probability and Random Processes - an introduction for applied scientists and engineers", McGraw-Hill Book Company Inc.
3. Fine, T. (2006), "Probability and Probabilistic Reasoning for Electrical Engineering", Prentice Hall.
4. Lima, E. L. et al. (2004), "A matemática do Ensino Médio", Volume I, Sociedade Brasileira de Matemática.
5. Lima, E. L. et al. (2004), "A matemática do Ensino Médio", Volume II, Sociedade Brasileira de Matemática.