Probabilidade

Mario A. Gneri, Hervé J. F. Guiol & Aluísio S. Pinheiro

Instituto de Matemática, Estatística e Computação Científica
UNICAMP, Brasil

8 de Março de 2002

Introdução

Este livro é o resultado de alguns anos de ensino de disciplinas de Probabilidade básica para alunos de cursos de Estatística, Física, Matemática e Matemática Aplicada na Universidade Estadual de Campinas. Este primeiro volume segue ponto por ponto a ementa do curso de Probabilidade I oferecido para os alunos do curso de Estatística. Seu objetivo principal é suprir alunos desse curso e afins com o material necessário para uma primeira disciplina de Probabilidade.

Tentamos cobrir os conceitos principais e basais da forma mais rigorosa possível nesse nível elementar, mas sem que o grau de abstração se colocasse como um obstaculo à absorção dos fundamentos tanto do ponto de vista conceitual como operacional.

O texto se apresenta dividido em seis capítulos e quatro apêndices. Os cinco primeiros capítulos formam a ementa das disciplinas em questão e, portanto, seu conteúdo deve ser entendido ao final de um semestre letivo. Uma quantidade razoável de exemplos é fornecida em cada capítulo, tanto como motivação para novos conceitos como para derivar uma aplicação dos conceitos já estudados. Além disso, ao final de cada capítulo, o(a) aluno(a) poderá encontrar um conjunto de exercícios, cuja solução é indispensável para reforçar e aprofundar as noções apresentadas.

No primeiro capítulo, introduzimos as noções mais importantes para uma correta apresentação do conceito de probabilidade e exemplos que motivam sua necessidade. A definição de Probabilidade é introduzida no Capítulo 2, por motivações e através de sua definição axiomática. No terceiro capítulo, apresenta-se a noção de variável aleatória e suas principais propriedades.

No nível desse material, utiliza-se a divisão usual das variáveis aleatórias em dois grupos: discretas e (absolutamente) contínuas. Essa divisão induz uma classificação de modelos probabilísticos em discretos e contínuos.

Apresentam-se os principais modelos probabilísticos discretos no capítulo 4, onde também sao estudadas sua propriedades, tendo o Capítulo 5 o mesmo teor para os principais modelos contínuos. Nesses dois capítulos, mostramos um número relativamente pequeno de modelos que, no entanto, sao utilíssimos para uma variedade enorme de problemas. Para tanto, basta-nos detectar analogias formais entre diferentes fenômenos e, com base nessas semelhanças, agrupá-los em classes.

No Capítulo 6, apresentam-se, de forma expositória, alguns dos resultados mais fundamentais da teoria probabilística: as Leis dos Grandes Numeros (Fracas e Fortes) e o Teorema Central do Limite. Esses formam a pedra sobre a qual todas as Ciências Estatísticas se assentam. Sao idéias sofisticadas, usualmente apresentadas com rigor numa segunda disciplina. Nosso objetivo no texto é apresentá-las conceitualmente numa primeira oportunidade, de forma a facilitar seu entendimento quando apresentadas formalmente.

O livro contem ainda quatro apêndices cujos respectivos títulos são: Teoria Ingênua dos Conjuntos, Ferramentas de Análise, Tabelas de Distribuições e Notação. Esses apêndices são formulários cujo conhecimento é necessário para a entendimento do curso.

O material do livro é tentativamente completo mas não fechado, no sentido de que se abrem questões aos leitores cujas respostas são encontradas em outros textos. Para isso, os leitores são fortemente convidados à consulta das referências. Sejamos bem claros: esse livro não pode ser visto (nem tem tal pretensão) como um substituto às obras-primas como os textos de Feller ou Shiryayev, só para citar alguns ...

A construção do conceito de probabilidade foi longa e controvertida, a tal ponto que passaram-se três séculos de história documentada, desde as primeiras incursões até uma definição universalmente aceita.

O conceito de probabilidade nasceu associado aos jogos e, em essência, assim continuou, até o comeco do século XIX. Laplace, em sua obra *Théorie Analitique des Probabilités* (1812), além de abordar a teoria dos jogos, mostrou que a noção de probabilidade poderia ser utilizada na resolução de grande variedade de problemas. Essa consideraçao ficou ainda mais evidente no decorrer do século XIX. Nesse mesmo século, o conceito de probabilidade teve papel decisivo do desenvolvimento da teoria dos erros, da matemática atuarial e da mecânica estatística.

O uso generalizado e crescente da Teoria das Probabilidades se deve ao fato de ela fornecer os elementos práticos e teóricos para lidar com a incerteza, seja essa induzida pela ignorância parcial ou total ou mesmo da natureza intrínseca do fenômeno em estudo. Estamos habituados ao uso de noções de probabilidade nas mais diversas áreas do conhecimento: Biologia, Economia, Engenharias, Genética, Medicina, Psicologia etc.

A Teoria de Amostragem e o Planejamento de Experimentos, grandemente desenvolvidos durante o século XX tornaram possível a utilização sistemática dos conceitos de probabilidade na experimentação científica, tanto nas ciências naturais como nas ciências humanas e biológicas. Mais ainda, propiciaram o surgimento de uma nova área do conhecimento: as Ciências Estatísticas.

E muito difícil nos dias atuais que se passe um dia sem obter alguma informação probabilística: cotação da Bolsa de Valores, média de Gols do Campeonato Brasileiro, Chances de Chuva no dia seguinte etc. Nosso objetivo é introduzir os alunos e alunas nesse conceito tão fundamental e moderno como a PROBABILIDADE.

Os autores gostariam de agradecer aos vários colegas e alunos do Departamento de Estatística da Unicamp, que gentilmente utilizaram o texto trazendo comentários e apontando erros. Essa colaboração foi fundamental para o correto desenvolvimento do projeto. Especiais agradecimentos devem ser feitos a Benilton de Sá Carvalho pelo apoio em várias dúvidas de LaTeX. Também agradecemos ao IMECC-Unicamp pela infraestrutura fornecida ao projeto.

HG agradece à FAPESP, Projeto Temático n. 99/11962-9 *Fenômenos Críticos em Processos Evolutivos e Sistemas em Equilíbrio*, o PRONEX n. 41.96.0923.00 *Fenômenos Críticos em Probabilidade e Processos Estocásticos* e a hospitalidade da Ecole Polytechnique Fédérale de Lausanne, Suiça, onde parte deste trabalho foi realizado.

Campinas e Lausanne, março de 2002

Conteúdo

Lista de Figuras

Lista de Tabelas

Capítulo 1

Noções básicas

Neste capítulo, introduzimos as duas noções fundamentais de *Experimento Aleatório* e de *Eventos*. Prosseguimos com elementos de análise combinatória.

1.1 Experimento Aleatório e Ω

A primeira noção fundamental da teoria da probabilidade, o experimento aleatório, é de fato aquela de maior importância, pois sua expressão correta permite construir um modelo capaz de modelá-la e, portanto, entender e prever seus resultados.

No entanto, para o entendimento da natureza de um experimento aleatório, devemos ter clara em nossa mente a definição de um experimento (que estaremos chamando aqui de *Experimento Determinístico*, isto é, não submetido ao acaso).

Definição 1.1.1 (Experimento Determinístico)

Sejam: $\Xi = \{\xi \ : \ \xi \in \Xi\}$ *um certo conjunto de ações;* R *um conjunto de resultados; e* r *uma aplicação de* Ξ *em* R*.* A aplicação *de cada elemento* $\xi \in \Xi$ *conduz a um resultado único* $r(\xi)$*, isto é, sempre que a ação* ξ *for tomada, tem-se como resultado* $r(\xi)$*. Então, dizemos que* (Ξ, R, r) *é um experimento determinístico.*

Deve-se entender da definição 1.1.1 que, num experimento determinístico, sempre que realizarmos uma ação controlada, temos condições de saber exatamente qual será o resultado obtido.

Exemplo 1.1.1 *Estou num área vazia e tenho uma caneta na mão. Decido jogá-la para ver o que acontece. Vamos supor (o que é bem natural) que a única força a que a caneta (uma vez fora de minha mão) está submetida seja a gravitação terrestre.*

O conjunto de ações Ξ tem por elementos: 'abrir a mão e deixar cair a caneta', 'lançar a caneta para cima', 'jogar com força a caneta no chão' etc. O resultado de qualquer uma dessas ações será invariavelmente 'a caneta cai no chão'.

Exemplo 1.1.2 *Temos à nossa frente torcedores fanáticos, distinguíveis, da Argentina, Brasil e Inglaterra. Pede-se-lhes que escolham qual foi o maior jogador, entre Bobby Charlton, Maradona e Pelé. O conjunto Ξ é constituído pelos elementos $\xi =$ 'fazer a pergunta para um torcedor argentino', $\eta =$'fazer a pergunta para um torcedor brasileiro' e $\zeta =$'fazer a pergunta para um torcedor inglês'. Vamos ter claramente $r(\xi) =$'Maradona', $r(\eta) =$'Pelé' e $r(\zeta) =$'Bobby Charlton'.*

O leitor poderá, com uma certa razão, reclamar da artificialidade dos exemplos 1.1.1 e 1.1.2. Na *realidade*, é difícil obter um exemplo de experimentos determinísticos pois, sob quaisquer circunstâncias, haverá sempre argumentos razoáveis para evidenciar outros possíveis resultados. Além disso, o tipo de resultado de um experimento determinístico não é tão interessante... Pensando no exemplo 1.1.1, resultados do tipo 'a caneta leva menos de cinco segundos para cair' seriam mais interessantes para caraterizar o fenômeno. Mas, mesmo numa situação muito idealizada, o número de parâmetros que deveríamos fixar para respondê-la (conhecimento total das condições iniciais) seria muito grande. Mais ainda, essas condições são, em sua maior parte, incontroláveis na prática. No caso do exemplo 1.1.2, um modelo bem razoável seria aquele cujos torcedores não fossem tão fanáticos e pudessem gostar de um jogador de um outro país, caso em que todos escolheriam Pelé.

Por essa razão, há a necessidade de um conceito mais geral de experimento, em que múltiplos resultados possam ser obtidos de uma única ação. Isto se dá com o conceito de **experimento aleatório**. Isto significa que uma determinada ação ξ resulta em um elemento de $r(\xi)$ que, desta vez, é um conjunto. Portanto, cada vez que se faz o experimento, o resultado pode ser diferente, mesmo tomando-se (aparentemente) a mesma ação. Para qualquer ξ, os conjuntos $r(\xi)$ têm várias cardinalidades possíveis. O estudo de probabilidade tem, exatamente, o objetivo de *relacionar* ξ e $r(\xi)$ de alguma forma quantitativa.

Definição 1.1.2 (Experimento Aleatório)

Um **experimento aleatório** *se descreve com o conjunto dos resultados possíveis de um experimento. Denota-se por ω um tal resultado e Ω o espaço formado por todos esses resultados. O espaço Ω é chamado de* **universo** *ou* **espaço amostral**.

Exemplo 1.1.3 (Lançamento de Dois Dados)

Lançamos dois dados, um verde e o outro azul; anota-se o resultado do dado verde e depois do dado azul. Um resultado possível será do tipo

$$\omega = (x, y), \quad 1 \leq x, y \leq 6,$$

onde x representa o resultado do dado verde e y o resultado do dado azul.

O espaço amostral pode então ser escrito como $\Omega = \{1, \ldots, 6\} \times \{1, \ldots, 6\}$. É um conjunto finito, com cardinalidade (número de elementos) $|\Omega| = 36$.

Exemplo 1.1.4 (Embaralhamento de n Cartas)

Por conveniência, vamos supor que as cartas estejam numeradas de 1 a n. Misturamos o baralho e observamos a seqüência de cartas assim obtida. Um resultado será do tipo

$$\omega = (\sigma_1, \ldots, \sigma_n),$$

onde $\sigma_i \in \{1, \ldots, n\}$ *e* $\sigma_i \neq \sigma_j$ *para quaisquer* $i \neq j$.

O universo é, então,

$$\Omega = \mathcal{S}_n,$$

o conjunto das permutações (aplicações bijectivas) de $\{1, \ldots, n\}$. *Este espaço é também finito: mostraremos que* $|\Omega| = n!$.

Exemplo 1.1.5 (Codificação de uma Imagem)

Considera-se a seqüência dos bytes que servem para codificar uma imagem a ser transmitida via satélite. Um resultado possível é uma seqüência (finita) de 0's e de 1's: a qualidade da imagem transmitida será melhor quando o tamanho dessa seqüência for maior (em geral). O espaço amostral é $\Omega = \{0, 1\}^n$, *onde* n *representa o* **tamanho** *da imagem. É um espaço finito, com* 2^n *elementos.*

No caso da figura 1.1, temos uma foto monocromática digitalizada com 890×597, onde cada *pixel* é um número binário variando de 0 (branco) a 1 (preto).

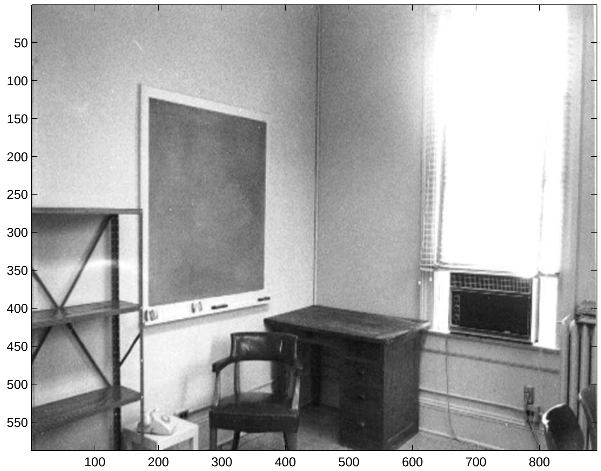
Exemplo 1.1.6 (Tempo de Vida)

Observação do tempo de vida de indivíduos de uma população biológica. Um resultado possível pode ser do tipo $\omega = 79$ *anos ou* $\omega = 29h45'57$"

Nesse caso, podemos usar $\Omega = \{1, \ldots, 200\}$, $\mathbb{N}$, $[0, T]$ *ou* $\mathbb{R}^+$. *O primeiro é um espaço finito, o segundo enumerável e os dois últimos, não-enumeráveis.*

Um dos maiores interesses para o estudo das características do tempo de vida de uma população (ou grupo) reside no planejamento de políticas de saúde pública, seguridade social, planos de proventos de aposentadoria ou pensão etc, para cuja realização faz-se fundamental um conhecimento preciso

Figura 1.1: Fotografia Codificada Digitalmente



da estrutura etária (principalmente mas não exclusivamente) da população ou grupo em questão. Os estudos variam bastante em suas perspectivas. A seguir, na tabela 1.1, apresentamos a tabela de mortalidade, para ambos os sexos da população brasileira, no ano de 1999. Entre as principais informações fornecidas por uma tal tabela, encontram-se a *Expectativa de Vida Restante* (sétima coluna), *Óbitos no próximo ano* (nas colunas dois e três), entre outras.

Tabela 1.1: BRASIL: TÁBUA COMPLETA DE MORTALIDADEAMBOS OS SEXOS - 1999

Idades Exatas	$Q(X,N)$ (por mil)	Óbitos $D(X,N)$	$l(X)$	$l(X,N)$	$T(X)$	$E(X)$
0	34,800	3480	100000	97036	6835840	68,4
1	2,341	226	96520	96407	6738804	69,8
2	1,225	118	96294	96235	6642397	69,0
3	0,884	85	96176	96134	6546162	68,1
4	0,801	77	96091	96053	6450028	67,1
5	0,664	64	96014	95982	6353976	66,2
6	0,522	50	95950	95925	6257994	65,2
7	0,412	40	95900	95880	6162068	64,3
8	0,337	32	95861	95845	6066188	63,3
9	0,297	28	95828	95814	5970343	62,3
10	0,293	28	95800	95786	5874529	61,3
11	0,325	31	95772	95756	5778743	60,3
12	0,402	39	95741	95721	5682987	59,4
13	0,513	49	95702	95678	5587265	58,4
14	0,656	63	95653	95622	5491588	57,4
15	0,821	78	95590	95551	5395966	56,4
16	0,992	95	95512	95465	5300415	55,5

17	1,153	110	95417	95362	5204950	54,5
18	1,292	123	95307	95246	5109588	53,6
19	1,414	135	95184	95117	5014343	52,7
20	1,540	146	95049	94976	4919226	51,8
21	1,667	158	94903	94824	4824250	50,8
22	1,773	168	94745	94661	4729426	49,9
23	1,852	175	94577	94489	4634765	49,0
24	1,909	180	94402	94311	4540276	48,1
25	1,961	185	94221	94129	4445965	47,2
26	2,018	190	94037	93942	4351836	46,3
27	2,082	195	93847	93749	4257894	45,4
28	2,158	202	93651	93550	4164145	44,5
29	2,246	210	93449	93344	4070595	43,6
30	2,340	218	93239	93130	3977250	42,7
31	2,439	227	93021	92908	3884120	41,8
32	2,548	236	92794	92676	3791212	40,9
33	2,668	247	92558	92434	3698536	40,0
34	2,801	259	92311	92182	3606102	39,1
35	2,950	272	92052	91917	3513920	38,2
36	3,113	286	91781	91638	3422004	37,3
37	3,286	301	91495	91345	3330366	36,4
38	3,467	316	91194	91036	3239021	35,5
39	3,663	333	90878	90712	3147985	34,6
40	3,870	351	90545	90370	3057273	33,8
41	4,110	371	90195	90009	2966903	32,9
42	4,380	394	89824	89627	2876894	32,0
43	4,710	422	89429	89219	2787268	31,2
44	5,080	452	89008	88782	2698049	30,3
45	5,480	486	88555	88312	2609267	29,5
46	5,910	521	88069	87809	2520955	28,6
47	6,360	558	87548	87269	2433146	27,8
48	6,830	595	86990	86693	2345877	27,0
49	7,330	633	86396	86079	2259184	26,1
50	7,850	674	85762	85425	2173105	25,3
51	8,430	718	85088	84730	2087680	24,5
52	9,060	765	84371	83989	2002950	23,7
53	9,750	816	83606	83198	1918961	23,0
54	10,510	870	82791	82355	1835763	22,2
55	11,330	928	81920	81456	1753407	21,4
56	12,210	989	80992	80497	1671951	20,6
57	13,140	1052	80003	79477	1591454	19,9
58	14,130	1116	78951	78393	1511977	19,2
59	15,180	1182	77835	77244	1433584	18,4
60	16,340	1253	76653	76027	1356339	17,7
61	17,600	1328	75401	74737	1280312	17,0
62	18,970	1406	74073	73370	1205576	16,3
63	20,460	1487	72667	71924	1132205	15,6
64	22,080	1572	71180	70394	1060282	14,9
65	23,800	1657	69608	68780	989888	14,2
66	25,730	1749	67951	67077	921108	13,6
67	28,040	1857	66202	65274	854031	12,9
68	30,860	1986	64345	63352	788758	12,2
69	34,140	2129	62359	61295	725406	11,6
70	37,730	2273	60230	59094	664111	11,0
71	41,580	2410	57958	56753	605017	10,4
72	45,870	2548	55547	54273	548264	9,9
73	50,640	2684	52999	51657	493991	9,3
74	55,920	2814	50315	48908	442333	8,8
75	61,760	2934	47501	46034	393425	8,3
76	68,140	3037	44568	43049	347390	7,8
77	75,060	3118	41531	39972	304341	7,3
78	82,540	3171	38413	36828	264369	6,9
79	90,680	3196	35242	33644	227542	6,5
80+	1,000	32045	32045	193898	193898	6,1

Na tabela acima, usa-se a seguinte notação:

$N = 1$.

$Q(X, N)$ - Probabilidades de morte entre as idades exatas X e X+N.

$l(X)$ - Número de sobreviventes à idade exata X.

$D(X, N)$ - Número de óbitos ocorridos entre as idades X e X+N.

$L(X, N)$ - Número de pessoas-anos vividos entre as idades X e X+N.

$T(X)$ - Número de pessoas-anos vividos a partir da idade X.

$E(X)$ - Expectativa de vida à idade X.

Exemplo 1.1.7 (Bovespa)

Observação dos volumes diários de negócios na Bolsa de São Paulo. Usam-se $\Omega = \mathbb{N}$ ou $\mathbb{Z}$, que são espaços enumeráveis.

Nesse caso, o interesse reside no comportamento dos volumes negociados. Na figura 1.2, representa-se uma situação real, em que se retratam os volumes diários negociados no período de 16/06/1999 a 16/08/2001.

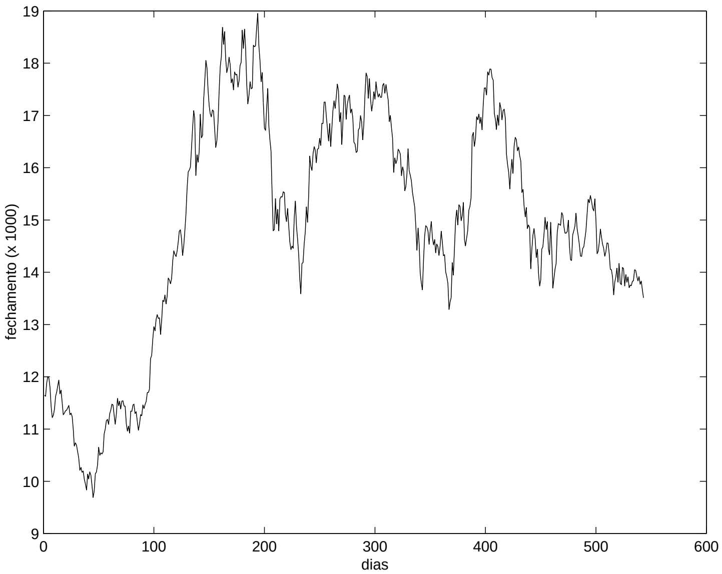
Exemplo 1.1.8 (Cara ou coroa)

Um jogo de cara ou coroa sem limite de tempo. Um resultado possível é uma seqüência de cara's ou coroa's, que podemos representar respectivamente por 1's e 0's. O espaço amostral pode ser então representado por $\Omega = \{0,1\}^{\mathbb{N}}$ que é um espaço não-enumerável. Uma típica realização de tal experimento seria dada por

$$010001100111\cdots,$$

ou seja, obteve-se cara no segundo, sexto, sétimo etc lançamentos e coroa no primeiro, terceiro, quarto, quinto, sexto etc lançamentos.

Figura 1.2: Índice de Fechamento da BOVESPA,
no período de 16/06/1999 a 16/08/2001



Exemplo 1.1.9 (Sinal Captado)

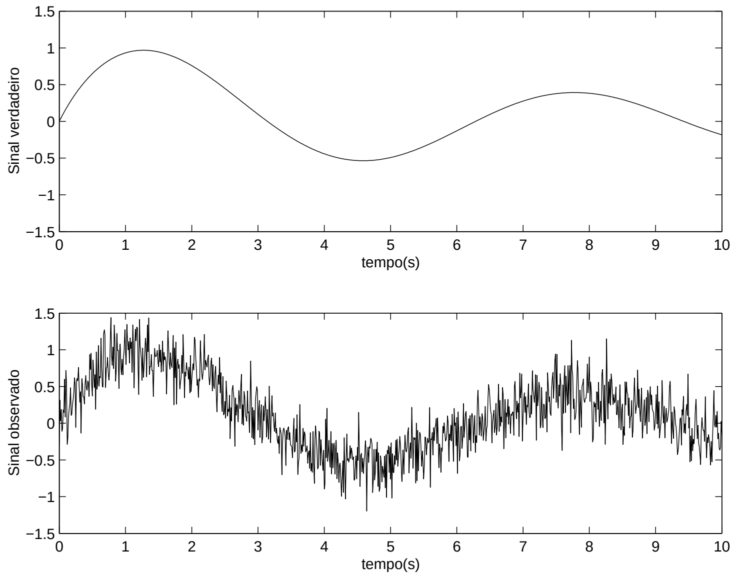
Observação durante um intervalo de tempo $[t_1, t_2]$ de um sinal. Um resultado possível é $\omega = f(t)$, onde f representa o sinal. Temos, por exemplo,

$$\Omega = \mathcal{C}([t_1, t_2]),$$

o espaço das funções contínuas sobre o intervalo $[t_1, t_2]$, com valores em $\mathbb{R}$. Este conjunto é não-enumerável.

Um exemplo típico de tal fenômeno é o apresentado na figura 1.3, cujo gráfico superior ilustra uma observação idealizada, isto é, onde não há qualquer ruído enquanto a segunda ilustra uma situação mais realista, onde a função do sinal do gráfico superior é observada com ruído.

Figura 1.3: Sinal Captado



1.2 Eventos

A segunda noção fundamental da Teoria de Probabilidade é aquela do *Evento Aleatório*. Consideramos que a realização (**sucesso**) ou não-realização (**fracasso**) de um evento aleatório depende exclusivamente do resultado do experimento associado.

Definição 1.2.1 (Evento Aleatório)

Um **evento aleatório** *A será representado como o conjunto dos resultados ω do experimento que o realizam:*

$$A = \{\omega \in \Omega : A \text{ é realizado se } \omega \text{ é o resultado do experimento}\}.$$

Exemplo 1.2.1 *Para o experimento (1.1.3): lançamento de dois dados, temos*

$$\Omega = \{1, \ldots, 6\} \times \{1, \ldots, 6\}$$

e o evento $A =$ 'a soma do resultado é par' pode ser descrito por

$$\begin{aligned} A \;=\; & \{(1,1);(1,3);(1,5);(2,2);(2,4);(2,6);(3,1);(3,3);(3,5); \\ & (4,2);(4,4);(4,6);(5,1);(5,3);(5,5);(6,2);(6,4);(6,6)\}. \end{aligned}$$

Exemplo 1.2.2 *No experimento aleatório dado no exemplo 1.1.9, onde um sinal é observado durante um intervalo de tempo $[t_1, t_2]$, o evento aleatório 'a amplitude do barulho não ultrapassa em variação absoluta a quantidade a' pode ser representado no espaço $\Omega = \mathcal{C}([t_1, t_2])$ pela bola*

$$\{\omega \in \Omega : \sup_{t \in [t_1, t_2]} |\omega(t)| \leq a\}.$$

A seguir, vamos descrever as operações lógicas que podem ser feitas com os eventos aleatórios. Do ponto de visto axiomático, a noção de evento aleatório é verdadeiramente definida por essas operações.

A cada evento A, é associado seu **complementar** (ou contrário), denotado porA^c:

$$A^c = \Omega - A.$$

Exemplo 1.2.3 *Para o experimento definido no exemplo 1.1.3: lançamento de dois dados, o complementar do evento $A =$ 'a soma dos resultados é par' é dado por:*

$$A^c = \Omega - A = \text{'a soma é ímpar'}.$$

Para cada par de eventos (A_1, A_2), o evento A_1 **e** A_2 (ou realização simultânea), denotado por $A_1 \cap A_2$, é o evento que é realizado se **ambos** os eventos A_1 e A_2 são realizados.

> **Exemplo 1.2.4** *Novamente para o lançamento de dois dados, o evento 'a soma é par' e 'o primeiro dado dá 6':=* $A \cap B$ *é dado por*
>
> $$A \cap B = \{(6,2); (6,4); (6,6)\}.$$

O evento **impossível** será denotado por $\emptyset$. A fórmula $A \cap B = \emptyset$ significa que os eventos A e B são **incompatíveis** ou mutuamente exclusivos.

Para cada par de eventos (A_1, A_2), o evento A_1 **ou** A_2, denotado por $A_1 \cup A_2$, é o evento realizado quando **pelo menos um** dos eventos A_1 ou A_2 é realizado.

O evento **certo** é denotado por Ω, porque é realizado para todo resultado ω do experimento aleatório.

Além de definir as relações entre dois específicos eventos, as operações precedentes podem ser utilizadas para combinar mais de dois eventos. Se, por exemplo, $(A_n)_{n \in \mathcal{N}}$ é uma seqüência finita ou infinita de eventos, $\cup_n A_n$ significa o evento A_1 *ou* A_2 *ou* $\ldots$, e $\cap_n A_n$ significa o evento A_1 *e* A_2 *e* $\ldots$.

Duas operações que fazem sentido apenas para dois eventos são, no entanto, extremamente úteis. A **diferença** de dois conjuntos, denotada por $-$, é definida por:

$$A - B = A \cap B^c.$$

Do ponto de visto de eventos, $A - B$ é o evento tal que A acontece mas B não. Conseqüentemente, se $A \cap B = \emptyset$, então $A - B = A$ e, se $A \subset B$, então $A - B = \emptyset$.

A **diferença simétrica** de dois conjuntos, denotada por $\triangle$, é definida por

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

Do ponto de visto de eventos, $A \triangle B$ é o evento tal que A ou B **acontecem exclusivamente**. Conseqüentemente, se $A \cap B = \emptyset$, então $A \triangle B = A \cup B$ e, se $A \subset B$, então $A \triangle B = B - A$.

1.3 Elementos de Análise Combinatória

1.3.1 Princípio da Multiplicação

Se uma *operação* A_1 pode ser realizada de n_1 maneiras e uma outra *operação* A_2 pode ser realizada de n_2 maneiras, então as duas *operações* podem ser realizadas, uma depois da outra de forma independente, de $n_1 \times n_2$ maneiras.

1.3.2 Número de Permutações

Uma permutação sobre um conjunto finito $E = \{e_1, \ldots, e_n\}$ é uma seqüência finita $(\sigma_1, \ldots, \sigma_n)$, onde $\sigma_i \in E$ e $\sigma_i \neq \sigma_j$, para $i \neq j$. Denotamos por $\mathcal{S}_n$ o conjunto das permutações de n elementos.

O número de permutações diferentes de n elementos é dado por

$$|\mathcal{S}_n| = n! = n.(n-1).\cdots.2.1. \tag{1.1}$$

Exemplo 1.3.1 *Tenho três CD's diferentes (I, II e III) para oferecer a três amigos (Fulana, Beltrano, Joaquim). O número de possibilidades de distribuir os CD's para eles é dado por* $3! = 6$.

1.3.3 Coeficientes Binomiais

Por definição, o coeficiente binomial é o número de subpopulações[1] (combinações) de tamanho k de uma população de tamanho $n \geq k$. Esse número é

[1] o termo **população de tamanho** n será usado para denotar um agregado de n elementos **sem que se leve em conta sua ordem**

denotado por $\binom{n}{k}$.

Neste caso, não estamos interessados na ordem. Iremos, a seguir, relacionar o número de combinações de k objetos sorteados em n com o número de permutações de n objetos. Pensemos que, definida uma permutação qualquer,

$$\underbrace{a_1, a_2, \ldots, a_{k-1}, a_k}_{k \text{ objetos}}, \overbrace{a_{k+1}, \ldots, a_n}^{(n-k) \text{ objetos}},$$

onde $a_1, \ldots, a_n \in \{1, 2, \ldots, n\}$, é-nos indiferente quais as particulares ordens em que $a_1, \ldots, a_k$ nos são apresentados **dentro do primeiro grupo** ou em que $a_{k+1}, \ldots, a_n$ nos são apresentados **dentro do segundo grupo**, além do simples fato de a qual grupo um particular objeto pertença. Sendo assim, cada elemento de uma subpopulação de k dos n objetos originais equivale à permutação de k objetos (tamanho do primeiro grupo) vezes a permutação de $(n-k)$ objetos (tamanho do segundo grupo) **diferentes** permutações dos n objetos originais, isto é,

$$\binom{n}{k}.k!.(n-k)! = n!$$

Então,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}. \tag{1.2}$$

Esses coeficientes têm as seguintes importantes propriedades:

Fórmula do Binômio

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}. \tag{1.3}$$

Triângulo de Pascal

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k} \tag{1.4}$$

1.3.4 Amostragem

Vamos aqui estudar o caso em que se sorteiam k bolas de uma urna que contém n bolas distinguíveis. Seja E um conjunto com n elementos distintos, $E = \{1, \ldots, n\}$.

Qualquer arranjo **ordenado**, denotado por $(e_1, e_2, \ldots, e_k)$, de k índices, onde e_i representa o número da i-ésima bola retirada, representa uma **amostra ordenada de tamanho** k retirada do conjunto E.

Qualquer subconjunto **desordenado**, denotado por $\{e_1, e_2, \ldots, e_k\}$, de k índices, é chamada de uma **amostra desordenada de tamanho** k retirada do conjunto E. Como a ordem não importa nesse caso, podemos supor que $e_1 \leq e_2 \leq \ldots \leq e_k$.

1.3.4.1 Amostras com Reposição Ordenadas

Nesse caso, cada seleção (de uma bola) é feita em toda a população, isto é, a cada extração, a bola retirada é recolocada na urna. Portanto, o mesmo elemento pode ser selecionado mais de uma vez.

O espaço amostral é

$$\Omega = \{\omega : \omega = (e_1, e_2, \ldots, e_k), e_i = 1, \ldots, n\}.$$

Cada um dos k elementos pode ser escolhido de n maneiras: o número de amostras ordenadas, com reposição, é, portanto

$$|\Omega| = n^k. \tag{1.5}$$

Exemplo 1.3.2

1) Lançamento de três dados (distinguíveis) não-viciados. Usa-se o modelo de urna $E = \{1, 2, 3, 4, 5, 6\}$. *O número de lançamentos distintos é, portanto, igual a* $6^3 = 216$.

2) Cara ou coroa. Lançamos uma moeda, não viciada, cinco vezes. O numero de resultados possíveis é $2^5 = 32$.

1.3.4.2 Amostras sem Reposição Desordenadas

Como no caso da amostragem sem reposição ordenada, um elemento, uma vez escolhido, é removido do conjunto e, portanto, não há qualquer repetição na seleção. O tamanho k deve ser também no máximo igual ao tamanho n do conjunto. O espaço amostral é

$$\Omega = \{\omega : \omega = \{e_1, e_2, \ldots, e_k\},\ e_i < e_j,\ i \neq j, e_l = 1, \ldots, n\}.$$

Estas amostras são exatamente o que se define como combinações (subpopulações) de k elementos de um conjunto de cardinalidade n:

$$|\Omega| = \binom{n}{k}. \tag{1.6}$$

Exemplo 1.3.3 (Jogo de cartas)

No Pôquer, quantas mãos distintas têm dois ases? Note que uma mão tem cinco cartas e devemos pegar duas que sejam ases entre os quatro possíveis (espadas, copas, ouros, paus), sorteando-se as três outras entre as 48 restantes. Então, têm-se

$$\tbinom{4}{2} . \tbinom{48}{3} = 103776$$

maneiras distintas de fazê-lo.

1.3.4.3 Amostras sem reposição ordenadas

No caso de amostras sem reposição ordenadas, um elemento, uma vez escolhido, é removido do conjunto e, portanto, não há qualquer repetição nos elementos selecionados. Obviamente, o tamanho amostral k deve ser no máximo igual ao tamanho populacional n.

O espaço amostral é

$$\Omega = \{\omega : \omega = (e_1, e_2, \ldots, e_k), (e_i \neq e_j, i \neq j), e_l = 1, \ldots, n\}.$$

Usando o princípio da multiplicação, temos:

$$|\Omega| = (n)_k = \frac{n!}{(n-k)!}.^2 \tag{1.7}$$

Exemplo 1.3.4 (Pódio da Copa do Mundo)

Numa Copa do Mundo, quantos pódios diferentes são possíveis? Note que um pódio é composto pelo campeão, o vice-campeão, o terceiro e o quarto colocados.

Temos $(32)_4 = 863.040$.

1.3.4.4 Amostras com Reposição Desordenadas

Uma amostra do tipo $(4, 3, 2, 1)$ será considerada idêntica a $(1, 3, 2, 4)$, sendo escrita $\{1, 2, 3, 4\}$. Como no primeiro caso, cada vez que uma bola é tirada da urna, ela é recolocada para o sorteio seguinte.

O espaço amostral é

$$\Omega = \{\omega : \omega = \{e_1, e_2, \ldots, e_k\},\ e_i \leq e_j,\ e_i = 1, \ldots, n\}.$$

O número de possíveis amostras diferentes é dado por:

[2] Na literatura, usam-se também os termos **Arranjo de n elementos k a k** ou também **n baixado a k** para $(n)_k$

$$|\Omega| = \binom{n+k-1}{k}. \tag{1.8}$$

Demonstração:

Suponha que representemos cada bola retirada por $\bigcirc$ e joguemos essas k $\bigcirc$'s numa caixa com n divisões. Para saber quantas vezes uma determinada bola foi retirada, basta contar o número de $\bigcirc$'s dentro da respectiva caixa.

Finalmente, o argumento se completa da seguinte forma: como as bolas têm que cair dentro de alguma caixa, basta controlar a posição relativa dos $n-1$ divisores internos (desprezando-se o primeiro e último) e as k $\bigcirc$'s. As possíveis realizações serão do tipo:

$$| \bigcirc \ ||| \ \bigcirc \bigcirc || \ \cdots \ | \bigcirc \bigcirc |.$$

No exemplo acima, houve uma retirada de bola 1, nenhuma de bolas 2 ou 3, duas de bola 4 etc.

O problema se resume, portanto, a escolher k posições (para os $\bigcirc$'s) de $n+k-1$ possíveis sem que a ordem de escolha seja importante , ou seja, $\binom{n+k-1}{k}$. ∎

Exemplo 1.3.5 *Número dos resultados (sem ordem) do lançamento de três dados (indistinguíveis). Temos* $\binom{6+3-1}{3} = 56$ *possibilidades.*

Exemplo 1.3.6 *Dentro de uma urna temos três bolas: 1,2,3, das quais sorteiam-se duas. Representamos todos os casos de amostras possíveis na tabela 1.2.*

1.3.5 Problemas de Ocupação

Os problemas de ocupação têm por objetivo o estudo de modelos associados à distribuição aleatória de k bolas em n compartimentos. Por exemplo, em

Tabela 1.2: Tipos de Amostragem

Distr. / Amostra	Ordenada	Desordenada
Com Reposição	$(1,1);(1,2);(1,3)$	$\{1,1\};\{1,2\};\{1,3\}$
	$(2,1);(2,2)(2,3)$	$\{2,2\};\{2,3\}$
	$(3,1);(3,2);(3,3)$	$\{3,3\}$
Sem Reposição	$(1,2);(1,3)$	
	$(2,1);(2,3)$	$\{1,2\};\{1,3\}$
	$(3,1);(3,2)$	$\{2,3\}$

Física Estatística, quando se estuda a distribuição de k partículas (que podem ser prótons, elétrons etc.) em n diferentes estados (que podem ser níveis de energia), utilizam-se as técnicas para problemas de ocupação. As condições impostas são:

1. Os compartimentos serão numerados de 1 a n;

2. As bolas são inicialmente distinguíveis e numeradas de 1 a k. Portanto, a distribuição de k bolas por n compartimentos é completamente descrita por um conjunto ordenado $(e_1, \ldots, e_k)$, onde e_i é o número do compartimento onde a i-ésima bola se encontra; e

3. No caso em que as bolas são indistinguíveis, sua distribuição entre os n compartimentos é completamente descrita pelo conjunto desordenado $\{e_1, \ldots, e_k\}$, onde e_i é o número do compartimento em que uma bola é colocada na i-ésima etapa (note que a *ordem* pode ser arbitrária).

Temos a seguinte correspondência:

- Amostras ordenadas $\longleftrightarrow$ Bolas Distinguíveis;
- Amostras desordenadas $\longleftrightarrow$ Bolas Indistinguíveis;
- Amostras com reposição $\longleftrightarrow$ Um compartimento pode ser ocupado por mais de uma bola; e

- Amostra sem reposição $\longleftrightarrow$ Um compartimento pode ser ocupado por no máximo uma bola.

  Exemplo 1.3.7 *Colocar dois objetos em três compartimentos (conforme Tabela 1.3).*

Tabela 1.3: Ocupação de Três Compartimentos por Dois Objetos

Distr. / Tipo de objetos	Distinguíveis	Indistinguíveis
Sem Exclusão	$(BP\vert.\vert.); (B\vert P\vert.); (B\vert.\vert P)$	$(++\vert.\vert.); (.\vert++\vert.); (.\vert.\vert++)$
	$(.\vert BP\vert.); (P\vert B\vert.); (.\vert B\vert P)$	$(+\vert+\vert.); (.\vert+\vert+)$
	$(P\vert.\vert B); (.\vert P\vert B); (.\vert.\vert PB)$	$(+\vert.\vert+)$
Com Exclusão	$(B\vert P\vert.); (B\vert.\vert P)$	
	$(P\vert B\vert.); (.\vert B\vert P)$	$(+\vert+\vert.); (.\vert+\vert+)$
	$(P\vert.\vert B); (.\vert P\vert B)$	$(+\vert.\vert+)$

Em muitas situações, é necessário considerarmos as bolas como indistinguíveis, como as citadas nos exemplos 1.3.8 e 1.3.9.

Exemplo 1.3.8 *Estudos estatísticos da distribuição de acidentes pelos dias da semana ou dos aniversários pelos dias do ano. Os dois casos são equivalentes, respectivamente, à ocupação de* $n = 7$ *(dias da semana) e* $n = 366$ *(dias do ano) compartimentos por um número aleatório (desconhecido, a princípio) de acidentes ou aniversários.*

Exemplo 1.3.9 *O lançamento de* k *dados, equivalente à distribuição de* k *bolas em* $n = 6$ *compartimentos.*

1.3.6 Coeficientes Multinomiais

Nessa seção, veremos uma natural extensão dos coeficientes binomiais, cuja interpretação amostral é a de alocação de n elementos em k grupos, tornando os coeficientes binomiais um caso particular, para $k = 2$.

Teorema 1.3.1 *Sejam $r_1, r_2, \ldots, r_k$ inteiros naturais ($r_i \in \mathbb{N}$) tais que $r_1 + r_2 + \ldots + r_k = n$. O número de maneiras em que uma população de n elementos pode ser dividida em k subpopulações, cujas respectivas freqüências absolutas são r_1, r_2, $\ldots$, r_k, é o coeficiente multinomial $\binom{n}{r_1, \ldots, r_k}$:*

$$\binom{n}{r_1, \ldots, r_k} = \frac{n!}{r_1! r_2! \cdots r_k!}. \tag{1.9}$$

Demonstração: Observe que

$$\frac{n!}{r_1! r_2! \cdots r_k!} = \binom{n}{r_1}\binom{n - r_1}{r_2}\binom{n - r_1 - r_2}{r_3}\cdots\binom{r_k}{r_k},$$

pois, para efetuarmos a partição desejada, temos que, em primeiro lugar, escolher r_1 elementos dos n disponíveis; a seguir, dos $n - r_1$ elementos restantes, selecionamos um grupo de tamanho r_2; ... ∎

1.3.7 Fórmula de Stirling

Os estudos de problemas de ocupação apresentam, em geral, a necessidade do cálculo de fatoriais. Quando o número n cujo fatorial deseja-se calcular é *grande*, enfrentam-se problemas para obtê-lo de forma direta. A Fórmula (ou aproximação) de Stirling nos permite fazê-lo de forma rápida e eficiente.

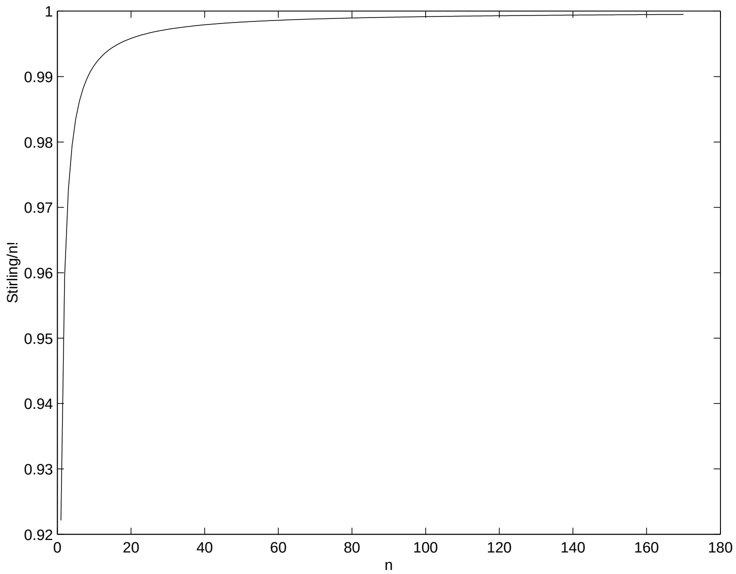
Teorema 1.3.2 (Fórmula de Stirling)

$$n! \asymp \sqrt{2\pi n}\left(\frac{n}{e}\right)^n, \tag{1.10}$$

onde o sinal $\asymp$ é usado para indicar que o quociente de um lado pelo outro de (1.10) converge para 1, quando $n \to \infty$.

Nota: A precisão da aproximação de Stirling é notável, mesmo para valores pequenos de n. Note na figura 1.3.7 como a razão entre a fórmula de Stirling e o valor de $n!$ se aproxima de 1 rapidamente.

Figura 1.4: Aproximação de Stirling para o Fatorial



Exemplo 1.3.10 *Suponha que tenhamos N objetos dos quais queiramos retirar uma amostra de tamanho n. Sabemos que a população se divide em dois grupos, de respectivos tamanhos Np e Nq (onde, obviamente, $N = Np + Nq$). A questão que se coloca é a da probabilidade*[3] *de que, dos n objetos amostrados, k sejam do tipo 1. A fórmula de Stirling pode ser usada para demonstrar que, se N é suficientemente grande, é irrelevante se calculamos tais probabilidades supondo reposição ou não.*

Seja $0 < p < 1$ e $q = 1 - p$, $N > n \in \mathbb{N}$. Para cada $k \in \{0, 1, \ldots, n\}$, a quantidade

[3] Para uma discussão detalhada do conceito de probabilidade, veja o capítulo 2.

$$h_k(N, n, Np) = \frac{\binom{Np}{k}\binom{Nq}{n-k}}{\binom{N}{n}} \tag{1.11}$$

é chamada de distribuição hipergeométrica de parâmetros (N, n, Np)[4]. Esta expressão pode ser escrita, usando a definição dos coeficientes binomiais:

$$\begin{aligned} & \frac{(Nq)!}{(Nq-(n-k))!(n-k)!} \times \frac{(Np)!}{(Np-k)!k!} \times \frac{n!\,(N-n)!}{(N)!} \\ = \quad & \binom{n}{k} \times \frac{(Nq)!}{(Nq-(n-k))!} \times \frac{(Np)!}{(Np-k)!} \times \frac{(N-n)!}{(N)!}. \end{aligned}$$

Agora, vamos usar a fórmula de Stirling para estimar cada quociente quando N cresce indefinidamente.

Temos

$$(Nq)! \asymp \sqrt{2\pi}(Nq)^{Nq+\frac{1}{2}} e^{-Nq}$$

e

$$(Nq-(n-k))! \asymp \sqrt{2\pi}(Nq-(n-k))^{Nq-(n-k)+\frac{1}{2}} e^{-Nq+(n-k)}.$$

Isto dá (rearranjando os termos)

$$\frac{(Nq)!}{(Nq-(n-k))!} \asymp (Nq)^{(n-k)} \times \left(\frac{Nq}{Nq-(n-k)}\right)^{Nq-(n-k)+\frac{1}{2}} e^{-(n-k)}.$$

Da mesma maneira, temos

$$\frac{(Np)!}{(Np-k)!} \asymp (Np)^k \times \left(\frac{Np}{Np-k}\right)^{Np-k+\frac{1}{2}} e^{-k},$$

e

[4] Veja a seção 4.6.

$$\frac{N!}{(N-n)!} \asymp (N)^n \times \left(\frac{N}{N-n}\right)^{N-n+\frac{1}{2}} e^{-n}.$$

Então, para N muito grande,

$$\frac{\binom{Nq}{n-k}\binom{Np}{k}}{\binom{N}{n}} \asymp \binom{n}{k} q^{n-k} p^k. \tag{1.12}$$

Essa última expressão é a Distribuição Binomial[5] de parâmetros n e p.

1.4 Exercícios

Exercício 1.1 *Descrever um resultado possível e escrever o conjunto Ω para os seguintes experimentos:*

1) Cara ou coroa: o jogo acaba quando a primeira coroa for encontrada; e

2) Um negociante está fazendo viagens entre São Paulo, Rio de Janeiro e Belo Horizonte. Cada cidade deve ser visitada exatamente duas vezes.

Exercício 1.2 *Continuação do exercício 1.1. Descrever os eventos:*

1) Não são observadas mais do que quatro caras antes de uma coroa; e

2) O negociante nunca visita a mesma cidade duas vezes seguidas.

Exercício 1.3 *Discorra (com suas próprias palavras) sobre fenômenos de natureza determinística e aleatória.*

Exercício 1.4 *Descrever o conjunto A^c no exemplo 1.2.3. Descrever o evento complementar no exemplo 1.2.2.*

Exercício 1.5 *Verificar as fórmulas*

a) **Distributiva**

[5] Veja a seção 4.3

$$A \cap (\cup_I B_i) = \cup_I (A \cap B_i) \ , \ A \cup (\cap_I B_i) = \cap_I (A \cup B_i); \ \ e$$

b) **Lei de de Morgan**

$$(\cup_I A_i)^c = \cap_I A_i^c \ , \ (\cap_I A_i)^c = \cup_I A_i^c.$$

Exercício 1.6 *Sejam Ω o conjunto universal, $\emptyset$ o conjunto vazio, A, B e C subconjuntos de Ω. Prove que:*

1. $\forall A \subset \Omega$, $A \cap \emptyset = \emptyset$, $A \cup \emptyset = A$;
2. $A \subset B \Leftrightarrow A \cup B = B$;
3. $A \subset B \Leftrightarrow A \cap B = A$;
4. $A \cup (B \cup C) = (A \cup B) \cup C$;
5. $A \cap (B \cap C) = (A \cap B) \cap C$;
6. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$;
7. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;
8. $A \cup A^c = \Omega$, $A \cap A^c = \emptyset$, $\emptyset^c = \Omega$, $\Omega^c = \emptyset$;
9. $(A^c)^c = A$, $A \subset B \Leftrightarrow B^c \subset A^c$;
10. $(A \cup B)^c = A^c \cap B^c$; *e*
11. $(A \cap B)^c = A^c \cup B^c$.

Exercício 1.7 *Escreva a união e a diferença de dois conjuntos, utilizando apenas as operações $\triangle$ e $\cap$.*

Exercício 1.8 *De quantas maneiras uma moça que tenha oito saias, quatro pares de sapatos e cinco blusas pode vestir-se.*

Exercício 1.9 *Demonstre (1.1).*

Exercício 1.10
(a) Mostre que $\forall n \in \mathbb{N}$ *o número de permutações de um conjunto de* n *elementos é* $n!$.

(b) Mostre que $\forall n \in \mathbb{N}^*$ *o número de subconjuntos distintos de um conjunto de* n *é* 2^n. *O que acontece se* $n \in \mathbb{N}$?

(c)Prove que $\forall n \in \mathbb{N}^*$ *um conjunto de* n *elementos tem* $\binom{n}{k}$ *subconjuntos distintos de* k *elementos,* $k \in \mathbb{N}^*$.

Exercício 1.11 *Quantas mãos distintas existem no Bridge? E no Pôquer?*

Exercício 1.12 (Números Binomiais I)
(a) Prove que $\forall n \in \mathbb{N}^*\ \forall k \in \mathbb{N}^*\quad k \leq n \quad \binom{n}{k} = \binom{n}{n-k}$;

(b) Prove a relação de Stifel: $\forall n \in \mathbb{N}^*\ \forall k \in \mathbb{N}^*\quad k \leq n$

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1};$$

(c) Prove que $\forall n \in \mathbb{N}^* \quad \sum_{k=0}^{n} \binom{n}{k} = 2^n$; *e*

(d) No caso geral (para r *real e* k *inteiro não-negativo), define-se* $\binom{r}{k}$ *por*

$$\binom{r}{k} = \frac{r(r-1)\dots(r-k+1)}{k!}.$$

Enuncie e prove a relação de Stifel neste caso.

Exercício 1.13 (Números Binomiais II)

(a) Prove o famoso Binômio de Newton: $\forall a \in \mathbb{R},\ \forall b \in \mathbb{R},\ \forall n \in \mathbb{N}^*$

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

(b) Ache $\sum_{k=0}^{n} \binom{n}{k} x^k$*, onde* n *é um número natural e* x *um número real;*

(c) Ache $\sum_{k=0}^{n} k\binom{n}{k} x^k$*, onde* n *é um número natural e* x *um número real;*

(d) Ache $\sum_{k=0}^{n} k\binom{n}{k}$*, onde* n *é um número natural;*

(e) Como seria a generalização do Binômio de Newton para o caso em que n *não seja um inteiro positivo; e*

(f) Proponha e prove o Polinômio de Leibniz (generalização do Binômio de Newton com n *ainda natural).*

Exercício 1.14 *Um técnico de futebol tem 22 jogadores a sua disposição. Quantos times (de onze jogadores) ele pode formar se*

(a) levar em consideração apenas quem entra em campo; ou

(b) considera-se além de quem entra em campo, qual a posição ocupada (suponha que haja onze posições).

Considere que o técnico dispõe de três goleiros, seis zagueiros, oito meio-campistas e cinco atacantes. Quantos times ele pode formar sob cada um dos seguintes esquemas táticos?
(c) 4-3-3;
(d) 4-2-4;
(e) 3-5-2;
(f) 3-4-3.

Assuma para os itens (c)-(f) que apenas a entrada em campo tenha importância.

Exercício 1.15 *Suponha que uma senhora esteja formando um grupo de amigos para uma viagem à Europa. Devido aos seus conhecimentos, ela* *dispõe de várias acomodações nas três cidades a que o grupo pretende ir: Roma, Londres e Madri. Em Roma, há dois apartamentos disponíveis, o primeiro com capacidade para quatro pessoas e o segundo com capacidade para seis pessoas. Em Londres, há três acomodações, com capacidade de cinco pessoas em cada e um outro com capacidade para quatro pessoas e em Madri, há duas casas, uma delas capaz de acomodar oito pessoas e a outra com capacidade de acomodar doze pessoas. O grupo será formado por quatro casais e dois homens solteiros. Quantas diferentes formações são possíveis:*

(a) quando os casais são indissolúveis;

(b) quando cada casal não sente ser imprescindível ficar na mesma casa; e

(c) quando um casal tenha brigado no avião e decidido não ficar na mesma casa durante todo o resto da viagem.

Capítulo 2

Probabilidade

2.1 Os Conceitos de Probabilidade

Probabilidade:

`A probabilidade (de um evento), em matemática: razão entre o número de casos favoráveis pelo número de casos possíveis para um evento qualquer, onde os casos possíveis têm supostamente o mesmo grau de probabilidade. Grande Enciclopédia Larousse Cultural, 1998.`

Compare o uso de probabilidade nos seguintes exemplos:

a) "A probabilidade de obter 6 com um dado comum é 1/6.";

b) "A probabilidade de que Shakespeare tenha escrito todas as peças que lhe foram atribuidas é quase nula."; e

c) "O experimento de Fresnel aumentou a probabilidade da teoria ondulatória da luz."

O sentido exato de probabilidade é, de fato, uma noção altamente polêmica, mesmo entre especialistas (filósofos, matemáticos ou estatísticos).

Na antigüidade, Aristóteles definiu a probabilidade como sendo *o que o homem sabe que pode ou não pode acontecer, ser ou não ser...*

Na epóca do Renascimento Europeu, na Itália, a importância econômica das grandes cidades foi um fator decisivo para o desenvolvimento de um interesse por seguros de vida e de bens, levando-se em conta o risco envolvido em cada situação. Alguns autores atribuem a fundação teórica dos seguros de vida ao inglês John Graunt. Em 1662, ele já estudava a estabilidade das séries estatísticas obtidas a partir de registros de óbitos. Pouco mais tarde, o astrônomo Edmund Halley (1656-1742) desenvolveu um método para calcular anuidades a partir das tabelas de mortalidade.

Simultaneamente, uma parte importante das pesquisas estava concentrada na área jurídica. Isto pode ser explicado pelo fato de existir interesse em discutir-se o peso de provas no resultado de um julgamento. Entre os matemáticos que, em algum ponto de suas carreiras, voltaram-se para esse assunto, destaca-se o famoso matemático alemão, G. Leibniz. Esse aspecto continuou importante até a metade do século XIX.

No entanto, foram os problemas matemáticos relacionados com os jogos os maiores responsáveis pelo desenvolvimento de uma teoria das probabilidades. Isto começou também no Renascimento, nos séculos XV e XVI. Um grande número de matemáticos italianos, incluindo Gerômino Cardano, Galileu e Tartaglia criaram vários métodos para entender e calcular as combinações vencedoras de alguns jogos.

Por volta de 1550, Cardano escreveu um livro "Sobre os jogos de azar" que, no entanto, só foi publicado em 1663, quando a Teoria das Probabilidades já recebia muita atenção. O assunto foi verdadeiramente desenvolvido numa rica correspondência entre os matemáticos franceses Blaise Pascal e Pierre de Fermat, no ano de 1654, sendo os problemas propostos resolvidos sob uma teoria geral de combinatória. O holandês Christiaan Huygens desenvolveu essas idéias num livro publicado em 1657, *De ratiociniis in ludo Aleæ*, logo seguido, em 1713, pelo suiço Jacob Bernoulli, com *Ars Conjectandi* (obra póstuma) e pelo francês Abraham de Moivre em *Doctrine of Chances* , 1718. Estes dois trabalhos incluíram as primeiras versões dos teoremas fundamentais da Probabilidade: a Lei dos Grandes Números e o Teorema Central do Limite.

No século XVIII, destacam-se os trabalhos realizados por Pierre Simon Laplace. Sua obra foi marcada por sua concepção *determinística* do universo: uma inteligência onisciente seria capaz de prever tudo o que deve acontecer na natureza em todos os detalhes. Neste aspecto, a Probabilidade só entraria na ciência como uma teoria dos erros, isto é, um estudo sistemático dos desvios de uma média, que aparecem numa seqüência de medidas de uma mesma quantidade. Esse estudo foi desenvolvido por Laplace numa série de livros entre 1774 e 1827. Muitos outros grandes matemáticos contribuíram para o seu desenvolvimento: Leonard Euler, A. M. Legendre, Lagrange, Simeon Denis Poisson e o *princípe* dos matemáticos, o alemão Carl Friedrich Gauss (Distribuição Normal , Método dos Mínimos Quadrados). É também nesse século que os conceitos de Probabilidade e Estatística entraram na área de Ciências Sociais, graças ao trabalho do filósofo de Condorcet.

Na segunda metade do século XIX, a Física se encontrava em grande desenvolvimento. James Clavel Maxwell, físico inglês, criou, por volta de 1860, a Teoria Cinética dos gases e, através dela, a Física Estatística. Em 1877, o físico austríaco Ludwig Boltzmann interpretou a irreversibilidade de um processo térmico como a tendência de as moléculas atingirem a distribuição de energias mais provável. Em 1900, com Max Planck, a Probabilidade também invadiu a teoria atômica na Física Quântica.

Antes do fim da primeira metade do século XX, a antiga concepção determinística de Laplace fora trocada para um ponto de visto totalmente probabilístico. Mesmo com todos os avanços, em tantas áreas do conhecimento, a teoria probabilística ainda sofria da ausência de base axiomática, fato argumentado pelo grande matemático David Hilbert, em 1900, em sua descrição dos grandes problemas a serem superados no século XX. Uma solução veio em 1933, quando outro grande matemático, o russo A. N. Kolmogorov, desenvolveu um cálculo abstrato, onde propunha-se a probabilidade como uma função de conjuntos: essa descrição é considerada a mais adequada e será a adotada neste curso.

Antes disto, porém, precisamos de ter algumas aplicações e interpretações de probabilidade que nos sugiram os axiomas propostos por Kolmogorov. Se voltamos à questão a) acima, duas das mais velhas interpretações de probabilidade se fazem importantes:

1. Uma expressão de simetria.

Acredita-se que um dado comum (normal) seja um objeto perfeitamente simétrico e, portanto, não deveria favorecer qualquer de suas faces. Isto, portanto, conduz à definição proposta na frase extraída da Enciclopédia, a famosa razão entre o número de casos favoráveis e casos possíveis. O leitor deve ser capaz de, rapidamente, recordar muitos exemplos práticos em que essa definição é suficiente. Mas não devemos esquecer que essa noção está relacionada com a hipótese de simetria (equiprobabilidade) entre os resultados e, portanto, não será difícil tampouco encontrar outros muitos casos onde ela não pode ser empregada.

2. Uma indicação da freqüência relativa

Essa idéia também está presente no exemplo a) pois acreditamos que, se lançarmos um grande número de vezes um dado, a freqüência relativa de cada face irá convergir para $1/6$ se o dado for normal. Por outro lado, essa interpretação também seria conveniente mesmo que não houvesse simetria (o dado é viciado): o limite de cada freqüência relativa descreveria de que maneira o dado seria viciado. Podemos também relacionar esta definição com a dada por Aristóteles: *o que a experiência diz que acontece ou não...*

Portanto, parece-nos que esta seria uma alternativa viável que funcionaria em casos, com presença de simetria ou não. Infelizmente, nem todos os eventos são reprodutíveis (pense no exemplo b) acima) e o cálculo de probabilidades tem que ser feito para esses casos também.

De fato, os exemplos b) e c) conduzem a interpretações bem mais complicadas de **probabilidade**. O matemático francês Emile Borel, no seu volume *Valeur pratique des probabilités*, observa que, na prática, os eventos *únicos* do tipo b) têm de fato pouco relevância pois sua probabilidade deveria ser tão pequena que sairia de toda escala prática de numeração.

O exemplo c) relaciona-se de forma mais evidente com uma interpretação chamada de *probabilidade subjetiva*, alternativa à visão clássica de probabilidade. Nesta área, destacam-se os trabalhos do italiano de Finetti e do americano Savage. Tal visão de probabilidade como uma noção subjetiva pode ser relacionada com as idéias físicas de um experimento quântico (interação experimentador - experimento) e formam o fundamento da teoria Bayesiana moderna. Este conceito de subjetividade é bastante controverso e sua utilização deve ser cuidadosa.

Os pontos centrais de muitas dessas *definições* de probabilidade podem ser, no entanto, recuperados dos axiomas de Kolmogorov , na maioria dos casos importantes. Terminamos enfatizando que as definições acima não são as únicas possíveis e que existem em diversas áreas, como na Mecânica Quântica, muitos problemas de qual conceito de probabilidade é o mais adequado.

2.2 Conceitos Fundamentais

Seja Ω um conjunto não vazio.

2.2.1 σ-álgebra

No primeiro capítulo, vimos que podemos identificar os eventos como subconjuntos de Ω. Será que qualquer subconjunto de Ω é um evento? A resposta é negativa, mas suas razões ultrapassam o nível deste curso.

De fato, é necessario pensar no conjunto dos eventos como sendo uma coleção $\mathcal{F}$ de sub-conjuntos de Ω com algumas propriedades. Em $\mathcal{F}$, deveríamos ter pelo menos Ω, que representa o evento certo. É também natural pedir que qualquer elemento de $\mathcal{F}$ tenha seu complementar em $\mathcal{F}$. Finalmente, vamos pedir que qualquer reunião de eventos seja também um evento.

Definição 2.2.1 (Sigma-álgebra)

Uma coleção $\mathcal{F}$ de subconjuntos de Ω é chamada de **σ-álgebra** *se temos (todas) as condições seguintes*

(a) $\Omega \in \mathcal{F}$;

(b) se $A_1, A_2, ... \in \mathcal{F}$, *então* $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$; *e*

(c) se $A \in \mathcal{F}$, *então* $A^c \in \mathcal{F}$.

Observação: É importante não perder de vista que uma σ-álgebra $\mathcal{F}$ é associada com o seu espaço Ω (veja o exemplo 2.2.3 seguinte).

Exemplo 2.2.1 *A menor σ-álgebra associada com Ω é* $\mathcal{F} = \{\emptyset, \Omega\}$.

Exemplo 2.2.2 *Se A é um subconjunto de Ω então* $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ *é uma σ-álgebra.*

Exemplo 2.2.3 *Seja $\mathcal{F}$ uma σ-álgebra de subconjuntos de Ω e $B \in \mathcal{F}$. Então $\mathcal{G} = \{A \cap B : A \in \mathcal{F}\}$ é uma σ-álgebra de subconjuntos de B.*

Deixamos a prova disto para o leitor (veja o exercício 2.1).

Observe que $\mathcal{G}$ não é uma σ-álgebra sobre Ω: se $C \in \mathcal{G}$ então $\exists C' \in \mathcal{F}$ tal que $C = C' \cap B$. O complementar de C **em** B é o conjunto $(C')^c \cap B$ que pertence ao $\mathcal{G}$, **mas** o complementar de C **em** Ω é o conjunto $(C')^c \cup B^c = (C' \cap B)^c$, que não é um elemento de $\mathcal{G}$. Então, **cuidado com os complementares**!!!

2.2.2 Probabilidade

Queremos achar um objeto para medir a *chance* de um evento acontecer. Referindo-se à primeira interpretação que damos na seção 2.1 (simetria), a probabilidade de um evento deve ser um número entre 0 e 1. O evento certo deve ter probabilidade 1. Também devemos pedir que a probabilidade da reunião de eventos **disjuntos** deva ser igual à soma das probabilidades individuais.

Definição 2.2.2 (Medida de Probabilidade)

Uma **Medida de Probabilidade** $\mathbb{P}$ *sobre* $(\Omega, \mathcal{F})$ *é uma função* $\mathbb{P} : \mathcal{F} \to [0, 1]$ *tal que*

(a) $\mathbb{P}(\Omega) = 1$;

(b) se $\{A_1, A_2, ...\}$ *é uma coleção (enumerável) de elementos disjuntos de* $\mathcal{F}$, *tal que* $A_i \cap A_j = \emptyset$ *para cada par* i, j, $i \neq j$, *então*

$$\mathbb{P}\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

A tripla $(\Omega, \mathcal{F}, \mathbb{P})$ *é chamada de* **Espaço de Probabilidade**.

Exemplo 2.2.4 (Cara ou Coroa)

Descreve-se o universo $\Omega = \{0, 1\}$ *(0="cara", 1="coroa") e* $\mathcal{F} = \{\emptyset, \{0\}, \{1\}, \Omega\}$. *Seja* $\mathbb{P} : \mathcal{F} \to [0, 1]$ *dada por*

$$\mathbb{P}(\emptyset) = 0,\ \mathbb{P}(\{0\}) = p,\ \mathbb{P}(\{1\}) = 1 - p,\ \mathbb{P}(\Omega) = 1,$$

onde p *é um número real fixo do intervalo* $[0, 1]$. *Se* $p = 1/2$, *a moeda é não viciada.*

Exemplo 2.2.5 (Dado)

Sejam $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ *o conjunto de partes (dos sub-conjuntos) de* Ω *e* $\mathbb{P}$ *dada por*

$$\mathbb{P}(A) = \sum_{i \in A} p_i, \text{ para qualquer } A \subset \Omega,$$

onde $p_1, p_2, ..., p_6$ *são números fixos do intervalo* $[0, 1]$, *cuja soma é 1. A probabilidade de que a face* i *apareça é* p_i. *O dado é normal se* $p_i = 1/6$, *para cada* i, *e, portanto,*

$$\mathbb{P}(A) = \frac{|A|}{6}, \text{ para qualquer } A \subset \Omega,$$

onde $|A|$ *representa o número de elementos (cardinalidade) de* A.

2.2.3 Propriedades Fundamentais

A partir de agora, os eventos A, B ou C serão elementos de uma σ-álgebra $\mathcal{F}$ sobre o espaço Ω.

Proposição 2.2.1

(a) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ *(em particular* $\mathbb{P}(\emptyset) = 0$*);*

(b) Se $A \subset B$, *então* $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B - A) \geq \mathbb{P}(A)$; *e*

(c) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Demonstração: Fácil (veja exercício 2.4). ■

A próxima propriedade é mais técnica e pode ser esquecida em uma primeira leitura. No entanto, é importantíssima, mostrando que a função $\mathbb{P}$ é *contínua* sobre os conjuntos. Ela é equivalente à condição (b) da definição 2.2.2 de medida de probabilidade (aditividade enumerável) e, portanto, em alguns livros, encontra-se em seu lugar.

Proposição 2.2.2 (Continuidade da Medida de Probabilidade)

Seja $\{A_1, A_2, ...\}$ uma seqüência crescente de eventos, isto é, tal que $A_1 \subset A_2 \subset A_3 \subset ...$, e A seu limite:

$$A = \cup_{i=1}^{\infty} A_i = \lim_{i \to \infty} A_i.$$

Então,

$$\mathbb{P}(A) = \lim_{i \to \infty} \mathbb{P}(A_i).$$

Em paralelo, seja $\{B_1, B_2, ...\}$ uma seqüência decrescente de eventos, tal que $B_1 \supset B_2 \supset B_3 \supset ...$, e B seu limite:

$$B = \cap_{i=1}^{\infty} B_i = \lim_{i \to \infty} B_i.$$

Então,

$$\mathbb{P}(B) = \lim_{i \to \infty} \mathbb{P}(B_i).$$

Demonstração:

$A = A_1 \cup (A_2 - A_1) \cup (A_3 - A_2) \cup ...$ é uma reunião de eventos disjuntos. Então, por definição,

$$\mathbb{P}(A) \quad \underset{(*)}{=} \quad \mathbb{P}(A_1) + \sum_{i=1}^{\infty} \mathbb{P}(A_{i+1} - A_i)$$

$$= \quad \mathbb{P}(A_1) + \lim_{n\to\infty} \sum_{i=1}^{n-1} [\mathbb{P}(A_{i+1}) - \mathbb{P}(A_i)]$$

$$\stackrel{(**)}{=} \quad \lim_{n\to\infty} \mathbb{P}(A_n).$$

Para o resultado com família decrescente, basta aplicar esse resultado à seqüência $\{B_1^c, B_2^c, \ldots\}$ $(***)$ (veja exercício 2.5). ■

2.3 Probabilidade Condicional e Independência

2.3.1 Probabilidade Condicional

É natural achar que a probabilidade de um evento A deva mudar quando uma informação sobre a ocorrência (ou não) de um evento B relacionado com A é dada. Por exemplo, a probabilidade de que *vai chover amanhã* pode mudar **dado que** *temos (ou não) muitas nuvens na noite de hoje.* Esta nova probabilidade de A dado B (isto é quando sabemos que B vai ocorrer ou ocorreu) é chamada de **Probabilidade Condicional** de A dado B e é denotada por $\mathbb{P}(A|B)$.

É claro que dado que um evento B vai ocorrer, o evento A pode ocorrer se e somente se $A \cap B$ ocorrer.

Levando-se em conta essa última observação, uma definição possível para a probabilidade condicional de A dado B seria considerá-la proporcional à $\mathbb{P}(A \cap B)$. Portanto, poderíamos escrever $\mathbb{P}(A|B) = \alpha\mathbb{P}(A \cap B)$, com $\alpha > 0$. Porém, a probabilidade condicional de Ω dado B deve ser igual a 1, então $\alpha\mathbb{P}(\Omega \cap B) = 1$ e, conseqüentemente, $\alpha = (\mathbb{P}(B))^{-1}$.

Isto é formalizado na definição seguinte.

Definição 2.3.1 (Probabilidade Condicional)

Sejam A e B elementos de $\mathcal{F}$. Se $\mathbb{P}(B) > 0$, então a **probabilidade condicional** *de A dado B (ou que A vai ocorrer dada a* *informação da ocorrência de B) é dada por*

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

Exemplo 2.3.1 (Experimento Laboratorial)

Vamos ver que a definição 2.3.1 corresponde à idéia freqüentista de probabilidade condicional. Suponha que o seguinte quadro represente a divisão de um grupo de ratos, com relação a um câncer de pele, após uma tratamento e de acordo com sua raça.

Tabela 2.1: Experimento Laboratorial

Raça / Condição	Sadia (S)	Doente (D)	TOTAIS
1	50	30	80
2	10	5	15
3	20	25	45
4	35	25	60
TOTAIS	115	85	200

Dado que um rato, escolhido ao acaso, seja da raça 3, *a probabilidade que esse rato sorteado ainda esteja doente após o tratamento, que denotaremos por D, é obviamente $\frac{25}{45}$.*
Por outro lado, lemos na tabela

$$\mathbb{P}(3 \cap D) = \frac{25}{200} = \frac{1}{8}$$

e

$$\mathbb{P}(3) = \frac{45}{200}.$$

Portanto, usando-se a definição 2.3.1, escrevemos

$$\mathbb{P}(D|3) = \frac{25}{45},$$

que corresponde ao resultado anterior.

2.3.2 Independência

No parágrafo anterior, vimos que a probabilidade $\mathbb{P}(A)$ de um evento A pode ser afetada, e mudar para $\mathbb{P}(A|B)$, com a informação de que um outro evento B da mesma experiência ocorre. Se essa probabilidade não muda, isto é, se $\mathbb{P}(A|B) = \mathbb{P}(A)$, então chamamos os eventos A e B de **independentes**. Essa interpretação só é bem definida se $\mathbb{P}(B) > 0$.

Usando a definição de probabilidade condicional 2.3.1, temos o seguinte.

Definição 2.3.2 (Eventos Independentes)
Dois eventos A e B são chamados de **independentes** *se*

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

Uma família de eventos $\{A_i : i \in I\}$ é chamada de **independente** *se*

$$\mathbb{P}(\cap_{i \in J} A_i) = \prod_{i \in J} \mathbb{P}(A_i),$$

para **toda** *subfamília finita de subconjuntos de índices J de I.*

CUIDADO:
1) Não confunda as noções de independência e incompatibilidade! Quando se diz que dois eventos são independentes, afirma-se que o conhecimento da ocorrência de um não modifica a expectativa (quantificada em sua probabilidade) do outro. Pelo contrário, quando se diz que dois eventos são incompatíveis, afirma-se que não existe a possibilidade de ocorrência simultânea de ambos (quantificada por uma probabilidade nula para a intersecção desses eventos), o que nos diz que o conhecimento da ocorrência de um evento implica na mudança da expectativa de ocorrência do outro, quantificada por uma probabilidade nula que, nos casos não-triviais, é diferente da probabilidade inicial!

2) Se a familia $\{A_i : i \in I\}$ só tem a propriedade de que

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j),\ i \neq j,$$

então os eventos são chamados *independentes dois a dois.*

Esta propriedade **não implica** a independência!!!

Exemplo 2.3.2 *No exemplo 2.3.1, do experimento laboratorial, tivemos*

$$\mathbb{P}(3 \cap D) = \frac{25}{85}, \mathbb{P}(3) = \frac{4}{5}200 \ e\ \mathbb{P}(D) = \frac{8}{5}200.$$

Então,

$$\mathbb{P}(3 \cap D) \neq \mathbb{P}(3)\mathbb{P}(D),$$

isto é, 3 *e* D *não são independentes.*

Exemplo 2.3.3 (Jogo de cara ou coroa)

Lançamos uma moeda duas vezes em seguida. Sejam: A *o evento 'obter cara na primeira vez'; e* B *o evento 'obter coroa na segunda vez'. Supondo que a probabilidade de obter cara com a moeda é* p *e que o primeiro lançamento não influencia o segundo, devemos ter:*

$$\mathbb{P}(B) = 1 - p \ e\ \mathbb{P}(B|A) = 1 - p.$$

Então,

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

isto é, A *e* B *são independentes.*

2.4 Fórmulas das Probabilidades

2.4.1 Fórmula de Poincaré

A Fórmula de Poincaré é também chamada de *Fórmula de Inclusão-Exclusão*. Na Proposição 2.2.1 (c), vimos que $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$. Vamos tentar achar uma fórmula análoga para a união de três ou mais eventos. Primeiramente,

$$\begin{aligned}
\mathbb{P}(A \cup B \cup C) &= \mathbb{P}(A \cup B) + \mathbb{P}(C) - \mathbb{P}((A \cup B) \cap C) \\
&= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) + \mathbb{P}(C) - \mathbb{P}((A \cap C) \cup (B \cap C)) \\
&= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) + \\
&\quad -\mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C).
\end{aligned}$$

Isto nos dá a idéia de uma fórmula mais geral.

Teorema 2.4.1 (Fórmula de Poincaré)

Sejam $A_1, A_2, \ldots, A_n \in \mathcal{F}$. *Então,*

$$\begin{aligned}
\mathbb{P}\left(\cup_{i=1}^n A_i\right) &= \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{1 \le i < j \le n} \mathbb{P}(A_i \cap A_j) + \sum_{1 \le i < j < k \le n} \mathbb{P}(A_i \cap A_j \cap A_k) \\
&\quad - \ldots + (-1)^{n+1} \mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n).
\end{aligned}$$

Demonstração: Por indução em n (veja exercício 2.42). ■

2.4.2 Probabilidades Totais

Definição 2.4.1 (Partição)

Uma seqüência $\{B_1, B_2, ..., B_n\}$ *de subconjuntos (não vazios) de* Ω *é chamada de* **partição** *de* Ω *se*

$$B_1 \cup B_2 \cup \cdots \cup B_n = \Omega$$

e

$$B_i \cap B_j = \emptyset,\ i, j \in \{1, 2, \ldots, n\},\ i \neq j.$$

Teorema 2.4.2 (Probabilidades Totais)
Para quaisquer eventos A e B,

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c).$$

Em geral, seja $\{B_1, B_2, ..., B_n\}$ uma partição de eventos de Ω. Então

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

Demonstração:
Usando $A = (A \cap B) \cup (A \cap B^c)$ e o fato que $(A \cap B)$ e $(A \cap B^c)$ são eventos incompatíveis *(porquê?)*, por definição de $\mathbb{P}$, temos

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

Usando agora que $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$ (veja exercício 2.37), temos o resultado para $n = 2$.

A prova no caso geral se dá por indução (veja os exercícios 2.41 e 2.43).■

Exemplo 2.4.1 *No exemplo 2.3.1, do experimento laboratorial, sendo Ω o conjunto de todos os ratos que participaram do estudo, $\{1, 2, 3, 4\}$ e $\{S, D\}$ são ambas partições de Ω, isto é, $\Omega = 1 \cup 2 \cup 3 \cup 4 = S \cup D$ e cada seqüência é formada por eventos disjuntos. Portanto, se quero calcular $\mathbb{P}(S)$, posso fazê-lo pelas probabilidades totais:*

$$\mathbb{P}(S) \quad = \quad \mathbb{P}(1)\mathbb{P}(S|1) + \cdots + \mathbb{P}(4)\mathbb{P}(S|4)$$

$$
\begin{aligned}
&= \frac{80}{200}\frac{50}{80} + \frac{15}{200}\frac{10}{15} + \frac{45}{200}\frac{20}{45} + \frac{60}{200}\frac{35}{60} \\
&= \frac{115}{200},
\end{aligned}
$$

que é exatamente o valor de $\mathbb{P}(S)$ dado pela tabela.

2.4.3 Fórmula de Bayes

Teorema 2.4.3 (Fórmula de Bayes)
Seja $\{B_1, B_2, ..., B_n\}$ uma partição de eventos de Ω e A um evento de Ω. Então

$$
\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}, \quad i = 1, 2, \ldots, n.
$$

Demonstração: Basta ver que

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

e usar o Teorema das Probabilidades Totais (ver exercício 2.44). ■

Exemplo 2.4.2 *Se no exemplo 2.3.1, sabemos que um determinado rato seja sadio, qual a probabilidade de ele ser raça 1?*

$$
\begin{aligned}
\mathbb{P}(1|S) &= \frac{\mathbb{P}(S|1)\mathbb{P}(1)}{\mathbb{P}(S|1)\mathbb{P}(1) + \cdots + \mathbb{P}(S|4)\mathbb{P}(4)} \\
&= \frac{\frac{50}{80}\frac{80}{200}}{\frac{50}{80}\frac{80}{200} + \frac{10}{15}\frac{15}{200} + \frac{20}{45}\frac{45}{200} + \frac{35}{60}\frac{60}{200}} \\
&= \frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{20} + \frac{1}{10} + \frac{7}{40}}
\end{aligned}
$$

$$= \frac{\frac{10}{40}}{\frac{23}{40}} = \frac{50}{115},$$

que é exatamente a probabilidade obtida diretamente da tabela.

2.5 Exercícios

Exercício 2.1 *Demonstre os resultados expostos nos exemplos 2.2.1, 2.2.2 e 2.2.3.*

Exercício 2.2 *Sejam $A, B \in \mathcal{F}$, $\mathcal{F}$ uma σ-álgebra. Demonstre que os conjuntos $A \cap B$, $A - B$ e $A \triangle B$ também pertencem a $\mathcal{F}$.*

Exercício 2.3 *Sejam $\mathcal{F}$ e $\mathcal{G}$ duas σ-álgebras de Ω.*
(a) Seja $\mathcal{H} = \mathcal{F} \cap \mathcal{G}$ a coleção dos subconjuntos de Ω que pertencem a ambas as σ-álgebras $\mathcal{F}$ e $\mathcal{G}$. Mostre que $\mathcal{H}$ é uma σ-álgebra.
(b) Demonstre que se $A_1, A_2, \ldots$ estão em $\mathcal{F}$ então $\cup_i A_i \in \mathcal{F} \cup \mathcal{G}$.
(c) Mostre que $\mathcal{F} \cup \mathcal{G}$ não é em geral uma σ-álgebra (Dica: Tome $\Omega = \{a, b, c\}$).

Exercício 2.4 *Demonstre a Proposição 2.2.1.*

Exercício 2.5 *Complete as passagens marcadas com asteriscos na demonstração da Proposição 2.2.2.*

Exercício 2.6 *Sejam A, B dois eventos com probabilidades $\mathbb{P}(A) = 3/4$ e $\mathbb{P}(B) = 1/3$. Demonstre que $1/12 \leq \mathbb{P}(A \cup B) \leq 1/3$. Achar as cotas para $\mathbb{P}(A \cap B)$.*

Exercício 2.7 *Suponha que 40% da população dum estado tenha o tipo de sangue O. Se quatro pessoas são sorteadas ao acaso neste estado, achar $\mathbb{P}$('pelo menos uma pessoa é do tipo O').*

Exercício 2.8 *Há duas rotas entre Campinas e São Paulo e duas entre São Paulo e Santos. Cada uma delas tem uma probabilidade p de estar engarrafada, independentemente de todas as outras.*

1) Qual é a probabilidade de viajar de Campinas até Santos sem parada?

2) Qual é a probabilidade de haver (pelo menos) uma rota sem engarrafamento entre Campinas e São Paulo dado que há engarrafamentos entre Campinas e Santos.

3) Supondo que há mais uma rota que ligue diretamente Campinas e Santos, com mesma probabilidade p de engarrafamento independentemente de todas as outras, achar a probabilidade condicional do item 2).

Exercício 2.9 Passeio aleatório simétrico (ou *Ruína do jogador*)

Um homem tenta economizar para comprar uma Ferrari nova que custa N unidades de moeda. Ele começa com k $(0 < k < N)$ unidades e tenta ganhar o resto no seguinte jogo com seu banco. Ele lança uma moeda não viciada: se o resultado for cara, então o banco dá para ele um unidade monetária mas, se o resultado for coroa, ele paga uma unidade para o banco. Isto é repetido até que um dos dois seguintes eventos ocorra: ele perde todo o seu dinheiro ou consegue chegar ao preço da Ferrari. Qual é a probabilidade de perder tudo?

Exercício 2.10 *Suponha que todos os conjuntos a seguir sejam elementos de alguma σ-álgebra $\mathcal{F}$ de um certo espaço dos resultados Ω.*

(a) Mostre, sem utilizar diretamente a Fórmula de Poincaré 2.4.1, que:

$$\begin{aligned}
\mathbb{P}(A \cup B \cup C \cup D) \quad = \quad & \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) + \mathbb{P}(D) - \mathbb{P}(A \cap B) \\
& -\mathbb{P}(A \cap C) - \mathbb{P}(A \cap D) - \mathbb{P}(B \cap C) - \mathbb{P}(B \cap D) \\
& -\mathbb{P}(C \cap D) + \mathbb{P}(A \cap B \cap C) + \mathbb{P}(A \cap B \cap D) + \\
& +\mathbb{P}(A \cap C \cap D) + \mathbb{P}(B \cap C \cap D) - \mathbb{P}(A \cap B \cap C \cap D)\,; \ \ e
\end{aligned}$$

(b) Faça um diagrama que reflita o resultado em (a);

Exercício 2.11 *Seja Ω o espaço dos resultados de um experimento aleatório e $\mathbb{P}$ uma probabilidade em Ω. Demonstre:*

(a) Sejam $n \in \mathbb{N}$, $A_1,\ A_2, \ldots,\ A_n$ subconjuntos de Ω, sendo $A_i \cap A_j = \emptyset$ se $i \neq j$. Então,

$$\mathbb{P}\left(\bigcup_{j=1}^{n} A_j\right) = \sum_{j=1}^{n} \mathbb{P}\left(A_j\right) \quad (o)$$

((o) quer dizer que $\mathbb{P}$ é aditiva);

(b)Sejam $n \in \mathbb{N}$, $A_1,\ A_2, \ldots,\ A_n$ subconjuntos de Ω. Então,

$$\mathbb{P}\left(\bigcup_{j=1}^{n} A_j\right) \leq \sum_{j=1}^{n} \mathbb{P}\left(A_j\right) \quad e,$$

portanto,

$$A = \bigcup_{j=1}^{n} A_j \Rightarrow \mathbb{P}(A) \leq \sum_{j=1}^{n} \mathbb{P}\left(A_j\right),$$

dizendo-se que $\mathbb{P}$ é subaditiva;

(c) Sejam $n \in \mathbb{N}$ e $A_1,\ A_2, \ldots,\ A_n$ elementos de $\mathcal{F}$. Portanto,

(i) $\mathbb{P}\left(A_j\right) = 0,\ j = 1, 2, \ldots, n \Rightarrow \mathbb{P}\left(\bigcup_{j=1}^{n} A_j\right) = 0$; e

(ii) $\mathbb{P}\left(A_j\right) = 1,\ j = 1, 2, \ldots, n \Rightarrow \mathbb{P}\left(\bigcap_{j=1}^{n} A_j\right) = 1$.

(d) Mostre que resultados análogos aos do item anterior valem para uniões e intersecções enumeráveis.

Nota Importante: $\mathbb{P}\left(A_j\right) = 0$ *não implica em que $A = \emptyset$ como tampouco $\mathbb{P}\left(A_j\right) = 1$ implica em que $A = \Omega$.*

Exercício 2.12 *Seja $\Omega = \{1, 2, 3, 4, 5, 6\}$ o espaço dos resultados obtidos ao jogar-se um dado e observar-se a face de cima. Mostre que existe uma única* *função de probabilidade* $\mathbb{P}$ *em* Ω *tal que:*

$$\begin{array}{lll} \mathbb{P}(\{1\}) = 0,20 & \mathbb{P}(\{2\}) = 0,30 & \mathbb{P}(\{3\}) = 0,10 \\ \mathbb{P}(\{4\}) = 0,15 & \mathbb{P}(\{5\}) = 0,11 & \mathbb{P}(\{6\}) = 0,14 \end{array}$$

Exercício 2.13 *Tome* Ω *e* $\mathbb{P}$ *como no Exercício 2.12. Suponha que exista uma função de probabilidade* $\mathbb{Q}$ *sobre a qual sabe-se apenas que:*

$$\mathbb{Q}(\{1,2,3\}) = 0,60 \quad \mathbb{Q}(\{4\}) = 0,15 \quad e \quad \mathbb{Q}(\{5,6\}) = 0,25. \tag{2.1}$$

(a) Mostre que é possível calcular as $\mathbb{Q}$*-probabilidades dos seguintes conjuntos:* $\emptyset$, $\{1,2,3\}$, $\{4\}$, $\{5,6\}$, $\{1,2,3,4\}$, $\{1,2,3,5,6\}$, $\{4,5,6\}$ *e* Ω*;*

(b) Mostre que, a partir do conjunto de informações (2.1), é impossível calcular $\mathbb{Q}(\{5\})$ *(e, portanto,* $\mathbb{Q}(\{6\})$*)*

Dica: *Defina uma outra probabilidade* $\mathbb{Q}'$ *em* Ω *por:*

$$\begin{array}{lll} \mathbb{Q}'(\{1\}) = 0,20 & \mathbb{Q}'(\{2\}) = 0,30 & \mathbb{Q}'(\{3\}) = 0,10 \\ \mathbb{Q}'(\{4\}) = 0,15 & \mathbb{Q}'(\{5\}) = 0,18 & \mathbb{Q}'(\{6\}) = 0,07. \end{array}$$

Verifique que tanto $\mathbb{Q}'$ *como* $\mathbb{P}$ *produzem 2.1;*

(c) Mostre que, a partir de (2.1), também é impossível de se calcularem as $\mathbb{Q}$*-probabilidades dos conjuntos:* $\{1\}$, $\{2\}$, $\{3\}$, $\{1,2\}$, $\{1,3\}$ *e* $\{2,3\}$*;*

(d) Utilize os resultados dos itens anteriores para concluir que é impossível calcular as $\mathbb{Q}$*-probabilidades de qualquer conjunto não incluso no item (a);*

(e) Prove que há um número infinito de probabilidades, dadas por $\mathbb{Q} \equiv \mathbb{Q}_{s,t,r}$ *tais que*

$$\begin{array}{lll} \mathbb{Q}(\{1\}) = s & \mathbb{Q}(\{2\}) = t & \mathbb{Q}(\{3\}) = 0,60 - s - t \\ \mathbb{Q}(\{4\}) = 0,15 & \mathbb{Q}(\{5\}) = r & \mathbb{Q}(\{6\}) = 0,25 - r, \end{array}$$

onde $s \geq 0$, $t \geq 0$, $s + t \leq 0,60$ *e* $0 \leq r \leq 0,25$*, que satisfazem o conjunto de condições (2.1).*

Exercício 2.14 *Seja Ω como no exercício 2.12. Defina $\mathbb{S}$ como uma função de probabilidade em Ω tal que:*

$$\mathbb{S}(\{1,2,3,4\}) = 0,75 \quad e \quad \mathbb{S}(\{4,5,6\}) = 0,40. \tag{2.2}$$

(a) Calcule as $\mathbb{S}$-probabilidades dos seguintes conjuntos: $\emptyset$, $\{1,2,3\}$, $\{4\}$, $\{5,6\}$, $\{1,2,3,4\}$, $\{1,2,3,5,6\}$, $\{4,5,6\}$ e Ω; e

(b) Mostre que os conjuntos em (a) são os únicos cujas $\mathbb{S}$-probabilidades podem ser calculadas a partir das condições (2.2).

Exercício 2.15 *Seja Ω como no exercício 2.12 e $\mathbb{T}$ uma função de probabilidade em Ω tal que:*

$$\mathbb{T}(\{1,2,3,4\}) = 0,75 \quad e \quad \mathbb{T}(\{3,4,5,6\}) = 0,50. \tag{2.3}$$

Mostre que os únicos conjuntos cujas $\mathbb{T}$-probabilidades podem ser calculadas a partir das condições (2.3) são os seguintes: $\emptyset$, $\{1,2\}$, $\{3,4\}$, $\{5,6\}$, $\{1,2,3,4\}$, $\{1,2,5,6\}$, $\{3,4,5,6\}$ e Ω.

Exercício 2.16 *Seja $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ o espaço de resultados de um experimento aleatório $\mathcal{E}$, com cardinalidade n. Seja $(p_1, p_2, \ldots, p_n) \in \mathbb{R}^n$ tal que $p_j \geq 0$, $1 \leq j \leq n$ e $\sum_{j=1}^{n} p_j = 1$. Prove que existe uma única função de probabilidade em Ω tal que $\mathcal{P}(\{\omega_j\}) = p_j$, para $1 \leq j \leq n$. Nesse caso, $\forall A \in \mathbb{P}(\Omega)$, vale que:*

$$\mathbb{P}(A) = \sum_{j:\omega_j \in A} p_j.$$

Exercício 2.17 *Suponha a seguinte definição.*

> **Definição:** *Seja $\Omega \neq \emptyset$. Uma família $\mathcal{A} \subset \mathcal{P}(\Omega)$ é uma álgebra se satisfaz algum dos seguintes conjuntos equivalentes de condições:*

$$\begin{array}{ll} (1) & \Omega \in \mathcal{A} \\ (2) & A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A} \\ (3) & A \in \mathcal{A},\ B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A} \end{array} \tag{2.4}$$

$$\begin{array}{ll} (1) & \Omega \in \mathcal{A} \\ (2) & A \in \mathcal{A},\ B \in \mathcal{A} \Rightarrow A - B \in \mathcal{A} \\ (3) & A \in \mathcal{A},\ B \in \mathcal{A},\ A \cap B = \emptyset \Rightarrow A \cup B \in \mathcal{A} \end{array} \tag{2.5}$$

$$\begin{array}{ll} (1) & \Omega \in \mathcal{A} \\ (2) & A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A} \\ (3) & A_j \in \mathcal{A},\ 1 \leq j \leq n \Rightarrow \cup_{j=1}^n A_j \in \mathcal{A} \end{array} \tag{2.6}$$

$$\begin{array}{ll} (1) & \Omega \in \mathcal{A} \\ (2) & A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A} \\ (3) & A_j \in \mathcal{A},\ 1 \leq j \leq n \Rightarrow \cap_{j=1}^n A_j \in \mathcal{A} \end{array} \tag{2.7}$$

$$\begin{array}{ll} (1) & \Omega \in \mathcal{A},\ \emptyset \in \mathcal{A} \\ (2) & A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A} \\ (3) & A \in \mathcal{A},\ B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}\ e \\ & A_j \in \mathcal{A}, 1 \leq j \leq n \Rightarrow \cup_{j=1}^n A_j \in \mathcal{A}\ e \cap_{j=1}^n A_j \in \mathcal{A} \end{array} \tag{2.8}$$

Mostre que as condições (2.4)-(2.8) da definição de uma álgebra são equi-

valentes. Sugere-se o seguinte esquema de prova:

$$(2.8) \Leftrightarrow (2.5) \Leftrightarrow (2.4) \Leftrightarrow (2.6) \Leftrightarrow (2.7).$$

Exercício 2.18 *Verifique que são álgebras as famílias de conjuntos cujas probabilidades são possíveis calcular a partir das condições (2.1), (2.2) e (2.3), nos exercícios 2.13, 2.14 e 2.15.*

Exercício 2.19 *Prove que toda σ-álgebra $\mathcal{A}$ (definição 2.2.1) é também uma álgebra.*

Exercício 2.20 *Seja $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$ $(\omega_i \neq \omega_j)$ o espaço de resultados de um experimento aleatório $\mathcal{E}$. Seja $\{p_1, p_2, \ldots, p_n, \ldots\}$ uma seqüência de números reais não-negativos tais que $\sum_{j=1}^{\infty} p_j = 1$. Prove que a função de conjunto definida por:*

$$\mathbb{P}(A) = \sum_{j:\omega_j \in A} p_j$$

é uma função de probabilidade em Ω tal que $\mathbb{P}(\{\omega_j\}) = p_j$, $j = 1, 2, \ldots$.

Exercício 2.21 *Enuncie e prove a versão do item (c) do exercício 2.11, para uma coleção enumerável de elementos $\{A_j\}$, $j = 1, 2, \ldots, \ldots$ de $\mathcal{A}$.*

Exercício 2.22

(a) Suponha que um experimento $\mathcal{E}$ seja executado n vezes. Para cada evento A dos espaço de resultados, seja $n(A)$ o número de vezes que o evento A ocorreu. Defina $f(A) = n(A)/n$. Mostre que $f(\cdot)$ satisfaz a definição axiomática de Probabilidade;

(b) Considere um experimento cujo espaço de resultados seja enumerável (e infinito). Argumente que equiprobabilidade de todos os pontos está fora de questão;

(c) Seja $(\Omega, \mathcal{A}, \mathbb{P})$ um espaço de probabilidade. Dados A e B elementos de $\mathcal{A}$, mostre que:

$$\mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1.$$

Descreva a utilidade prática deste resultado (conhecido como Desigualdade de Bonferroni) e os casos em que ele a perde completamente; e

(d) Mostre que a desigualdade de Bonferroni também é válida para n eventos:

$$\mathbb{P}\left(\bigcap_{j=1}^{n} A_j\right) \geq \sum_{j=1}^{n} \mathbb{P}\left(A_j\right) - (n-1)$$

e faça comentários análogos aos do item anterior.

Exercício 2.23 *Colocam-se cinco bolas vermelhas, seis azuis e oito verdes em uma urna. Suponha que não seja possível distinguir entre elas antes de retiradas e que todas as dezenove bolas tenham chances iguais de sorteio. Se um conjunto de três bolas for aleatoriamente escolhido, qual a probabilidade de que:*

(a) todas as bolas tenham a mesma cor; e

(b) cada bola seja de um cor diferente?

Dica: *Calcule estas probabilidades sob a hipótese de que o processo de seleção seja feito com reposição (cada bola é retirada, sua cor é anotada e ela é colocada de volta na urna).*

Exercício 2.24 *Uma professora distribui para sua classe um conjunto de dez problemas, sendo que o exame final consistirá de uma seleção aleatória de cinco desses problemas. Se um aluno tiver resolvido corretamente sete problemas, qual a probabilidade de que:*

(a) ele consiga nota máxima no exame; e

(b) ele consiga acertar pelo menos quatro problemas, no exame?

Nota: Nos exercícios 2.25-2.40, a não ser que explicitamente definido de outra maneira, considere a existência de um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$. Sempre que possível, escreva todos os eventos de interesse como função de eventos mais simples, cujas probabilidades tenham um cálculo mais fácil. Interprete todas as expressões matemáticas em termos de ocorrência de eventos.

Exercício 2.25 *Uma urna contém duas bolas brancas e três bolas vermelhas. Duas bolas são extraídas ao acaso, em seqüência, e suas cores são registradas. Definam-se:*

$$\begin{aligned} B_i &= \{i\text{-ésima bola retirada é branca}\}, \quad i = 1,2 \\ V_j &= \{j\text{-ésima bola retirada é vermelha}\}, \quad j = 1,2 \end{aligned}$$

Calcule $\mathbb{P}(B_2|V_1)$, $\mathbb{P}(V_2|V_1)$ e a probabilidade de cada uma das possíveis seqüências nas seguintes situações:

(a) bolas extraídas com reposição; e
(b) bolas extraídas sem reposição.

Exercício 2.26 *Tome $\mathbb{P}(\cdot|A)$ como a definição usual de probabilidade condicional, dada por:*

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

quando $A, B \in \mathcal{F}$, $\mathbb{P}(A) > 0$.

Prove que $\mathbb{P}(\cdot|A)$ é uma função de probabilidade em seu sentido axiomático.

Exercício 2.27 *Demonstre que $\mathbb{P}(A|B) = \mathbb{P}(B|A)\mathbb{P}(A)/\mathbb{P}(B)$ se $\mathbb{P}(A)\mathbb{P}(B) \neq 0$. Mostre que, se $\mathbb{P}(A|B) > \mathbb{P}(A)$, então $\mathbb{P}(B|A) > \mathbb{P}(B)$.*

Exercício 2.28 *Suponha que $\Omega = \{aaa, abc, acb, bbb, bac, bca, ccc, cab, cba\}$ e cada um desses nove eventos elementares ocorra com probabilidade $\frac{1}{9}$. Seja, para $k = 1, 2, 3$, A_k = 'a k-ésima letra é a'. Mostre que A_1, A_2, A_3 são independentes dois a dois, mas não independentes.*

Exercício 2.29
(a) Sejam A e B dois eventos independentes; mostre que A^c e B também são independentes, assim como o são A^c e B^c;

(b) Mostre que A *e* A^c *são independentes se e somente se* $P(A) = 0$ *ou* $P(A) = 1$*; e*

(c) Num domingo, Ponte Preta e Guarani disputam jogos de futebol eliminatórios, num campeonato cujas regras não admitem empates. Sejam:

$$A = \{Ponte\ ganha\} \text{ e } B = \{Guarani\ ganha\}.$$

Observe que:

(i) Se os times jogarem um contra o outro, então A *e* B *são complementares; e*

(ii) Se os times não estiverem em disputa direta, A *e* B *podem ser considerados independentes.*

Defina os espaços de resultados adequados para as situações (i) e (ii).

Exercício 2.30 *Uma moeda viciada é lançada com repetição. A cada lançamento, há uma probabilidade* p *de obter-se 'cara'. Seja* p_n *a probabilidade de que um número par de caras ocorra depois de* n *lançamentos (0 é par). Prove que* $p_0 = 1$ *e* $p_n = p(1 - p_{n-1}) + (1 - p)p_{n-1}$*, se* $n \geq 1$*. Resolver essa equação em diferenças.*

Exercício 2.31 *Joga-se uma moeda cinco vezes. A probabilidade de sair cara é 0,6. Definem-se:*

$$\begin{aligned} H_i &= \{i\text{-ésimo lançamento é cara}\}, \quad i = 1, 2, 3, 4, 5; \text{ e} \\ T_j &= \{j\text{-ésimo lançamento é coroa}\}, \quad j = 1, 2, 3, 4, 5. \end{aligned}$$

Suponha que os lançamentos sejam feitos de maneira independente e calcule:

(a) $\mathbb{P}(H_1 H_2 T_3 H_4 T_4)$*;*

(b) Probabilidade de se obterem três caras;

Generalize (a) e (b) para qualquer probabilidade de cara em um lançamento $0 \leq p \leq 1$.

Exercício 2.32 *Achar na tabela 2.1, do exemplo 2.3.1, todas as probabilidades possíveis:* $\mathbb{P}(1)$, $\mathbb{P}(1 \cap S)$, $\mathbb{P}(D|1)$ *etc.*

Exercício 2.33 *Sejam* A *e* B *elementos de* $\mathcal{F}$ *t.q.* $0 < \mathbb{P}(A), \mathbb{P}(B) < 1$. *Demonstrar que* $\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B)$ *mas* $\mathbb{P}(A|B^c) \neq 1 - \mathbb{P}(A|B)$.

Exercício 2.34 *O que você acha da seguinte* **prova**, *dada por Lewis Carroll, de que, dentro duma urna, não pode haver duas bolas da mesma cor?*

> Suponha que uma urna tenha duas bolas, cada uma delas podendo ser branca ou preta. As distribuições possíveis são $\{B, B\}$, $\{B, P\}$ ou $\{P, P\}$. Por simetria, temos $\mathbb{P}(BB) = \mathbb{P}(BP) = \mathbb{P}(PP) = \frac{1}{3}$.
>
> Adiciona-se à urna uma bola preta; então, $\mathbb{P}(PBB) = \mathbb{P}(PBP) = \mathbb{P}(PPP) = \frac{1}{3}$. Agora, sorteamos uma bola ao acaso; a probabilidade de que a bola seja preta é (usando as probabilidade condicionais) $\frac{1}{3}.\frac{1}{3} + \frac{2}{3}.\frac{1}{3} + 1.\frac{1}{3} = \frac{2}{3}$.
>
> Porém, se temos uma probabilidade de $\frac{2}{3}$ de sortear uma bola preta de uma urna com três bolas, usando o argumento de simetria, devemos ter duas bolas pretas e uma branca, o que é a mesma coisa que dizer que, no início do experimento, a urna tinha uma bola branca e outra preta!

(Dica: Qual é o experimento?)

Exercício 2.35 *Sejam* A, B *e* C *eventos tais que* $\mathbb{P}(A) = 1/3$, $\mathbb{P}(B) = 1/4$, $\mathbb{P}(C) = 1/6$, $\mathbb{P}(A \cap B) = 1/6$, $\mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = 1/10$ *e* $\mathbb{P}(A \cap B \cap C) = 1/12$. *Determine as probabilidades da ocorrência de:*

(i) exatamente um dos eventos A, B *e* C*;*
(ii) exatamente dois dos eventos A, B *e* C*;*
(iii) pelo menos dois desses eventos;
(iv) no máximo dois desses eventos; e
(v) no máximo um desses eventos.

Exercício 2.36 (Pôquer com Dados)

O pôquer com dados é jogado da seguinte forma. Cada jogador lança simultaneamente um conjunto de cinco dados de seis faces e os resultados são classificados em:

Nada (A_1)*: todas as faces superiores diferentes, sem formar uma seqüência (vide* A_6*) como, por exemplo,* 23146*;*
Trinca (A_2)*: três faces superiores iguais e as outras duas diferentes entre si e daquelas como, por exemplo,* 33312*;*
Par (A_3)*: duas faces superiores iguais e as outras diferentes destas e entre si como, por exemplo,* 22134*;*
Quíntupla (A_4)*: todas as faces superiores iguais, como, por exemplo,* 55555*;*
Dois Pares (A_5)*: dois conjuntos de duas faces superiores iguais e a restante diferente de ambas como, por exemplo,* 66114*;*
Seqüência (A_6) *: faces superiores formam uma seqüência ininterrupta como, por exemplo,* 12345*;*
Pôquer (A_7)*: quatro faces superiores iguais e a restante diferentes daquelas como, por exemplo,* 11151*; e*
Full Hand (A_8)*: dois conjuntos de faces superiores iguais entre si e diferentes da outra, sendo um de tamanho três e o outro de tamanho dois como, por exemplo,* 65566*.*

Calcule a probabilidade de A_i*,* $i = 1, 2, \ldots, 8$ *e, de acordo com esses valores, proponha e justifique uma hierarquia entre as diversas combinações possíveis.*

Exercício 2.37 *Sejam* A *e* B *eventos de um determinado experimento aleatório* $\mathcal{E}$ *tais que* $\mathbb{P}(A)\mathbb{P}(B) > 0$*. Verifique que:*

(i) $\mathbb{P}(A|B)\,\mathbb{P}(B) = \mathbb{P}(A \cap B) = \mathbb{P}(B|A)\,\mathbb{P}(A)$ *e, conseqüentemente*

(ii) $\mathbb{P}(B|A) = \dfrac{\mathbb{P}(A|B)\,\mathbb{P}(B)}{\mathbb{P}(A)}$.

Exercício 2.38

(a) Sejam A e B eventos independentes tais que $\mathbb{P}(A) = 1/3$ e $\mathbb{P}(B) = 1/2$. Calcule: $\mathbb{P}(A \cup B)$, $\mathbb{P}(A^c \cup B^c)$ e $\mathbb{P}(A^c \cap B)$.

(b) Sejam A e B eventos independentes tais que $\mathbb{P}(A) = 1/4$ e $\mathbb{P}(A \cup B) = 1/3$. Calcule $\mathbb{P}(B)$.

(c) Suponha que A, B e C sejam eventos independentes. Prove que também o são: (i) A e B^c e (ii) A^c, B e C^c.

Exercício 2.39 (Problema dos Encontros de Montmort)

Forma-se uma permutação simples dos números $\{1, 2, \ldots, n\}$. Caso a i-ésima posição seja ocupada pelo número i, diz-se haver um encontro na posição i. Calcule a probabilidade dos seguintes eventos:

(i) haver exatamente k $(k \leq n)$ encontros;

(ii) haver um encontro na i-ésima posição dado que há exatamente k encontros;

(iii) haver um encontro na posição i e não ocorrer um na posição j $(i \neq j)$; e

(iv) haver um encontro na posição i dado que não há um na posição j $(i \neq j)$.

Exercício 2.40 *Uma pessoa com um molho de n chaves tenta abrir uma porta. Apenas uma das chaves consegue abrí-la. Calcule a probabilidade de que ela só consiga abrir a porta na k-ésima tentativa, supondo:*

(a) que, após cada tentativa mal sucedida, ela descarta a chave usada; ou

(b) que nenhuma chave seja descartada.

Exercício 2.41 *Demonstre o seguinte lema:*

> **Lema 1** *Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e suponha que uma coleção de elementos de $\mathcal{F}$, $\{A_1, A_2, \ldots, A_n\}$, seja uma partição de Ω (isto é, $\Omega = \bigcup_{i=1}^{n} A_i$ e $A_i \cap A_j = \emptyset$, se $i \neq j$). Então,*

$\forall B \in \mathcal{F}$,

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

Exercício 2.42 *Demonstre a Fórmula de Poincaré (Teorema 2.4.1).*

Exercício 2.43 *Demonstre o seguinte lema:*

> **Lema 2: Fórmula da Probabilidades Totais**: *Seja* $(\Omega, \mathcal{F}, \mathbb{P})$ *um espaço de probabilidade e suponha que a coleção de elementos de* $\mathcal{F}$, $\{A_1, A_2, \ldots, A_n\}$, *seja uma partição de* Ω. *Suponha, ainda que* $\mathbb{P}(A_i) > 0$, $i = 1, 2, \ldots, n$. *Então,* $\forall B \in \mathcal{F}$,
>
> $$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_i)\, \mathbb{P}(B|A_i).$$

Como reduzir as n desigualdades introduzidas no lema 2 para apenas uma?

Dica: *Use a relação entre* $\mathbb{P}(A)$ *e* $\mathbb{P}(B)$ *quando* $A, B \in \mathcal{F}$ *e* $A \subset B$.

Exercício 2.44 *Complete a demonstração de*

> **Lema 3: Fórmula de Bayes** *Seja* $(\Omega, \mathcal{F}, \mathbb{P})$ *um espaço de probabilidade e suponha que uma coleção de elementos de* $\mathcal{F}$, $\{A_1, A_2, \ldots, A_n\}$, *seja uma partição de* Ω. *Suponha, ainda que* $\mathbb{P}(B) > 0$ *e* $\mathbb{P}(A_i) > 0$, $i = 1, 2, \ldots, n$. *Então,*
>
> $$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(B|A_j)\, \mathbb{P}(A_j)}{\sum_{i=1}^{n} \mathbb{P}(A_i)\, \mathbb{P}(B|A_i)}.$$

Capítulo 3

Variáveis Aleatórias

3.1 Introdução

Seja $\mathcal{E}$ um experimento aleatório e Ω seu espaço de resultados. Até o momento, temos estudado $\mathcal{E}$ a partir do próprio Ω. Entretanto, em muitas situações, o resultado observado do experimento $\omega \in \Omega$ não intervém de forma explícita, pois podemos estar interessados apenas em alguma característica de ω e não necessariamente em sua individualização.

Uma forma conveniente de representar uma característica de interesse é mediante a definição de uma função X com domínio em Ω. Estudaremos aqui características numéricas das observações, ou seja, as dadas através de funções $X\colon \Omega \to \mathbb{R}$. Estas funções são chamadas de variáveis aleatórias e sua definição formal será vista um pouco mais à frente.

A seguir, veremos alguns exemplos para ilustrar os aspectos que acabamos de abordar.

Exemplo 3.1.1 *Todos os* 86 *alunos da sétima série de uma escola de primeiro grau são submetidos a um teste para avaliar seus conhecimentos de matemática. Há apenas duas notas: aprovado* (A) *ou reprovado* (R)*. O espaço dos resultados pode ser representado através de uma lista com o nome e a nota de cada um dos alunos.*

Equivalentemente, na posse de uma lista em ordem alfabética, um elemento do espaço dos resultados será individualizado através de uma seqüência formada por 86 *símbolos* A *ou* R*, onde uma letra* A *ou* R *na* j*-ésima posição significará que o* j*-ésimo aluno da* *lista foi aprovado ou reprovado, respectivamente* $(1 \leq j \leq 86)$.

Entretanto, para se ter uma idéia do desempenho geral da escola no teste, poderá ser necessário apenas considerar o número de alunos aprovados (ou número de alunos aprovados, dos 86 *inscritos), sem se importar com os desempenhos individuais.*

Exemplo 3.1.2 *Um fabricante de calçados masculinos para adultos pretende atender ao consumo local de uma pequena cidade, ou seja,* $\Omega = \{$*os homens adultos que moram na cidade*$\}$. *É obvio que, para dimensionar sua produção, ele não precisa conhecer os nomes nem outras características pessoais dos seus clientes potenciais, com exceção do tamanho do sapato.*

Em resumo, podemos dizer que o fabricante deve possuir algum conhecimento a respeito da variável X *definida por:* $X(\omega) =$*'número de sapato utilizado pela pessoa* ω*', para cada* $\omega \in \Omega$.

Mas qual é o conhecimento necessário a respeito de X*? Por exemplo, o fabricante deve saber que fabricando sapatos desde o número* 32 *até o* 48, *ele cobre todas as possibilidades e também que existem mais clientes potenciais para os números* 40 *e* 41 *do que para* 32 *ou* 48.

Com maior generalidade, é necessário que o fabricante conheça as probabilidades dos conjuntos $\{\omega / X(\omega) = x\}$, *denotados por* $[X = x]$ *para* $32 \leq x \leq 48$.

Exemplo 3.1.3 *Graciliano é aposentado da SANASA e uma vez ao mês costuma se encontrar com seus antigos colegas no Largo do Carmo. Ele pega o ônibus que vai para o centro de Campinas, no terminal de Barão Geraldo. Nos horários em que ele costuma viajar, há um intervalo de exatamente cinco minutos entre duas saídas consecutivas. Graciliano é muito calmo e nem tenta acertar a chegada no ponto para minimizar a espera. Ele tampouco gosta de correr, de forma que, mesmo que o ônibus esteja na plataforma* *prestes a partir, ele continuará andando no seu ritmo. O tempo X que Graciliano fica no ponto aguardando o ônibus depende do horário de chegada ao ponto que, por sua vez depende de muitos fatores: hora em que Graciliano acordou, tempo demorado no café da manhã, tempo demorado para tomar banho etc. Assim sendo, desistiremos de considerar o possível conjunto de situações ω que determina o tempo final X.*

Alternativamente, tentaremos trabalhar diretamente sobre X: com base na informação disponível a respeito dos costumes de Graciliano, formularemos algumas hipóteses que simplifiquem o problema e que nos permitam resolvê-lo. [1]

Em primeiro lugar, é evidente que $0 \leq X < 5$. O mínimo que ele espera é claramente 0. Por que $X < 5$ e não $X \leq 5$? O que significa $X = 5$? Uma situação em que X poderia ser exatamente igual a 5 seria caso Graciliano chegasse na hora certa e simultaneamente dormisse no ponto.

Parece razoável pensar que, se $0 \leq a < b < 5$, a probabilidade de $\{\omega \ / \ a \leq X(\omega) < b\}$ seja proporcional ao comprimento de $[a, b)$, ou seja, $(b - a)/5$.[2]

Veja que se $a = 0$ e $b = 5$, teremos probabilidade para $[0 < X < 5]$ igual a 1, o que é correto. Como conseqüência, se $A \cap [0 < X < 5]\emptyset$, a probabilidade de $[X \in A]$ é igual a 0.

[1] Qualquer tentativa de estudo de um problema é baseada apenas parcialmente nos *dados*, pois outra parte importante é formada por hipóteses ou suposições a respeito da situação sob análise. Dado que, na prática, é impossível considerar todos os aspectos de uma determinada realidade, o uso dessas hipóteses se torna necessário para permitir a modelagem, ou abordagem matemática, do problema. Estas suposições são racionalizações matemáticas de conhecimentos imprecisos ou até de palpites, não sendo, portanto, estritamente verdadeiras. A arte na escolha de tais hipóteses reside em que as restrições matemáticas por elas impostas não sejam conflitantes com aspectos relevantes da realidade.

[2] Uma observação deve ser feita sobre o problema de notação. Serão utilizados como equivalentes $\{\omega/a < X(\omega) < b\}$, $X^{-1}((a, b))$ ou $[a < X < b]$. Em geral, se $A \subset \mathbb{R}$, o conjunto $\{\omega/X(\omega) \in A\}$ será denotado por $[X \in A]$ ou também por $X^{-1}(A)$ e denominado imagem inversa de A (referente à função X). Se $A = \{a\}$, utilizaremos também $[X = a]$.

Outro fato interessante é o de como seria a probabilidade de $[X = c]$, *para* $0 \le c < 5$*? Observe que o conjunto* $[X = c]$ *pode ser pensado como*

$$[X = c] = \lim_{n\to\infty}[\max(0, c - 1/n) \le X \le \min(5, c + 1/n)],$$

que é decrescente e cujas probabilidades são menores do que ou iguais a $2n^{-1}$.

Portanto, é natural que, se a probabilidade estiver definida de forma consistente, $[X = c]$ *tenha probabilidade* 0.

Isto poderia, à primeira vista, parecer contraditório com os fatos $[0 \le X < 5] = \cup_{0\le c<5}[X = c]$ *e* $\mathbb{P}([0 \le X < 5]) = 1$. *Note, no entanto, que o conjunto* $[0 \le X < 5]$ *é união de uma família não-enumerável de conjuntos disjuntos e, portanto, aqui não é imediata a aplicação de propriedades do tipo aditividade ou* σ-*aditividade. Observe também que você já enfrentou este tipo de contradições aparentes: por exemplo, o retângulo* $[0 \le x \le 1] \times [0 \le y \le 1] \in \mathbb{R}^2$ *pode ser escrito como união disjunta da família não-enumerável dos segmentos* $[0 \le x \le 1] \times \{y\}$ *onde* $0 \le y \le 1$, *sendo nula a área de cada um dos segmentos e igual a 1 a área total do retângulo.*

Também, como conseqüência de que a probabilidade de $[X = c]$ *é igual a* 0 *se* $0 \le c < 5$, *tem-se que: se* $0 \le a < b < 5$, *ou seja, as probabilidades dos conjuntos* $[a \le X \le b]$, $[a \le X < b]$, $[a < X \le b]$ *e* $[a < X < b]$ *são todas* $(b - a)/5$.

Observe que, diferentemente do exemplo 3.1.2,onde a probabilidade estava concentrada em um número finito de valores de X, *aqui a probabilidade de qualquer conjunto* $[X = c]$ *é nula e, conseqüentemente, o interesse está nas probabilidades de outros conjuntos como, por exemplo, as imagens inversas de intervalos, tal como vimos no parágrafo precedente.*

Exemplo 3.1.4 *Um indivíduo atira num alvo a uma distância de vinte metros. O alvo consiste de um círculo de madeira de 2 m de raio com um ponto C marcado no seu centro. Assuma que o atirador sempre acerta o plano Π determinado pelo círculo e que, mediante algum dispositivo, seja possível detectar o ponto $\mathcal{P}$ em que o disparo atravessa Π. Neste caso, o espaço dos resultados pode ser representado pelo conjunto $\Omega = \{\mathcal{P}/\mathcal{P} \in \Pi\}$. Define-se a variável X = distância ($\mathcal{P}$,C) e é claro que Imagem(X) = $\mathbb{R}^+ = \{r \in \mathbb{R}/r \geq 0\}$.*

Este exemplo tem alguns aspectos semelhantes ao exemplo 3.1.3, no sentido de que a probabilidade de qualquer subconjunto A de Ω deve ter alguma relação com a área de A.

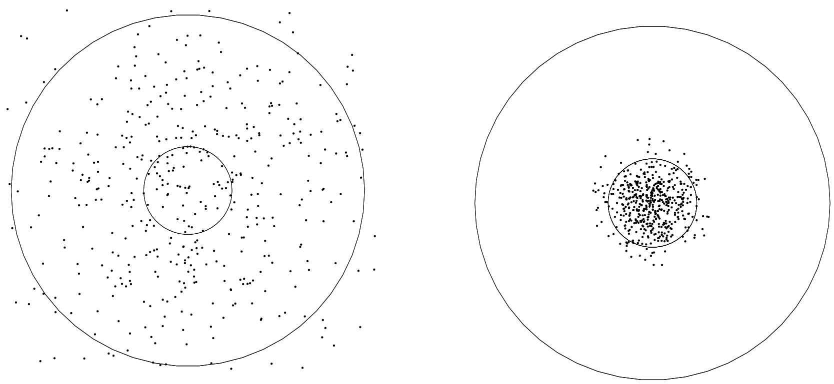
Em primeiro lugar, é razoável pensar que se área(A) $= 0$, então $\mathbb{P}(A) = 0$, como ocorre no exemplo 3.1.3. Mas, neste caso, não mais se sustenta a hipótese de que a probabilidade deva ser diretamente proporcional à área, já que um atirador razoável acertará com maior freqüência o círculo de centro C e raio 5 metros (cuja área é $25\pi m^2$) do que o seu complementar (cuja área é infinita)[3].

Neste momento, não preocupar-nos-emos em criar um modelo mais ou menos preciso para X, como fizéramos no exemplo 3.1.3, por sua maior complexidade, mas apresentamos, na figura 3.1, um simulação do desempenho comparado de quinhentos atiradores amadores e quinhentos profissionais. Note que, no caso dos amadores, a distribuição espacial das flechas não é proporcional à área. À primeira vista, no entanto, existiria certa proporcionalidade no caso dos profissionais. Essa idéia inicial é um equívoco, fruto da escala dos gráficos, tendo os profissionais comportamento análogo ao dos amadores desde que a escala seja escolhida de forma *adequada*. As escalas *adequadas* são diferentes pelo simples fato de que os profissionais são mais *precisos* e,

[3] Uma situação semelhante: o preço de um terreno depende de sua área, mas não é diretamente proporcional, pois também depende fundamentalmente de sua localização. Quanto você pagaria por um terreno de área nula? A semelhança está na negação da proporcionalidade do preço ao tamanho do terreno

portanto, suas flechas se *concentram* de forma bem mais acentuada em torno do centro do alvo. Note que a grande maioria de suas flechas acertaram uma região menor (centro preto do exemplo 3.1.5), enquanto os amadores tem suas flechas espalhadas por todo o alvo e várias caindo fora do mesmo.

Figura 3.1: Tiro ao Alvo



Exemplo 3.1.5 *Consideremos uma situação igual à do exemplo 3.1.4, com exceção de que agora o alvo tem um centro preto de* 50 *cm de diâmetro e o resto do círculo (um anel) de madeira está pintado de branco. O atirador ganha* 8 *pontos se acertar no centro,* 2 *se acertar na parte branca e* -5 *se não acertar na placa do alvo.*

Seja Ω *o mesmo conjunto do exemplo 3.1.4. Agora, no entanto, definiremos a variável* Y *(pontuação) da seguinte forma:*

$$Y(\omega) = \begin{cases} 8 & \textit{se } \omega \textit{ estiver no centro preto;} \\ 2 & \textit{se } \omega \textit{ estiver na parte branca da placa; e} \\ -5 & \textit{se } \omega \textit{ estiver fora da placa.} \end{cases}$$

Esta situação é semelhante à descrita no exemplo 3.1.2 e seria, portanto, interessante determinar as probabilidades dos conjuntos: $[Y = -5]$, $[Y = 2]$ *e* $[Y = 8]$. *Uma observação que o leitor atento já se deve ter feito é a de que esse exemplo do alvo e da importância factual apenas dos valores de* Y, *contrastada com o exato ponto em que a flecha atinge o plano* π *nos mostra a abundância excessiva de informação de* Ω *para os nossos propósitos e que a introdução de uma variável aleatória nos possibilita calcular probabilidades apenas para os casos de interesse, em vez de fazê-lo para uma classe de subconjuntos de* Ω *grande em demasia.*

Exemplo 3.1.6 *Os alunos inscritos na disciplina 'Probabilidade I' são submetidos a uma prova cuja duração máxima é de duas horas. Anota-se o tempo demorado* X *por cada aluno para entregá-la. Defina formalmente* Ω *e* X. *É fácil que haja empates nos tempos de entrega? Em que valor eles ocorreriam?*

Exemplo 3.1.7 *No sistema público de saúde de certa cidade, pretende-se organizar o atendimento às gestantes na hora do parto. Uma semana antes da data prevista, é importante prever se o parto será ou não de risco, já que, em caso afirmativo, a gestante será encaminhada a uma unidade de atendimento especial.*

Ω *é o conjunto das gestantes atendidas pelo sistema público de saúde. É bastante claro que um elemento de* Ω *- uma gestante - é um ente bastante complexo com todas as peculiaridades de qualquer ser humano. Portanto, é óbvia a necessidade de simplificar o problema.*

Na prática, a previsão sobre o parto é feita através de algumas características da gestante: idade em anos completos, peso, altura, condição de fumante, ser o primeiro parto, tempo transcorrido desde o parto anterior (se esse não for o primeiro), pressão arterial, temperatura, modificações no peso na última semana etc.. Na posse das informações, ou seja da particular combinação dos va-

lores observados das características consideradas, a equipe médica procederá ao encaminhamento adequado da paciente.

No presente caso, como na maioria dos problemas reais, há várias variáveis que devem ser utilizadas em forma conjunta para analisar a situação da gestante. Entretanto, em uma primeira fase, consideraremos apenas uma variável de cada vez.

A propósito desse exemplo, deve-se salientar que a escolha das variáveis relevantes para se fazer um trabalho desse tipo é não trivial e fica fora dos objetivos deste livro.

Definição 3.1.1 (Variável Aleatória)

Seja $(\Omega, \mathcal{F}, \mathbb{P})$ o espaço de probabilidade associado a um experimento aleatório. Uma função $X : \Omega \to \mathbb{R}$ é dita uma variável aleatória se:

$$\{\omega / X(\omega) \in I\} = [X \in I] = X^{-1}(I) \in \mathcal{F},$$

para todo intervalo $I \subset \mathbb{R}$.

Observação 3.1.1 *A definição acima foi feita para possibilitar o cálculo das probabilidades dos conjuntos $[X \in I]$, para todo intervalo $I \subset \mathbb{R}$.*

Definição 3.1.2 (Pontos Isolados)

Um subconjunto A de $\mathbb{R}$ é um conjunto de pontos isolados se satisfaz a seguinte propriedade: para todo $a \in A$, existe um intervalo aberto I_a tal que $I_a \cap A = \{a\}$.

Observação 3.1.2 *Conhecer as probabilidades dos conjuntos $[X \in I]$ para todo intervalo $I \subset \mathbb{R}$ é fundamental no exemplo 3.1.3; o mesmo acontece nos exemplos 3.1.4, 3.1.6 e na maioria das variáveis do 3.1.7, sendo óbvio que a definição 3.1.1 nos possibilita trabalhar nesses casos. Também em todos os exemplos citados* *neste parágrafo, com exceção do exemplo 3.1.6 e algumas variáveis do 3.1.7, as probabilidades dos conjuntos $[X = x]$ são nulas.*

Já no exemplo 3.1.2 (também nos exemplos 3.1.1, 3.1.5, 3.1.6 e na variável idade em anos completos do exemplo 3.1.7), o interesse maior estaria centrado em calcular as probabilidade de conjuntos do tipo $[X = x]$, para $x \in \mathbb{R}$. Em todos estes casos, Imagem(X) é um conjunto finito e, portanto, todos estes casos também são contemplados pela definição 3.1.1, dado que:

- *se $x \notin$ Imagem(X), então $[X = x] = \emptyset \in \mathcal{F}$;*
- *se Imagem(X) for um conjunto finito e $x \in$ Imagem(X), então existe um intervalo $I \subset \mathbb{R}$ tal que $[X = x] = [X \in I]$ - basta tomar $I = (x - \varphi, x + \varphi)$, onde*

  $$0 < \varphi \leq \min |x - y|$$

  (onde essa minimização é feita para o conjunto de y's tais que $y \in$ Imagem$(X), y \neq x$)

  e, portanto, $[X = x] \in \mathcal{F}$;
- *mais geralmente, se Imagem(X) for um conjunto de pontos isolados e $x \in$ Imagem(X), é obvio que $[X = x] \in \mathcal{F}$, dado que basta considerar qualquer intervalo $I \subset \mathbb{R}$ tal que $x \in I$ e $I \cap$ Imagem(X) $= \{x\}$ (veja definição 3.1.2) pois, neste caso, $[X = x] = [X \in I]$.*

Posteriormente, estudaremos diversos modelos probabilísticos que utilizam variáveis aleatórias cuja imagem é um conjunto de pontos isolados não necessariamente finito.

Portanto, podemos dizer que a definição 3.1.1 é apta para responder a todas as perguntas de interesse relativas aos exemplos 3.1.1-3.1.7.

Uma pergunta que o leitor ainda deve ter é sobre a utilidade da definição 3.1.1 para lidar com variáveis categóricas, isto é, que dividam as possibilidades de ocorrências em categorias como, por exemplo, ser ou não fumante ou o primeiro parto, como no exemplo 3.1.3. Não é nosso objetivo exaurir todas as questões relevantes que se podem colocar diante dos leitores mas, nesse caso, uma possível solução seria a de se criarem variáveis binárias que em conjunto representem a variável categórica de interesse e, através daquelas, utilizar a definição 3.1.1.

3.2 Função de Distribuição Acumulada

O conhecimento das probabilidades dos conjuntos $[X \leq r] = \{\omega / X(\omega) \leq r\}$ tem interesse próprio em qualquer um dos problemas mencionados acima e, mais geralmente, em qualquer situação que envolva variáveis aleatórias.

Dada uma variável aleatória X, é fácil verificar que, para todo $r \in \mathbb{R}$, o conjunto $[X \leq r] = \{\omega / X(\omega) \leq r\} \in \mathcal{F}$, dado que se $m = \max\{z \in \mathbb{Z} / z \leq r\}$, então:

$$[X \leq r] = \bigcup_{n \in \mathbb{Z}, n \leq m} [n-1 < X \leq n] \cup [m < X \leq r],$$

(se r for inteiro, então $m = r$ e $[m < X < r] = \emptyset$).

Também, dado que o conjunto $[X \leq r]$ foi escrito como a união disjunta de uma subfamília enumerável de conjuntos da família $\{[a < X \leq b], a, b \in \mathbb{R}, a < b\}$, é possível calcular as probabilidades dos conjuntos $[X \leq r] = \{\omega / X(\omega) \leq r]$, $r \in \mathbb{R}$ a partir das probabilidades dos conjuntos $\{[a < X \leq b], a, b \in \mathbb{R}, a < b\}$.

Reciprocamente, veremos que, a partir das probabilidades dos conjuntos $\{[X \leq r, r \in \mathbb{R}]\}$, será possível obter as probabilidades dos conjuntos $[X \in I]$, para qualquer intervalo $I \subset \mathbb{R}$. Por exemplo, se $a < b$, então $[X \leq b] = [X \leq a] \cup [a < X \leq b]$ e, portanto, $\mathbb{P}([a < X \leq b]) = \mathbb{P}([X \leq b]) - \mathbb{P}([X \leq a])$.

Além de se obterem as probabilidades dos conjuntos $[X \in I]$, $I \subset \mathbb{R}$, será possível, a partir das probabilidades de conjuntos do tipo $[X \leq a]$, achar as probabilidades de outros muitos conjuntos interessantes (veja teorema 3.2.3 e exercício 3.2.3).

Finalmente, para condensar a informação contida nos conjuntos $\{[X \leq r], \ r \in \mathbb{R}\}$, define-se a função de distribuição acumulada da variável aleatória

X a seguir.

Definição 3.2.1 (Função de Distribuição Acumulada)

A função de distribuição acumulada da variável aleatória X, $F : \mathbb{R} \to [0,1]$, é

$$F(r) = \mathbb{P}([X \leq r]) = \mathbb{P}(\{\omega / X(\omega) \leq r, r \in \mathbb{R}\}),$$

para todo $r \in \mathbb{R}$.

Muitas vezes, será conveniente denotar F por F_X , para salientar que se trata da função de distribuição acumulada associada à variável aleatória X (por exemplo, quando se trabalha com várias variáveis aleatórias simultaneamente). A função de distribuição acumulada receberá a abreviação fd .

Teorema 3.2.1 (Propriedades da Função de Distribuição)

Seja X uma variável aleatória definida no espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$. Então, a função de distribuição acumulada F de X tem as seguintes propriedades:

a) F é monótona não decrescente;

b) F é contínua à direita, ou seja, $\lim_{t \to r^+} F(t) = F(r)$; e

c) $\lim_{t \to -\infty} F(t) = 0$ e $\lim_{t \to \infty} F(t) = 1$.

Demonstração:
a) É conseqüência do seguinte fato: se $a < b$, então $[X \leq a] \subset [X \leq b]$;
b) Dado que F é monótona, existe $\lim_{t \to r^+} F(t)$ e também vale que $\lim_{t \to r^+} F(t) = \lim_{n \to \infty} F(t_n)$, para qualquer seqüência tal que $t_n \to r^+$. Consideremos, então, a seqüência $t_n = r + 1/n$: verifica-se que $\lim_{t \to r^+} F(t) = \lim_{n \to \infty} F(t_n) = F(r)$, dado que a seqüência decrescente de conjuntos $\{[X \leq r + 1/n], n \in \mathbb{N}\}$ converge para o conjunto $[X \leq r]$; e

c) Sendo F monótona existem $\lim_{t\to\ -\infty} F(t)$ e $\lim_{t\to\infty} F(t)$ e estes podem ser calculados através de quaisquer seqüências cujos 'limites' sejam $-\infty$ e $+\infty$, respectivamente. Por exemplo: $\lim_{t\to-\infty} F(t) = \lim_{m\to-\infty} F(m)$ e $\lim_{t\to\infty} F(t) = \lim_{n\to\infty} F(n)$, onde $n \in \mathbb{N}$ e $m = -n$ e basta-nos então considerar as seqüências de conjuntos $\{[X \leq -n], n \in \mathbb{N}\}$ e $\{[X \leq n], n \in \mathbb{N}\}$ e observar que a primeira delas decresce para $\emptyset$ e a que a segunda cresce para Ω (para detalhes, veja o exercício 3.3). ∎

Teorema 3.2.2 (Recíproco do Teorema 3.2.1)

Seja $F : \mathbb{R} \to \mathbb{R}$ tal que:

a) F é monótona não-decrescente;

b) F é contínua à direita, ou seja, $\lim_{t\to r^+} F(t) = F(r)$; e

c) $lim_{t\to-\infty}F(t) = 0$ e $lim_{t\to\infty}F(t) = 1$.

Nessas condições:

i) Existe uma variável aleatória X definida em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ cuja função de distribuição acumulada é F;

ii) A probabilidade $\mathbb{P}_X$ definida nos borelianos de $\mathbb{R}$ por:

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{\omega / X(\omega) \in B\})$$

é única e obviamente determinada pela função F.

O teorema 3.2.2 é de natureza puramente técnica (veja o exercício 3.4), de certa forma não-construtiva, pois não sabemos de qual característica (variável aleatória) F é a função de distribuição acumulada. Além disso, sua demonstração exige conhecimentos matemáticos bem além dos exigidos neste curso e será, portanto, omitida. No entanto, é fundamental para a Inferência Estatística, em validar existência de distribuições associadas à características amostrais.

Teorema 3.2.3 (Descontinuidades da Função de Distribuição) *Seja F a função de distribuição acumulada da variável aleatória X. Então:*

a) Existem e são finitos os limites laterais de F para todo $t \in \mathbb{R}$; além disso, satisfazem $\lim_{t\to r^-} F(t) \leq \lim_{t\to r^+} F(t)$;

b)$\lim_{t\to r^+} F(t) = F(r)$;

c) F é descontínua em $r \in \mathbb{R}$ se e somente se $\lim_{t\to r^-} F(t) < F(t)$ e, portanto, toda descontinuidade consiste em um salto;

d) para todo $r \in \mathbb{R}$, vale que $\mathbb{P}(X = r) = F(r) - \lim_{t\to r^-} F(t) =$ salto de F em r (este salto pode ser inclusive 0, em caso de um ponto de continuidade);

e) existe, no máximo, um número enumerável de descontinuidades.

Demonstração:
a) a existência e finitude dos limites laterais decorre do fato de F ser monótona e limitada; também, sendo F monótona não decrescente, então, para todo $r \in \mathbb{R}$, $\lim_{t\to r^-} F(t) \leq \lim_{t\to r^+} F(t)$;
b) é conseqüência da continuidade à direita de F;
c) decorre do fato de ser F monótona não decrescente;
d) para todo $r \in \mathbb{R}$, o conjunto $[X = r]$ é limite da seqüência decrescente $(r - 1/n < X \leq r]$; e
e) por definição, salto de F em t é igual a

$$\lim_{t\to r^+} F(t) - \lim_{t\to r^-} F(t);$$

então, sendo F monótona,

$$\# \left(\{t/\text{salto de } F \text{ em } t > \frac{1}{n}\} \right) \times \frac{1}{n} \leq 1 =$$
$$\sup_{t\in\mathbb{R}} F(t) - \inf_{t\in\mathbb{R}} F(t),$$

para todo $n \in \mathbb{N}$.

Portanto, $\#(\{t/\text{salto de } F \text{ em } t > 1/n\})$ é finito para todo $n \in \mathbb{N}$ e, conseqüentemente,

$$\{\text{saltos de } F\} = \bigcup_{n \in \mathbb{N}} \{\text{salto de } F/ \text{ salto} > 1/n\}$$

é no máximo enumerável (para detalhes, veja o exercício 3.5). ■

Observação 3.2.1 *As demonstrações dos teoremas 3.2.2 e 3.2.3 utilizam conceitos de análise. Em caso de dúvidas, seria bom que o leitor interessado procurasse um bom livro de análise como, por exemplo, o excelente Rudin (referência [29]).*

Os pontos e amplitudes dos saltos da fd têm grande importância na Teoria de Probabilidade e seu estudo motiva a seguinte definição associada.

Definição 3.2.2 (Função de Massa)

Seja X uma variável aleatória definida no espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$, com função de distribuição acumulada F. A **função de massa de probabilidade** *da variável aleatória X é a função $p : \mathbb{R} \to \mathbb{R}$ definida por:*

$$p(r) = \mathbb{P}(X = r) = \mathbb{P}(\{\omega/X(\omega) = r\}).$$

Lembre que, pelo item d do teorema 3.2.3, para todo $r \in \mathbb{R}$, tem-se $\mathbb{P}(X = r) = F(r) - \lim_{t \to r^-} F(t) =$ salto de F em r.

Outro ponto de interesse sobre a imagem de uma variável aleatória é relacionado à existência de limites reais. Para isso, nas definições 3.2.3 e 3.2.4, introduxem-se os conceitos de variáveis aleatórias limitadas e essecialmente limitadas, respectivamente.

Definição 3.2.3 (Variável Aleatória Limitada)

Seja X uma variável aleatória em $(\Omega, \mathcal{F}, \mathbb{P})$.

i) X é dita inferiormente limitada se existe $m \in \mathbb{R}$ tal que $X \geq m$ (ou seja, tal que $X(\omega) \geq m$ para todo $\omega \in \Omega$); m é dita uma cota inferior de X;

ii) X é dita superiormente limitada se existe $k \in \mathbb{R}$ tal que $X \leq k$ (ou seja, tal que $X(\omega) \leq k$ para todo $\omega \in \Omega$); k é dita uma cota superior de X; e

iii) X é dita limitada se for inferiormente e superiormente limitada.

As noções de cotas expressas na definição 3.2.3 e resultados como os do exercício 3.8 induzem a definição 3.2.4 e os exercícios 3.9 e 3.10. O que se deve entender por esse 'relaxamento' da definição 3.2.3 é que, pela natureza aleatória de X, não há necessidade de estudar suas cotas para todos os valores de ω mas apenas para aqueles que, em conjunto tenham probabilidade 1. Pense no seguinte: se eu jogar uma moeda para cima e definir como possível trajetória o $\mathbb{R}^3$, eu preciso de me preocupar com essa moeda destruindo um satélite de comunicação em órbita?

Definição 3.2.4 (Variável Aleatória Essencialmente Limitada)

Seja X uma variável aleatória em $(\Omega, \mathcal{F}, \mathbb{P})$.

i) X é dita essencialmente inferiormente limitada se existe $m \in \mathbb{R}$ tal que $\mathbb{P}(X \geq m) = 1$; m é dita uma cota inferior essencial de X;

ii) X é dita essencialmente superiormente limitada se existe $k \in \mathbb{R}$ tal que $\mathbb{P}(X \leq k) = 1$; k é dita uma cota superior essencial de X; e

iii) X é dita essencialmente limitada se existe $c \in \mathbb{R}$ tal que $\mathbb{P}(|X| \leq c) = 1$; c é dita uma cota essencial de X.

3.3 Variáveis Aleatórias Discretas

Estudaremos aqui aquelas variáveis aleatórias $X : \Omega \to \mathbb{R}$ cujas imagens sejam algum conjunto finito ou enumerável. No caso enumerável, acrescentaremos a condição de que os pontos da Imagem(X) sejam isolados.

Definição 3.3.1 (Variável Aleatória Discreta)

Seja $(\Omega, \mathcal{A}, \mathbb{P})$ o espaço de probabilidade associado a um experimento aleatório. Uma função $X : \Omega \to \mathbb{R}$ é dita uma variável aleatória discreta se:

i) Imagem(X) é um conjunto finito ou um conjunto enumerável de pontos isolados; e

ii) $[X = x] = X^{-1}(x) \in \mathcal{A}$, para todo $x \in$ Imagem(X).

Observação 3.3.1 *Se Imagem$(X) = \{\alpha_i\}_{i \in \mathbb{N}}$, então é claro que: $\{[X = \alpha_i]\}_{i \in \mathbb{N}}$ é uma partição de Ω e, portanto,*

$$\sum_{i \in \mathbb{N}} \mathbb{P}(X = \alpha_i) = 1.$$

Neste contexto, elementos $\omega \in \Omega$ e $\omega' \in \Omega$, tais que $X(\omega) = X(\omega')$, são equivalentes para todos os efeitos. Esta equivalência ocorre também no sentido estrito do termo, já que, para qualquer função X, a relação $\equiv$ definida em Domínio(X) através de: $u \equiv v$ se e somente se $X(u) = X(v)$, é uma relação de equivalência (ou seja, uma relação reflexiva, simétrica e transitiva). Essa última afirmação é equivalente à existência da partição mencionada na observação 3.3.1.

Um conjunto de resultados muito importantes pode ser resumido pela seguinte proposição.

Proposição 3.3.1 (Propriedades de uma Variável Discreta)

Seja X uma variável aleatória discreta (definição 3.3.1). Então:

(i) X satisfaz a definição 3.1.1 de uma variável aleatória;

ii) X satisfaz a condição iv) do exercício 3.14, ou seja:

$$\sum_{r\in\mathbb{R}} p(r) = \sum_{r\in\mathbb{R}} \mathbb{P}(X = r) = 1,$$

onde p é a função de massa de probabilidade de X e $\mathbb{P}(X = r) > 0$ se e somente se $r \in Imagem(X)$; e

iii) Se $Imagem(X) = \{\alpha_i\}_{i\in\mathbb{N}}$, então:

$$X = \sum_{i\in\mathbb{N}} \alpha_i \mathbf{1}_{A_i},$$

onde $A_i = X^{-1}(\alpha_i)$ para todo $i \in \mathbb{N}$ e, se F é a função de distribuição acumulada de X, tem-se que:

$$F(t) = \sum_{r\leq t} p(r) = \sum_{\alpha_i\leq t} \mathbb{P}(X = \alpha_i) = \sum_{i\in\mathbb{N}} \mathbb{P}(X = \alpha_i)\mathbf{1}_{[\alpha_i,\infty)}(t).$$

A demonstração da proposição 3.3.1 é deixada a cargo do leitor, no exercício 3.15. Notemos, como sua conseqüência mais importante que, para verificarmos se uma determinada seqüência $\{p_i, i \in I\}$ (seja I um conjunto finito ou não, desde que enumerável) representa as probabilidades de alguma variável aleatória discreta, basta-nos verificar que

$$\sum_{i\in I} p_i = 1; \text{ e} \tag{3.1}$$

$$p_i \geq 0, \ \forall i \in I, \tag{3.2}$$

ou seja, toda a probabilidade deve estar concentrada em I (por (3.1)) e todos os p_i's devem ser não-negativos (por (3.2)), para que possam ser realmente probabilidades. Lembrem-se que $\forall A \in \mathcal{F}$, $0 \leq \mathbb{P}(A) \leq 1$. Caso ambas as condições sejam satisfeitas, diz-se que $\{p_i, i \in I\}$ é uma **função de massa fidedigna**. Uma última observação interessante é a de que podemos substituir a condição de não-negatividade por uma de positividade, desde que façamos o devido ajuste no conjunto I. A escolha entre a condição de positividade ou não-negatividade será feita convenientemente em cada caso.

3.4 Distribuições Contínuas

Definição 3.4.1 (Variável Aleatória Contínua)

Seja X uma variável aleatória. X é dita contínua se sua função de distribuição acumulada F_X for uma função contínua. Equivalentemente, X é contínua se e somente se $p(t) = \mathbb{P}(X = t) = 0$ para todo $t \in \mathbb{R}$.

Exemplo 3.4.1 *Seja X a variável aleatória definida no exemplo 3.1.3. Foi visto que F_X é dada por:*

$$F_X(t) = \begin{cases} 0 & t < 0, \\ t/5 & 0 \leq t < 5, \\ 1 & t \geq 5; \end{cases}$$

Exemplo 3.4.2 *Seja V uma variável aleatória cuja função de distribuição acumulada $F_V(t)$ é dada por:*

$$F_V(t) = \frac{1}{2} + \frac{1}{\pi} arctg(t),$$

para todo $t \in \mathbb{R}$, onde $arctg$ é a função inversa de $\tan : (-\pi/2, \pi/2) \to \mathbb{R}$.

Exemplo 3.4.3 *Seja T uma variável aleatória cuja função de distribuição acumulada $F_T(t)$ é dada por:*

$$F_T(t) = \begin{cases} 0 & t < 0, \\ 1 - e^{-t} & t \geq 0. \end{cases}$$

Note que, no caso da variável V, do exemplo 3.4.2, para todo $t \in \mathbb{R}$, existe a derivada de $F_V(t)$, sendo

$$f_T(t) = (F_T)'(t) = \frac{1}{\pi}\frac{1}{1+t^2}.$$

Quando existe $f_T(t) = (F_T)'(t)$ para todo t, o teorema fundamental do cálculo integral afirma que, para todo $s, t \in \mathbb{R}$, $s < t$:

$$\mathbb{P}(s < T \leq t) = F_T(t) - F_T(s) = \int_s^t f_T(u)du.$$

Tomando-se $s \to -\infty$ na equação acima, obtém-se

$$F_T(t) = \int_{-\infty}^t f_T(u)du.$$

Observe-se que, sendo F_T monótona não decrescente e f_T contínua, $f_T(t) \geq 0$ para todo $t \in \mathbb{R}$. [4]

Muitas vezes, a função de distribuição acumulada, F, de uma certa variável aleatória satisfaz $F(t) - F(s) = \int_s^t f(u)du$, onde f é uma função não-negativa, integrável, mas não necessariamente contínua. Isto ocorre, por exemplo, no caso da função de distribuição acumulada $F_X(t)$ da variável X do exemplo 3.1.3. Lembre-se que $F_X(t)$ é dada por:

[4] Seja F tal que, para todo $t \in \mathbb{R}$, existe a derivada $F'(t) = f(t)$. Nessas condições, o teorema fundamental do cálculo integral garante que se $s, t \in \mathbb{R}$, $s < t$, então $F(t) - F(s) = \int_s^t f(u)du$. Se, além disso, acrescentarmos as hipóteses de que F é monótona não decrescente em $\mathbb{R}$ e que f é contínua em $\mathbb{R}$, então necessariamente $f(t) \geq 0$ para todo $t \in \mathbb{R}$.

Com efeito, seja $t_0 \in \mathbb{R}$ tal que $f(t_0) = c < 0$. Então, sendo f contínua em t_0, dado $\epsilon = -c/2$ existe $\delta > 0$ tal que se $t \in (t_0 - \delta, t_0 + \delta)$, então $f(t) \in (f(t_0 - \epsilon, f(t_0) + \epsilon) = (3c/2, c/2)$. Portanto, se $t(t_0 - \delta, t_0 + \delta)$, então $f(t) < -c/2$. Mas essa última afirmação implica em que

$$\int_{t_0-\delta}^{t_0+\delta} f(t)dt \leq 2\delta(c/2) = \delta c < 0,$$

o que é contraditório com o fato de F ser monótona não decrescente e, conseqüentemente, $F(to + \delta) - F(to - \delta) \geq 0$ para todo $\delta > 0$.

$$F_X(t) = \begin{cases} 0 & t < 0, \\ t/5 & 0 \leq t < 5, \\ 1 & t \geq 5. \end{cases}$$

Portanto, verifica-se que

$$(F_X)'(t) = \begin{cases} 0 & t < 0, \\ 1/5 & 0 < t < 5, \\ 0 & t > 5 \end{cases}$$

e que $(F_X)'(t)$ não existe em 0 e 5; também observa-se que, além de não estar definida em 0 e 5, $(F_X)'$ tem descontinuidades essenciais em 0 ou 5, já que, nesses pontos, não coincidem os limites laterais.

Entretanto, é fácil verificar que $F_X(t)$ satisfaz:

$$F_X(t) - F_X(s) = \int_s^t (F_X)'(u)du,$$

se $s, t \in \mathbb{R}$, $s < t$, sendo irrelevante para a validade da igualdade acima o fato de $(F_X)'$ não estar definida em 0 e 5 (ou em qualquer conjunto finito de pontos).

Também, se considerarmos a variável aleatória T, do exemplo 3.4.3, poderemos verificar que F_T satisfaz

$$F_T(t) - F_T(s) = \int_s^t f_T(u)du,$$

onde $f_T(t)$ é dada por:

$$f_T(t) = \begin{cases} 0 & t < 0, \\ e^{-t} & t \geq 0. \end{cases}$$

Novamente, observe-se que f_T não é contínua em 0. Esses últimos fatos motivam a seguinte definição.

Definição 3.4.2 (Função de Densidade)

Seja X uma variável aleatória contínua e F_X sua função de distribuição acumulada. Uma função f_X não-negativa e integrável em $\mathbb{R}$ é dita **função de densidade de probabilidade** *de X se*

$$F_X(t) - F_X(s) = \int_s^t f_X(u)du,$$

para $s, t \in \mathbb{R}$, $s < t$.

De forma análoga à notação da função de distribuição acumulada , f é denotada por f_X, quando se fizer necessária a distinção entre várias variáveis aleatórias. Além disso, nessas condições, a variável aleatória X é dita absolutamente contínua.

Veja que, caso exista uma densidade para a variável aleatória contínua X com função de distribuição acumulada F_X, ela não será única, no sentido de que, se a modificarmos num número finito de pontos, esta nova função continua a satisfazer o teorema fundamental do cálculo integral para F_X e, portanto, também é uma função de densidade de X. Com isso, é usual utilizarmo-nos da função de densidade de X que tiver menos pontos de descontinuidades e, caso possível, que seja contínua.

Como f é uma função limitada (por ser integrável), se F satisfaz

$$F(t) - F(s) = \int_s^t f(u)du,$$

para $s, t \in \mathbb{R}$, $s < t$, necessariamente F será uniformemente contínua, já que, se $s, t \in \mathbb{R}$ e $s < t$,

$$|F(t) - F(s)| = F(t) - F(s) = \int_s^t f(u)du \leq (t - s) \sup_{x \in \mathbb{R}} f(x).$$

De forma análoga ao caso discreto, iremos chamar de uma **função de densidade fidedigna** toda função f tal que:

$$f(x) \geq 0 \quad \forall x \in \mathbb{R} \quad \text{e} \quad \int_{-\infty}^{+\infty} f(x)dx = 1. \tag{3.3}$$

Claramente a não-negatividade de f é fundamental para que F, definida em f, no espírito da definição 3.4.2 seja monótona não-decrescente mas que sua integral seja um pode ser facilmente contornada por uma transformação desde que sua integral seja estritamente positiva.

Finalmente, existem distribuições contínuas que não são absolutamente contínuas mas está fora do escopo deste livro a construção de um exemplo ou uma discussão mais aprofundada do assunto. Nesse espírito, iremos intercambiar livremente os termos *absolutamente contínua* e *contínua* sem maiores preocupações, ao longo do texto.

3.5 Momentos de uma Variável Aleatória

3.5.1 Esperança de uma Variável Aleatória

Exemplo 3.5.1 (Trad. Livre de *Carta de C. Huygens, 1669*) *Usando observações feitas em Londres com extrema exatidão, temos os seguintes fatos. De 100 pessoas nascidas, morrem: 36, após 6 anos; 24, entre 6 e 16 anos; 15, entre 16 e 26 anos; 9, entre 26 e 36 anos; 6, entre 36 e 46 anos; 4, entre 46 e 56 anos; 3, entre 56 e 66 anos; 2, entre 66 e 76 anos; 1, entre 76 e 86 anos.*

Portanto, de 100 pessoas, aqueles que atingem a idade de: 6 anos são 64; 16 anos são 40; 26 anos são 25; 36 anos são 16; 46 anos são 10; 56 anos são 6; 66 anos são 3; 76 anos, 1; e 86 anos, 0.

Quem apostasse que uma criança concebida viveria até 6 anos, poderia apostar 64 contra 36 ou 16 contra 9.

Por outro lado, quem apostasse que uma criança nascida viveria até 16 anos, só apostaria 40 contra 60 ou 2 contra 3, porque, de 100, só iria haver 40 que viveriam até a idade de 16 anos.

> ...
>
> *De cem crianças concebidas, morrem 36 antes da idade de 6 anos. Daquelas, podemos dizer que viveram, compensando o outro, 3 anos.*
>
> *Dos 64 que chegam aos 6 anos, morrem 24 antes dos 16 anos. Neste caso, daqueles, viveram, compensando o outro, 11 anos.*
>
> *E assim por diante...*
>
> *Portanto, para uma criança nascida, há 36 chances de viver 3 anos, 24 de viver 11 anos etc.*
>
> *Por minha regra dos jogos de azar, temos que multiplicar cada número de chances pelos anos correspondentes e dividir a soma desses produtos (que é aqui 1822) pela soma de todas as chances (que aqui é 100). Sobre o quociente (que é aqui 18 anos e aproximadamente 2 meses), não se pode afirmar que haja evidências de que uma pessoa iria viver tanto, pois é muito mais aparente que ela iria morrer antes disso.*

No exemplo histórico acima, C. Huygens define o conceito de *Esperança* (observe-se que aqui trata-se de esperança de vida: daí vem o nome) de uma variável aleatória discreta. Em linguagem matemática moderna (introduzida ao longo deste livro), seu raciocínio seria colocado da seguinte forma.

Definição 3.5.1 (Esperança de uma V.a. com Imagem Finita) *Considere-se uma variável aleatória discreta X, com imagem $\{x_1, \ldots, x_n\}$, e seja p sua função de massa. Então, o 'valor esperado' (ou valor médio) da variável X, denotada por $\mathbb{E}(X)$, é a seguinte média ponderada:*

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

Podemos estender sem dificuldades a definição para v.a.'s discretas com imagem infinita.

Definição 3.5.2 (Esperança de uma Variável Aleatória Discreta)

Seja X uma variável aleatória discreta, com imagem $\{x_1, x_2, \ldots\}$, e função de massa p. Sua esperança é dada por

$$\mathbb{E}(X) = \sum_{i=1}^{+\infty} x_i p(x_i),$$

quando essa soma (infinita) converge.

Uma questão natural é a da extensão de tal conceito para variáveis aleatórias de qualquer natureza, isto é, que não sejam necessariamente discretas. Como já discutido anteriormente, foge aos objetivos deste livro, o tratamento sistemático de variáveis aleatórias que não sejam discretas ou absolutamente contínuas. Portanto, casos como variáveis mistas não serão tratadas nesta seção. No entanto, falta-nos ainda um definição de conceito equivalente para o caso de variáveis aleatórias (absolutamente) contínuas.

Para motivar uma definição análoga, devemos atentar para dois fatos.

Primeiramente, qual a interpretação que se pode dar a conceitos como os expostos nas definições 3.5.1 e 3.5.2? Claramente, se utilizarmos como motivação o texto de Huygens, vemos que sua tentativa na utilização de um valor *esperado* era o de caracterizar um valor *relevante* ou de alguma forma *central* no conjunto de todos os valores possíveis. Ele o fazia no espírito da noção clássica de probabilidade, isto é, em que há um número finito de eventos equiprováveis[5] e, dessa forma, o valor central de todos os possíveis nada mais é do que a média de todos eles. Quando distanciamo-nos da noção clássica de probabilidade, aceitando probabilidades diferentes para eventos discretos, por exemplo, a idéia de um valor central se torna mais complicada, como é, numa

[5] O leitor deve notar que a idéia de eventos equiprováveis é ancilar ao problema de mortalidade em questão, isto é, não há razão para se crer em qualquer equiprobabilidade entre idades. O conceito de equiprobabilidade é inteligentemente utilizado por Huygens como uma ferramenta meramente técnica, através daqueles anos vividos por um pelo outro, de forma a contar chances da única forma utilizada na época.

primeira leitura, o texto de Huygens. O que ele faz, no entanto, é simplesmente, susbstituir a média geral por uma média ponderada pelas respectivas probabilidades, o que é exatamente a interpretação da definição 3.5.1. Como interpretar a esperança de v.a.'s discretas (mas de infinitos valores), como o exposto na definição 3.5.2? Se pensarmos, intuitivamente, em que podemos definir valores esperados para cada conjunto finito de valores possíveis: $\{x_1, x_2, \ldots, x_k\}$ e tomar seu limite quando $k \to \infty$[6][7], teríamos que a esperança de uma v.a. discreta com conjunto de valores possíveis infinito seria resultado de uma reavaliação sistemática das esperanças das v.a's definidas em conjuntos finitos, quando do aumento do conjunto de valores possíveis. Para quem possa estar considerando essa idéia extremamente abstrata, reflita sobre o seguinte fato: toda a análise de Huygens está errada se num aumento de disponibilidade de dados populacionais, descobrimos pessoas que viveram além dos 86 anos? Qualquer precisa formulação de uma valor característico (como deve ser a esperança) deveria ser robusto a tais *mudanças* e é isto que a definição 3.5.2 nos traz sobre a definição 3.5.1.

Embuídos de tal espírito, devemos também lembrar que a noção de densidade e função de massa tem similaridades que por nós devem ser utilizadas para que possamos definir a esperança de uma variável contínua analogamente à de uma discreta.

Exemplo 3.5.2 (Termômetro Analógico de Laboratório)

Nos laboratórios farmacêuticos, científicos e fotográficos, entre outros, é muito importante o controle da qualidade do ar sob dois aspectos principais: temperatura e umidade relativa.

Utilizemo-nos da temperatura, T, e suponhemos que nosso interesse se concentre na temperatura média diária. Outra suposição importante é a de que tenhamos procedimentos laboratoriais

[6] Conceito equivalente à definição de uma série como o limite de somas quando o conjunto de índices cresce.

[7] Há que se tomar cuidado com a idéia de fazer tal operação de limite mas esses problemas fogem ao grau de complexidade deste livro.

históricos de forma que possamos tomá-los como representativos de valores característicos das **verdadeiras temperaturas**.

Se essa série fosse composta de medições horárias das temperaturas, numa resolução de $0,1^oC$ *e o sistema de condicionamento do ar fosse tal que não houvesse valores abaixo de* 18^oC *ou acima de* 23^oC*, teríamos* $\Omega_{0,1}$*, o espaço amostral da seguinte forma:*

$$\Omega_{0,1} = \{18; 18,1; \ldots, 22,9; 23\}$$

e poderíamos utilizar a definição 3.5.2 para obter a esperança da temperatura laboratorial como:

$$\mathbb{E}(T) = \sum_{j=0}^{50}(18 + 0,1 \times j)f_j,$$

onde f_j, $j = 1, 2, \ldots, 50$ *representam as respectivas freqüências relativas das temperaturas* $18 + 0,1 \times j$.

No entanto, podemos apontar um defeito grave nessa esperança. Na prática, podemos tomar os valores sob as condições em que foram observados e calcular quantidades mas, do ponto de vista conceitual, precisamos de refletir mais acentuadamente sobre **o que** *estamos calculando. O leitor já deve ter notado que a resistência de uma tal definição de esperança é nula, isto é, qualquer mudança da escala dos termômetros a faz errada.*

O que temos nela, na realidade, é uma excelente aproximação discreta da esperança quando sabemos sob que resolução estamos usando. Podemos chamá-la então de $E_{0,1}(T)$*. Assim, teríamos* $E_r(T)$*, onde r indicaria o grau de precisão de nossas medidas. O que as* $E_r(T)$ *teriam em comum e quais suas diferenças? Estas são bem fáceis de notar: quanto menor o r maior seria o espaço amostral a ele associado,* Ω_r *e conseqüentemente, mais somandos seriam utilizados para definir* E_r*. Por outro, todas essas esperanças têm em comum que não há um vício em seus valores, isto* *é, a escolha de um r não significa nada mais do que uma escolha do número de somandos e da precisão (em torno do verdadeiro valor).*

Na realidade, sabemos que o espaço amostral mais adequado é de natureza contínua pois cada temperatura pode assumir qualquer valor real[8]. *Portanto, supondo que os possíveis valores da temperatura estejam entre* 18 *e* 23*, vejamos alguns exemplos de discretizações, na figura 3.2.*

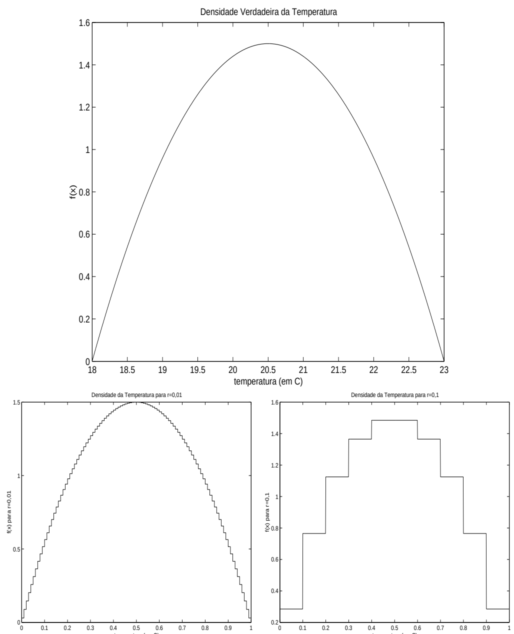
Notemos que a diferença entre os gráficos a da cardinalidade da densidade. Quando $r = 0,01$*, temos dez valores possíveis, quando* $r = 0,1$*, temos cem e a verdaeira* f *tem uma quantidada inumerável de valores. Esses gráfico ilustram o fato de que podemos pensar no* r *como o indexador de uma seqüência de valores* $E_r(T)$ *que converge para o verdadeiro valor* $E(T)$*: estamos aqui simplesmente repetindo a definição da* **Integral de Riemann***, com* r *nos dando quão fina nossa partição será.*

Uma maneira alternativa de pensar nessa integral como nossa esperança é a seguinte. Em geral, utilizam-se nos laboratórios os chamados Papéis de Controle, que vêm na forma de um disco, onde cada círculo representa uma temperatura diferente, como se vê na figura 3.3. A Temperatura média é dada, como visto, num curso usual de cálculo pela área dentro da curva de temperaturas.

Inspirando-se na definição 3.5.2 e no exemplo 3.5.2, podemos definir a esperança de uma v.a. (absolutamente) contínua da seguinte forma.

[8] É importante notar que o valor da temperatura pode ser qualquer número real, mesmo que nossa capacidade de observação não seja suficiente para registrá-lo. Dessa incapacidade prática, vem a noção de grau de precisão, r, e conseqüente ilusão de que a temperatura possa ser discreta. Notemos ainda que esse tipo de discretização é inerente à análise de de dados contínuos.

Figura 3.2: Termômetro Digital



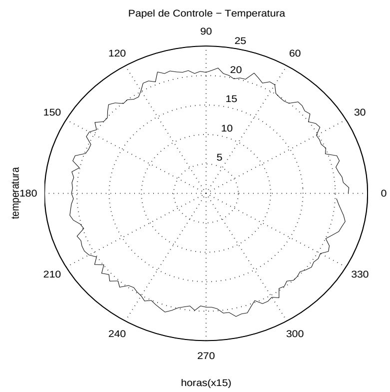
Definição 3.5.3 (Esperança de uma Variável Aleatória Contínua)

Seja X uma variável aleatória contínua com densidade f. Define-se a esperança de X, denotada por $\mathbb{E}(X)$, como

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x)\ dx,$$

quando esta integral existe.

Figura 3.3: Papel-Controle para Temperatura



3.5.2 Propriedades da Esperança

A esperança é um *operador* linear positivo[9], isto é

Proposição 3.5.1 (Propriedades da Esperança)

(a) Seja X uma variável aleatória positiva. Então, $\mathbb{E}(X) > 0$ *(Positividade).*

(b) $\mathbb{E}(1) = 1$*;*

(c) Se $a \in \mathbb{R}$ *então,* $\mathbb{E}(aX) = a\mathbb{E}(X)$*; e*

(d) $\mathbb{E}(X + b) = \mathbb{E}(X) + b$, $\forall b \in \mathbb{R}$.

Demonstração:

Primeiramente, o resultado (b) é muito simples: 1 pode ser visto como uma v.a. discreta assumindo somente um valor: 1. Tem, portanto, função de massa dada por $p(1) = 1$, $p(x) = 0$ caso $x \neq 1$ e o resultado é aplicação imediata da definição 3.5.2.

[9]Uma versão alternativa da proposição 3.5.1 é dada pelo exercício 3.21.