

Notas de Aula do Curso  
PGE950: Probabilidade

**Leandro Chaves Rêgo, Ph.D.**

2014.1

# Prefácio

Estas notas de aula foram feitas para compilar o conteúdo de várias referências bibliográficas tendo em vista o conteúdo programático da disciplina PGE950-Probabilidade do curso de mestrado em Estatística da Universidade Federal de Pernambuco. Em particular, elas não contém nenhum material original e não substituem a consulta a livros textos. Seu principal objetivo é dispensar a necessidade dos alunos terem que copiar as aulas e, deste modo, poderem se concentrar em entender o conteúdo das mesmas.

*Recife, março de 2014.*

*Leandro Chaves Rêgo, Ph.D.*

# Conteúdo

<b>Prefácio</b>	<b>i</b>
<b>1 Introdução à Probabilidade</b>	<b>1</b>
1.1 Experimento Aleatório . . . . .	1
1.2 Espaço Amostral . . . . .	2
1.3 Eventos e Coleção de Eventos . . . . .	3
1.3.1 Partição . . . . .	3
1.3.2 Álgebra de Eventos . . . . .	4
1.3.3 Função Indicadora . . . . .	6
1.4 Fundamentos de Probabilidade . . . . .	8
1.4.1 Hierarquia de Conceitos Estruturais de Probabilidade . . . . .	10
1.4.2 Interpretações de Probabilidade . . . . .	11
1.5 Frequências Relativas . . . . .	11
1.6 Axiomas de Kolmogorov . . . . .	13
1.6.1 Exemplos de Medidas de Probabilidade . . . . .	15
1.6.2 Propriedades de uma Medida de Probabilidade . . . . .	16
<b>2 Probabilidade Condicional</b>	<b>22</b>
2.1 Probabilidade Condicional . . . . .	22
2.2 Independência . . . . .	29
<b>3 Variável Aleatória</b>	<b>33</b>
3.1 Introdução . . . . .	33
3.2 Função de Distribuição Acumulada . . . . .	35
3.3 Tipos de Variável Aleatória . . . . .	37
3.3.1 Variável Aleatória Discreta . . . . .	37
3.3.2 Variável Aleatória Contínua . . . . .	38
3.3.3 Variável Aleatória Singular . . . . .	38
3.3.4 Decomposição de uma Variável Aleatória . . . . .	39
3.4 Principais Distribuições de Probabilidade . . . . .	40
3.5 Variáveis Aleatórias Multidimensionais . . . . .	47
3.5.1 Função de Distribuição Acumulada Conjunta . . . . .	48
3.5.2 Independência entre Variáveis Aleatórias. . . . .	49
3.5.3 Exemplos de Distribuições Multivariadas . . . . .	51

3.6	Funções de Variáveis Aleatórias . . . . .	51
<b>4</b>	<b>Esperança e Momentos de Variáveis Aleatórias</b>	<b>56</b>
4.1	O Conceito de Esperança . . . . .	56
4.2	Definição da Esperança - Caso Discreto . . . . .	56
4.3	As integrais de Riemman-Stieltjes e de Lebesgue-Stieltjes . . . . .	59
4.3.1	Propriedades da Integral de Lebesgue-Stieltjes . . . . .	61
4.4	Definição da Esperança - Caso Geral . . . . .	62
4.4.1	Interpretação Geométrica da Esperança . . . . .	64
4.5	Esperança de Funções de Variáveis Aleatórias . . . . .	66
4.5.1	Caso Discreto . . . . .	66
4.5.2	Caso Geral . . . . .	67
4.6	Propriedades da Esperança . . . . .	67
4.7	Momentos . . . . .	70
4.7.1	Momentos Centrais . . . . .	71
4.8	Momentos Conjuntos . . . . .	74
<b>5</b>	<b>Distribuição e Esperança Condicionais</b>	<b>77</b>
5.1	Distribuição condicional de $X$ dada $Y$ discreta . . . . .	77
5.2	Distribuição condicional de $X$ dada $Y$ : caso geral . . . . .	79
5.3	Esperança Condicional . . . . .	83
<b>6</b>	<b>Convergência Estocástica</b>	<b>87</b>
6.1	Seqüência de Eventos . . . . .	87
6.1.1	Borel-Canteli . . . . .	89
6.2	Covergência de Variáveis Aleatórias . . . . .	91
6.2.1	Tipos de Convergência . . . . .	92
6.2.2	Relação Entre os Tipos de Convergência . . . . .	98
6.3	Convergência de Vetores Aleatórios . . . . .	102
<b>7</b>	<b>Funções Características</b>	<b>104</b>
7.1	Motivação . . . . .	104
7.2	Definição . . . . .	105
7.2.1	Propriedades . . . . .	105
7.2.2	Exemplos de Funções Características . . . . .	110
7.3	Teorema da Continuidade de Levy . . . . .	111
7.4	Soma de um Número Aleatório de Variáveis Aleatórias . . . . .	115
7.5	Função Característica de um Vetor Aleatório . . . . .	117
7.6	Funções Geratrizes de Momento . . . . .	120
7.7	Teorema de Slutsky . . . . .	120
<b>8</b>	<b>Lei dos Grandes Números</b>	<b>123</b>
8.1	Motivação . . . . .	123
8.2	Lei Fraca dos Grandes Números . . . . .	125
8.3	Lei Forte dos Grandes Números . . . . .	127

8.4	Um Exemplo de Divergência das Médias . . . . .	134
<b>9</b>	<b>Teorema Central do Limite</b>	<b>136</b>
9.1	Motivação . . . . .	136
9.2	Teoremas e provas . . . . .	136
9.3	Teorema Central do Limite: Caso Multivariado . . . . .	145
9.4	Método Delta . . . . .	145
	<b>Referências Bibliográficas</b>	<b>149</b>

# Capítulo 1

## Introdução à Probabilidade

### 1.1 Experimento Aleatório

Um dos maiores objetivos de um estatístico é chegar a conclusões sobre certa população de objetos através da realização de um experimento. Um *experimento* é qualquer processo de observação. Em muitos experimentos de interesse, existe um elemento de incerteza, ou chance, que não importa quanto nós sabemos sobre o passado de outras performances deste experimento, nós essencialmente não somos capazes de prever seu comportamento em futuras realizações. As razões para nossa falta de habilidade para prever são varias: nós podemos não saber de todas as causas envolvidas; nós podemos não ter dados suficientes sobre as condições iniciais do experimento; as causas podem ser tão complexas que o cálculo do seu efeito combinado não é possível; ou na verdade existe alguma aleatoriedade fundamental no experimento. Estamos interessados em uma classe particular de experimentos, chamados *experimentos aleatórios*. Os seguintes traços caracterizam um experimento aleatório:

- (a) Se for possível repetir as mesmas condições do experimento, os resultados do experimento em diferentes realizações podem ser diferentes. Por exemplo, jogar uma moeda diversas vezes com bastante cuidado para que cada jogada seja realizada da mesma maneira.
- (b) Muito embora não sejamos capazes de afirmar que resultado particular ocorrerá, seremos capazes de descrever o conjunto de todos os possíveis resultados do experimento.<sup>1</sup>
- (c) Quando o experimento for executado repetidamente, os resultados individuais parecerão ocorrer de uma forma acidental. Contudo, quando o experimento for repetido um grande número de vezes, uma configuração definida ou regularidade surgirá. É esta regularidade que torna possível construir um modelo probabilístico. Por exemplo,

---

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

pense nas repetidas jogadas de uma moeda, muito embora caras e coroas apareçam sucessivamente, em uma maneira arbitrária, é fato empírico conhecido que, depois de um grande número de jogadas, a proporção de caras e de coroas serão aproximadamente iguais (assumindo que a moeda é simétrica).

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

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

## 1.2 Espaço Amostral

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

SS1. listar os possíveis resultados do experimento;

SS2. fazê-lo sem duplicação;

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

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

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

Podemos classificar espaços amostrais em dois tipos de acordo com o número de elementos que eles contêm. Espaços amostrais podem ser enumeráveis ou não enumeráveis; se os elementos do espaço amostral podem ser colocados em uma correspondência 1-1 com um subconjunto dos inteiros, o espaço amostral é enumerável. Em um nível filosófico, pode-se argumentar que só existem espaços amostrais enumeráveis, visto que medidas não podem ser feitas com infinita precisão. Enquanto na prática isto é verdadeiro, métodos estatísticos e probabilísticos associados com espaços amostrais não enumeráveis são, em geral, menos complicados que aqueles para espaços amostrais enumeráveis, e proporcionam uma boa aproximação para a situação (enumerável) real.

## 1.3 Eventos e Coleção de Eventos

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

**Definição 1.3.1:** Os eventos  $A$  e  $B$  são *disjuntos* ou *mutuamente excludentes* ou *mutuamente exclusivos* se não puderem ocorrer juntos, ou, em linguagem de conjuntos,  $A \cap B = \emptyset$ .

**Exemplo 1.3.2:** Sejam  $A$ ,  $B$ , e  $C$  eventos em um mesmo espaço amostral  $\Omega$ . Expresse os seguintes eventos em função de  $A$ ,  $B$ , e  $C$  e operações Booleanas de conjuntos.

(a) Pelo menos um deles ocorre:

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

(b) Exatamente um deles ocorre:

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

(c) Apenas  $A$  ocorre:

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

(d) Pelo menos dois ocorrem:

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

(e) No máximo dois deles ocorrem:

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

(f) Nenhum deles ocorre:

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

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

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

### 1.3.1 Partição

**Definição 1.3.3:** Dado um espaço amostral  $\Omega$ , uma partição  $\Pi = \{A_\alpha, \alpha \in \mathcal{I}\}$  de  $\Omega$  é uma coleção de eventos (subconjuntos de  $\Omega$ ) (neste caso, indexados por  $\alpha$  que toma valores no conjunto de índices  $\mathcal{I}$ ) e satisfaz:

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



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

■

Deste modo os eventos de uma partição são mutuamente excludentes (ou disjuntos) e cobrem todo o espaço amostral. Portanto, cada elemento  $\omega \in \Omega$  pertence a um, e somente um, dos eventos  $A_\alpha$  de uma partição.

**Exemplo 1.3.4:** Se  $\Omega = \{1, 2, 3, 4\}$ , então  $\{A_1, A_2\}$ , onde  $A_1 = \{1, 2, 3\}$  e  $A_2 = \{4\}$ , é uma partição de  $\Omega$ . ■

**Exemplo 1.3.5:** A coleção de intervalos  $\{(n, n + 1] : n \in \mathbb{Z}\}$  é uma partição dos números reais  $\mathbb{R}$ . ■

### 1.3.2 Álgebra de Eventos

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

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

**Definição 1.3.6:** Uma álgebra de eventos  $\mathcal{A}$  é uma coleção de subconjuntos do espaço amostral  $\Omega$  que satisfaz:

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

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

**Exemplo 1.3.7:**

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

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

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

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

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

**Lema 1.3.8:** *Se  $\mathcal{A}$  é uma álgebra, então  $\Omega \in \mathcal{A}$*

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

**Teorema 1.3.9:** *Sejam  $\mathcal{A}_1$  e  $\mathcal{A}_2$  álgebras de subconjuntos de  $\Omega$  e seja  $\mathcal{A} = \mathcal{A}_1 \cap \mathcal{A}_2$  a coleção de subconjuntos comuns as duas álgebras. Então,  $\mathcal{A}$  é uma álgebra.*

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

É fácil ver que a prova do Teorema 1.3.9 pode ser estendida para o caso de uma intersecção de um número arbitrário de álgebras. O seguinte corolário usa este fato para provar que sempre existe uma menor álgebra contendo uma família qualquer de eventos.

**Corolário 1.3.10:** *Existe uma menor (no sentido de inclusão) álgebra contendo qualquer família dada de subconjuntos de  $\Omega$ .*

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

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

Pelo Teorema 1.3.9,  $\mathcal{A}(\mathcal{C})$  é uma álgebra de eventos, e consequentemente é a menor álgebra de eventos contendo  $\mathcal{C}$ .  $\mathcal{A}(\mathcal{C})$  é conhecida como a *álgebra de eventos gerada por  $\mathcal{C}$* . ■

**Teorema 1.3.11:** *Se  $\mathcal{A}$  é uma álgebra de eventos, então*

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

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

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

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

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

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

**Observação 1.3.12:** Uma maneira de construir uma álgebra de eventos, é primeiro particionar  $\Omega$  em um número finito subconjuntos e depois considerar álgebra que consiste dos eventos que são uniões finitas dos subconjuntos da partição. ■

**Exemplo 1.3.13:** Por exemplo,  $\Omega = \{a, b, c, d\}$ . Considere a partição,  $\{\{a, c\}, \{b, d\}\}$ , então considere a coleção de eventos que consiste de uniões finitas dos eventos desta partição:  $\mathcal{A} = \{\emptyset, \Omega, \{a, c\}, \{b, d\}\}$ . É fácil ver que  $\mathcal{A}$  é uma álgebra de eventos.

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

**Exemplo 1.3.14:** Se  $\Omega = \{a, b, c, d, e, f\}$ , encontre a álgebra gerada por  $\mathcal{C} = \{\{a, b, d\}, \{b, d, f\}\}$ . Os átomos de  $\mathcal{C}$  são  $\{\{a\}, \{f\}, \{c, e\}, \{b, d\}\}$ . Logo,

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

### 1.3.3 Função Indicadora

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

**Definição 1.3.15: Função Indicadora.** A função indicadora  $I_A : \Omega \rightarrow \{0, 1\}$  de um evento  $A$  é dada por

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

■

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

**Exemplo 1.3.16:** Se  $I_A(\omega)$  for identicamente igual a 1, ou seja,  $I_A(\omega) = 1, \forall \omega \in \Omega$ , então  $A$  é igual ao espaço amostral  $\Omega$ . Se  $I_A(\omega)$  for identicamente igual a 0, então  $A$  é igual ao conjunto vazio  $\emptyset$ . Se  $I_A(\omega)$  for igual a 1 somente quando  $\omega = \omega_0$ , então  $A$  é o evento  $\{\omega_0\}$  que contém somente o elemento  $\omega_0$ . ■

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

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

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

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

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

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

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

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

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

**Exemplo 1.3.17:** Utilizando funções indicadoras, verifique que  $A \subseteq B \Leftrightarrow B^c \subseteq A^c$ .

**Solução:** Temos que

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

■

**Exemplo 1.3.18:** As seguintes questões não estão relacionadas umas com as outras.

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

d. Se  $I_A - I_B$  for identicamente igual a 1, o que podemos concluir sobre  $A$  e  $B$ ?

e. Se  $A \cap B = B \cup A$ , o que podemos concluir sobre  $A$  e  $B$ ?

**Solução:** Exercício. ■

**Exemplo 1.3.19:** Utilizando funções indicadoras, determine se  $(A - C) \cup (B - C) = (A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c)$ . (Sugestão: Faça um Diagrama de Venn.)

**Solução:** Seja  $\omega \in A \cap B \cap C^c$ . Então,  $I_A(\omega) = I_B(\omega) = I_{C^c}(\omega) = 1$ . Portanto, temos

$$I_{(A-C) \cup (B-C)} = I_{A-C} + I_{B-C} - I_{A-C}I_{B-C} = I_A I_{C^c} + I_B I_{C^c} - I_A I_{C^c} I_B I_{C^c}.$$

De onde conclui-se que  $I_{(A-C) \cup (B-C)}(\omega) = 1$ . Por outro lado,

$$\begin{aligned} I_{(A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c)} &= I_{(A \cap B^c \cap C^c)} + I_{(A^c \cap B \cap C^c)} - I_{(A \cap B^c \cap C^c)} I_{(A^c \cap B \cap C^c)} \\ &= I_A I_{B^c} I_{C^c} + I_{A^c} I_B I_{C^c} - I_A I_{B^c} I_{C^c} I_{A^c} I_B I_{C^c} \end{aligned}$$

De onde conclui-se que  $I_{(A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c)}(\omega) = 0$ . Logo,  $I_{(A-C) \cup (B-C)} \neq I_{(A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c)}$ , o que implica que  $(A - C) \cup (B - C) \neq (A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c)$ . ■

## 1.4 Fundamentos de Probabilidade

Raciocínio probabilístico aparece em uma ampla variedade de fenômenos de chance e incerteza, ele é lugar comum em nosso dia-a-dia. Nós expressamos julgamentos probabilísticos tanto através da linguagem como através de nossas ações. Ultrapassar um carro em uma estrada com outro carro vindo em direção oposta implica que calculamos as distâncias e velocidades, e calculamos os riscos de uma batida ocorrer e estamos conscientes das graves consequências de erros nos nossos julgamentos, mas os consideramos pequenos o suficiente. Raciocínio probabilístico no dia-a-dia enquanto não desenvolvido matematicamente precisa ser levado seriamente em conta se desejamos tomar decisões racionais.

Nota-se que, em geral, precisamos incorporar conhecimento probabilístico que seja tanto qualitativo e expresso linguisticamente como também o conhecimento quantitativo que pode ser expresso numericamente. Antes de focarmos em uma teoria probabilística, vamos explorar o espaço de alternativas. Nós podemos classificar as formas de raciocínio probabilístico nas seguintes dimensões:

- grau de precisão: o conceito estrutural
- o significado, ou interpretação a ser dada a probabilidade
- estrutura matemática formal de probabilidade dada por um conjunto de axiomas

O conceito estrutural determina a precisão com que podemos esperar que probabilidade represente fenômenos aleatórios. A interpretação proporciona a base com a qual probabilidade deve ser determinada e indica o que podemos esperar aprender com ela, ou seja, o que uma afirmação probabilística significa. O conceito estrutural e a interpretação guiam a

escolha dos axiomas. O conjunto de axiomas, contudo, pode somente capturar uma parte do que entendemos da interpretação.

Compreensão de fundamentos de probabilidade é importante, pois aplicações de teoria da probabilidade dependem fortemente de seus fundamentos. Por exemplo, os fundamentos influem na escolha dos métodos estatísticos a serem utilizados (Frequentistas, Bayesianos, ...) e na interpretação dos resultados obtidos. Os próximos exemplos ajudam a motivar um pouco a importância do estudo de fundamentos de probabilidade.

**Exemplo 1.4.1:** Suponha que Alice tenha uma moeda honesta e que ela e Bob saibam que a moeda é honesta. Alice joga a moeda e olha o resultado. Após a moeda ser jogada, qual a probabilidade de cara segundo Bob? Um argumento diria que a probabilidade ainda é  $1/2$ , pois Bob não aprendeu nada sobre o resultado da jogada, então ele não deve alterar o valor de sua probabilidade. Um outro argumento, questiona se realmente faz sentido falar sobre probabilidade de cara depois que a moeda foi jogada. Segundo este argumento, a moeda ou caiu cara ou coroa, então o melhor que Bob pode afirmar é que a probabilidade de cara ou é 0 ou é 1, mas ele não sabe discernir entre esses valores.

**Exemplo 1.4.2:** Suponha agora que Alice tenha duas moedas, uma honesta e outra tendenciosa e é duas vezes mais provável dar cara que coroa com esta moeda. Alice escolhe uma das moedas (suponha que ela sabe distinguir as moedas) e está prestes a jogá-la. Bob sabe que uma moeda é honesta e que a outra é tendenciosa e que é duas vezes mais provável cair cara que coroa com a moeda tendenciosa, mas ele não sabe que moeda Alice escolheu nem lhe foi dada a probabilidade com que Alice escolhe a moeda honesta. Qual a probabilidade de cara segundo Bob?

**Exemplo 1.4.3: Paradoxo de Ellsbergue.** Suponha que existam duas urnas cada uma com 60 bolas. A urna 1 contém 30 bolas azuis e 30 bolas verdes. Tudo que se sabe sobre a urna 2 é que ela contém bolas azuis e verdes, mas não sabe-se a distribuição das bolas. Considere que existem duas loteria com prêmios baseados no sorteio de bolas dessas urnas. Loteria  $L_1$  paga R\$1.000,00 se uma bola azul for sorteada na urna 1, e R\$0,00 caso contrário. Loteria  $L_2$  paga R\$1.000,00 se uma bola azul for sorteada na urna 2, e R\$0,00 caso contrário. A maioria das pessoas quando questionada se prefere um bilhete da Loteria  $L_1$  ou  $L_2$  prefere um bilhete da loteria  $L_1$ . Suponha agora que temos duas outras loterias  $L_3$  e  $L_4$ , onde a primeira paga R\$1.000,00 somente se uma bola verde for sorteada da urna 1, e a segunda para R\$1.000,00 somente se uma bola verde for sorteada da urna 2. Também, é verificado que a maioria das pessoas que preferiram a loteria  $L_1$  a loteria  $L_2$  preferem a loteria  $L_3$  a loteria  $L_4$ . Com estas preferências, não é possível que o decisor possua uma única distribuição de probabilidade subjetiva sobre as cores das bolas na urna 2, pois a primeira preferência ( $L_1$  sobre  $L_2$ ) indica que o decisor considera que existam mais bolas verdes que azuis na urna 2, e a segunda ( $L_3$  sobre  $L_4$ ) indica que o decisor considera que existam mais bolas azuis que verdes na urna 2. Esse fenômeno é conhecido na literatura como *aversão a ambiguidade*, e pode-se modelar a incerteza do decisor por um conjunto de medidas de probabilidade ao invés de uma única medida de probabilidade.

Nós discutiremos uma variedade de conceitos estruturais e interpretações de probabilidade. Depois nós focaremos na probabilidade numérica tradicional que satisfaz os famosos axiomas de Kolmogorov e em uma interpretação baseada em frequências de ocorrência.

### 1.4.1 Hierarquia de Conceitos Estruturais de Probabilidade

Os seguintes são exemplos de uma variedade de conceitos estruturais de probabilidade:

**Possivelmente.** “Possivelmente  $A$ ” é o conceito mais rudimentar e menos preciso, e o usado pelos antigos Gregos para distinguir entre o que era necessário e o que era contingente. Existe um número de conceitos de possibilidade que incluem os seguintes:

**possibilidade lógica**, no sentido que não se contradiz logicamente;

**possibilidade epistêmica**, segundo a qual ocorrência de  $A$  não contradiz nosso conhecimento, que inclui, mas estende mais que mera lógica;

**possibilidade física**, a ocorrência de  $A$  é compatível com leis físicas, contudo ela pode ser extremamente improvável — por exemplo, uma moeda parando e ficando equilibrada na borda em uma superfície rígida;

**possibilidade prática**, a noção do dia-a-dia segundo a qual  $A$  é praticamente possível se ele tem pelo menos uma verossimilhança não tão pequena de ocorrer.

**Provavelmente.** Provavelmente  $A$  é um fortalecimento da noção de possibilidade que significa mais que provável que não. Enquanto ela pode corresponder ao caso que a probabilidade numérica de  $A$  seja maior que  $1/2$ , este conceito não requer nenhum comprometimento com probabilidade numérica nem com o preciso estado de conhecimento que probabilidade numérica requer.

**Probabilidade Comparativa.** “ $A$  é pelo menos tão provável quanto  $B$ ”. A probabilidade comparativa inclui “provavelmente  $A$ ” através de “ $A$  é pelo menos tão provável quanto  $A$ ”. Pode ser relacionada com probabilidade numérica através de  $P(A) \geq P(B)$ ; embora como nos dois exemplos anteriores, probabilidade comparativa não requer nenhum comprometimento com probabilidade numérica.

**Probabilidade Intervalar.** “ $A$  tem probabilidade intervalar, ou probabilidade inferior e superior ( $\underline{P}(A), \overline{P}(A)$ )”. Isto permite um grau de indeterminação variável sem nenhum comprometimento com que exista um “verdadeiro” valor no intervalo.

**Probabilidade Numérica.** “A probabilidade de  $A$  é o número real  $P(A)$ .” Este é o conceito usual com o qual nos ocuparemos neste curso. Enquanto este conceito absorveu quase toda atenção de pessoas envolvidas com fenômenos de chance e incerteza e provou ser frutífero na prática científica, este não é o único conceito utilizado em linguagem ordinária e no raciocínio probabilístico do dia-a-dia. É duvidoso que probabilidade numérica seja adequada a todas as aplicações que ela é utilizada, e é provável que ela tenha inibido o desenvolvimento de teorias matemáticas apropriadas para outros fenômenos aleatórios.

De agora em diante focaremos no conceito estrutural mais utilizado e preciso que é a probabilidade numérica.

### 1.4.2 Interpretações de Probabilidade

Parece não ser possível reduzir probabilidade a outros conceitos; ela é uma noção em si mesma. O melhor que podemos fazer é relacionar probabilidade a outros conceitos através de uma interpretação. Os cinco mais comuns grupos de interpretação são os seguintes:

1. **Lógica:** grau de confirmação da hipótese de uma proposição que “ $A$  ocorre” dada uma evidência através da proposição que “ $B$  ocorreu”. Esta interpretação está ligada a um sistema lógico formal e não, digamos, ao mundo físico. Ela é usada para tornar o raciocínio indutivo quantitativo. Quando as evidências ou premissas são insuficientes para deduzir logicamente a hipótese ou conclusão, podemos ainda medir quantitativamente o grau de suporte que uma evidência dá a uma hipótese através de probabilidade lógica.
2. **Subjetiva:** se refere ao grau de crença pessoal na ocorrência do evento  $A$  e é medida através da interpretação comportamental de disposição a apostar ou agir.
3. **Frequentista:** se refere ao limite da frequência relativa de ocorrência do evento  $A$  em repetidas realizações não relacionadas do experimento aleatório  $\mathcal{E}$ . Note que limites de frequência relativas são uma idealização, pois não se pode repetir infinitas vezes um experimento.
4. **Propensidade:** tendência, propensidade, ou disposição para um evento  $A$  ocorrer. Por exemplo, considerações de simetria, podem levar a conclusão que um dado tem a mesma propensão ou tendência a cair em qualquer uma de suas faces.
5. **Clássica:** baseada em uma enumeração de *casos igualmente prováveis*.

## 1.5 Frequências Relativas

Resta-nos discutir o terceiro elemento para modelagem do raciocínio probabilístico, a associação de uma medida numérica a eventos que representam a verossimilhança com que eles ocorrem. As propriedades desta associação são motivadas em grande parte pelas propriedades de frequência relativas. Considere uma coleção de experimentos aleatórios  $\mathcal{E}_i$  que possuem a mesma álgebra de eventos  $\mathcal{A}$  e tem resultados individuais não necessariamente numéricos  $\{\omega_i\}$ . Seja  $X(\omega)$  uma função real dos resultados, com  $X_i = X(\omega_i)$  sendo o valor associado com o resultado  $\omega_i$  do  $i$ -ésimo experimento. Seja  $Av_n X = \frac{1}{n} \sum_{i=1}^n X_i$  a média dos resultados dos  $n$  primeiros experimentos. Por simplicidade matemática, assumiremos que a função  $X$  é escolhida de uma família  $\mathcal{F}$  de funções que podem assumir apenas um número finito de valores numéricos. Fixando uma dada sequência de resultados  $\{\omega_i\}$ , é fácil verificar as seguintes propriedades de  $Av_n$ :

$$Av_0. Av_n : \mathcal{F} \rightarrow \mathbb{R}.$$



Av1. Se para todo  $\omega$ ,  $X(\omega) \geq 0$ , então  $Av_n \geq 0$ .

Av2. Se  $X$  é uma função constante, então  $Av_n X = X$ .

Av3. Para todo  $X, Y \in \mathcal{F}$ , para todo  $\alpha, \beta \in \mathbb{R}$ ,

$$Av_n(\alpha X + \beta Y) = \alpha Av_n X + \beta Av_n Y.$$

Em particular, se estamos interessados em um dado evento  $A$  e escolhemos  $X(\omega) = I_A(\omega)$ , uma função binária, então a média é conhecida como a frequência relativa de  $A$ .

**Definição 1.5.1:** A frequência relativa de um evento  $A$ , determinada pelos resultados  $\{\omega_1, \dots, \omega_n\}$  de  $n$  experimentos aleatórios, é

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

Propriedades chaves da frequência relativa são:

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

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

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

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

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

Pode-se expressar  $Av_n$  em termos de  $r_n$ . Dada uma função  $X$  que assume valores no conjunto finito  $\{x_1, x_2, \dots, x_k\}$ , considere os  $k$  eventos  $\{A_i = \{\omega : X(\omega) = x_i\}, i = 1, 2, \dots, k\}$ . Podemos rearranjar os termos em  $Av_n X$  e reescrevê-la da seguinte forma:

$$Av_n X = \sum_{i=1}^k x_i r_n(A_i) = \sum_{i=1}^k x_i r_n(X = x_i).$$

Em particular, se para cada  $i$ , temos convergência da sequência  $r_1(X = x_i), r_2(X = x_i), \dots, r_n(X = x_i)$  para um limite  $p_i$ , então também temos convergência da média  $Av_n X$ ,

$$\lim_n Av_n X = \sum_{i=1}^k x_i p_i.$$

Este limite das médias, quando existe, serve como interpretação para o conceito essencial de esperança ou média de uma quantidade aleatória numérica  $X$ . Veremos mais sobre esperança neste curso.

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

## 1.6 Axiomas de Kolmogorov

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

**Definição 1.6.1:** Uma  $\sigma$ -álgebra  $\mathcal{A}$  é uma álgebra de eventos que também é fechada com relação a uma união enumerável de eventos,

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

**Exemplo 1.6.2:** A coleção de conjuntos de números reais finitos e co-finitos é uma álgebra que não é uma  $\sigma$ -álgebra.

**Exemplo 1.6.3:** A  $\sigma$ -álgebra de Borel  $\mathcal{B}$  de subconjuntos reais é, por definição, a menor  $\sigma$ -álgebra contendo todos os intervalos e é a  $\sigma$ -álgebra usual quando lidamos com quantidades reais ou vetoriais. Em particular, temos que uniões enumeráveis de intervalos (por exemplo, o conjunto dos números racionais), seus complementos (por exemplo, o conjunto dos números irracionais), e muito mais está em  $\mathcal{B}$ .

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

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

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

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

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

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

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

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

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

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

Um forma equivalente de K4 é a seguinte:

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

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

**Teorema 1.6.4:** *Se  $P$  satisfaz K0-K3, então  $P$  satisfaz K4' se, e somente se, ela satisfaz K4.*

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

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

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

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

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

Por definição de série numérica,

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

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

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

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

---

<sup>2</sup>K4 (ou equivalentemente K4') é uma idealização que não é aceita por alguns tratamentos subjetivistas de probabilidade, em especial não é aceita por uma escola de estatísticos liderados por deFinetti (1972). Assumir apenas aditividade finita, embora pareça mais plausível, pode levar a complicações inesperadas em teoria estatística. Portanto, nós prosseguiremos sobre a suposição que o axioma da continuidade (K4) é válido.

Então, por K4' temos que

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

Como por K4',

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

temos que

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

logo K4 é verdadeiro. ■

**Definição 1.6.5:** Uma função que satisfaz K0-K4 é chamada de uma medida de probabilidade.

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

Eventos são os elementos de  $\mathcal{A}$ , aos quais se pode atribuir probabilidade. Probabilidade é uma função cujo argumento é um conjunto. Portanto, não somente conjuntos, como também as operações sobre eles, têm uma importância fundamental em teoria da probabilidade.

### 1.6.1 Exemplos de Medidas de Probabilidade

**Exemplo 1.6.6:** Se  $\Omega$  for um conjunto finito, então temos que a probabilidade clássica que assume que todos os resultados são igualmente prováveis, é um exemplo de uma medida de probabilidade. Neste caso, temos que

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

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

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

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

**Exemplo 1.6.7:** Se  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$  um conjunto finito, e seja  $P(\{\omega_i\}) = p_i$ , onde  $p_i \geq 0, i \geq 1$  e  $\sum_{i=1}^n p_i = 1$ , e  $P(A) = \sum_{\omega_i \in A} P(\{\omega_i\})$ . Neste caso, também é fácil verificar que  $P$  é uma medida de probabilidade verificando os axiomas.

**Exemplo 1.6.8:** Seja  $\Omega = [0, 1]$  e  $\mathcal{B}_0$  a  $\sigma$ -álgebra de Borel restrita a eventos contidos em  $[0, 1]$ . Pode-se provar que existe uma medida de probabilidade  $\mu$  em  $(\Omega, \mathcal{B}_0)$  tal que para todo intervalo  $I$  em  $[0, 1]$   $\mu(I)$  é igual ao comprimento de  $I$ . Esta medida de probabilidade  $\mu$  é conhecida como *medida de Lebesgue*.

### 1.6.2 Propriedades de uma Medida de Probabilidade

**Teorema 1.6.9:** *Se  $P$  é uma medida de probabilidade, então*

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

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

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

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

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

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

**Teorema 1.6.10: Monotonicidade.** *Se  $A \subseteq B$ , então  $P(A) \leq P(B)$ .*

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

**Corolário 1.6.11:**  $P(A \cup B) \geq \max(P(A), P(B)) \geq \min(P(A), P(B)) \geq P(A \cap B)$ .

**Teorema 1.6.12:** *Uma expressão exata para a probabilidade de uma união não-disjunta é dada por*

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

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

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

■

**Teorema 1.6.13: Probabilidade de Partições.** *Se  $\{A_i\}$  é uma partição enumerável de  $\Omega$  feita de conjuntos em  $\mathcal{A}$ , então para todo  $B \in \mathcal{A}$*

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

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

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

O resultado segue então por K4'. ■

**Teorema 1.6.14: Desigualdade de Boole.** *Para  $n$  eventos arbitrários  $\{A_1, \dots, A_n\}$ , a desigualdade de Boole é*

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

**Prova:** Provaremos por indução matemática em  $n$ . A desigualdade é trivialmente verdadeira para  $n = 1$  e verdadeira para  $n = 2$ , pois é uma consequência imediata do Teorema 1.6.12. Assuma que a desigualdade é válida para  $n = k$  e vamos provar que ela é válida para  $n = k+1$ . Para ver isto, escrevemos  $\cup_{i=1}^{k+1} A_i = A_{k+1} \cup \cup_{i=1}^k A_i$ .

Pela desigualdade para  $n = 2$ ,

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

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

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

portanto, a desigualdade de Boole é verdadeira. ■

**Corolário 1.6.15:** *Para  $n$  eventos arbitrários  $\{A_1, \dots, A_n\}$ ,*

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

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

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

Logo,

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

■

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

**Teorema 1.6.16: Princípio da Inclusão-Exclusão.** *Seja  $I$  um conjunto genérico de índices que é um subconjunto não-vazio qualquer de  $\{1, 2, \dots, n\}$ . Para eventos arbitrários  $\{A_1, \dots, A_n\}$ ,*

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

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

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

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

**Prova:** A prova é por indução matemática em  $n$ . O resultado é trivialmente verdadeiro para  $n = 1$  e já foi provado para  $n = 2$  no Teorema 1.6.12. Assuma que o resultado vale para  $n = k$  e vamos provar que ele é verdadeiro para  $n = k + 1$ . Como na prova da desigualdade de Boole,  $\cup_{i=1}^{k+1} A_i = A_{k+1} \cup \cup_{i=1}^k A_i$ . Usando o resultado para  $n = 2$ , temos

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

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

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

O resultado segue ao rearranjarmos os termos destes somatórios. ■

**Exemplo 1.6.17:** Professor Leônidas está tentando calcular a probabilidade  $p = P(A)$  do evento  $A$ , e determinou que ela é uma raiz do seguinte polinômio de grau cinco:

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

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

**Exemplo 1.6.18:** Se  $\Omega = \{a, b, c\}$ , e a álgebra  $\mathcal{A}$  é o conjunto das partes de  $\Omega$ , e a medida de probabilidade  $P$  é parcialmente definida por

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

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

**Exemplo 1.6.19:** Se  $\{A_i\}$  for uma partição enumerável de  $\Omega$  e  $P(A_i) = ab^i$ ,  $i \geq 1$ , então quais as condições que  $a$  e  $b$  devem satisfazer para que  $P$  seja uma medida de probabilidade?

**Exemplo 1.6.20:** Em um grupo de  $r$  pessoas qual a probabilidade de haver pelo menos duas pessoas que façam aniversário no mesmo dia, assumindo que a distribuição de aniversários é uniforme ao longo do ano e desprezando a existência de anos bissextos?

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

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

Logo, a probabilidade deste evento é:

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

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

**Exemplo 1.6.21:** Em uma loteria de  $N$  números há um só prêmio. Salvador compra  $n$  ( $1 < n < N$ ) bilhetes para uma só extração e Sílvio compra  $n$  bilhetes, um para cada uma de  $n$  extrações. Qual dos dois jogadores têm mais chances de ganhar algum prêmio?

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

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

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

Para  $n = 2$ , temos:

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

Suponha que para  $n = k$ , temos que

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

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

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

**Exemplo 1.6.22:** Doze pessoas são divididas em três grupos de 4. Qual é a probabilidade de duas determinadas dessas pessoas ficarem no mesmo grupo?



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

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

**Exemplo 1.6.23:** Suponha que temos em uma sala  $n$  mães cada uma com um filho. Suponha formemos duplas aleatoriamente, onde cada dupla contém uma mãe e um filho, qual a probabilidade de que pelo menos uma mãe forme uma dupla com seu próprio filho?

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

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

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

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

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

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

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

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

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

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

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

**Exemplo 1.6.24:** Demonstre que se  $P(A_i) = 1$  para  $i = 1, 2, \dots$ , então  $P(\cap_{i=1}^{\infty} A_i) = 1$ .

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

**Exemplo 1.6.25:** Demonstre: se  $A_1, A_2, \dots$  e  $B_1, B_2, \dots$  são eventos aleatórios do mesmo espaço de probabilidade tais que  $P(A_n) \rightarrow 1$  e  $P(B_n) \rightarrow p$ , então  $P(A_n \cap B_n) \rightarrow p$ .

**Solução:** Note que

$$\begin{aligned} P(A_n \cap B_n) &= 1 - P((A_n \cap B_n)^c) = 1 - P(A_n^c \cup B_n^c) \\ &\geq 1 - P(A_n^c) - P(B_n^c) = P(A_n) + P(B_n) - 1. \end{aligned} \tag{1.1}$$

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

# Capítulo 2

## Probabilidade Condicional

### 2.1 Probabilidade Condicional

Como vimos no capítulo anterior, existem várias possíveis interpretações de probabilidade. Por exemplo, pode-se interpretar probabilidade de um evento  $A$  como um limite das frequências relativas de ocorrência do evento  $A$  em realizações independentes de um experimento. Por outro lado, a interpretação subjetiva de probabilidade associa a probabilidade de um evento  $A$  com o grau de crença pessoal que o evento  $A$  ocorrerá. Em ambos os casos, probabilidade é baseada em informação e conhecimento. Revisão desta base de informação ou conhecimento pode levar a revisão do valor da probabilidade. Em particular, conhecimento que determinado evento ocorreu pode influenciar na probabilidade dos demais eventos.

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

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

$$P(B^c|B) = 0 \tag{2.1}$$

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

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

$$\frac{P(A_1)}{P(A_2)} = \frac{P(A_1|B)}{P(A_2|B)} \quad (2.2)$$

Segue que (2.1) e (2.2) determinam completamente  $P(\cdot|B)$  se  $P(B) > 0$ .

**Teorema 2.1.1:** *Se  $P(B) > 0$  e  $P(\cdot|B)$  é uma medida de probabilidade em  $\Omega$  que satisfaz (2.1) e (2.2), então*

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

**Prova:** Como  $P(\cdot|B)$  é uma medida de probabilidade e satisfaz  $P(B^c|B) = 0$ , nós temos que  $P(B|B) = 1 - P(B^c|B) = 1$ . Considerando  $A_1 = A$  e  $A_2 = B$  em (2.2), temos então  $P(A|B) = \frac{P(A \cap B)}{P(B)}$  para  $A \subseteq B$ . Se  $A$  não é um subconjunto de  $B$ , temos que  $A = (A \cap B) \cup (A \cap B^c)$ . Como  $(A \cap B)$  e  $(A \cap B^c)$  são eventos disjuntos, temos  $P(A|B) = P(A \cap B|B) + P(A \cap B^c|B)$ . Como  $A \cap B^c \subseteq B^c$  e  $P(B^c|B) = 0$ , temos que  $P(A \cap B^c|B) = 0$ . Como  $A \cap B \subseteq B$ , usando o caso anterior

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

■

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

**Definição 2.1.2:** *Seja  $(\Omega, \mathcal{A}, P)$  um espaço de probabilidade. Se  $A, B \in \mathcal{A}$  e  $P(B) > 0$  a probabilidade condicional de  $A$  dado  $B$  é definida por*

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

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

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

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

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

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

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

A probabilidade condicional também satisfaz as seguintes propriedades:

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

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

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

Utilizando indução matemática, pode-se facilmente provar que

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

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

**Teorema 2.1.3:** *Seja a seqüência de eventos  $B_1, B_2, \dots$  uma partição de  $\Omega$ , então para todo  $A \in \mathcal{A}$*

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

**Prova:**

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

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

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

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

■

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

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

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

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

Para outro exemplo, suponha que os eventos  $B_1, B_2, \dots$  formam uma partição do espaço amostral, e os eventos  $B_i$ 's descrevem diferentes mensagens emitidas em um sistema de comunicações e  $A$  descreve uma mensagem recebida pelo sistema.  $P(A|B_i)$  determina a probabilidade que a mensagem  $B_i$  seja emitida e a mensagem  $A$  seja recebida por este sistema. Essas probabilidades condicionais especificam o modelo do canal de comunicações. Caso, as probabilidades  $P(B_i)$ 's de cada mensagem ser enviada e as probabilidades condicionais que descrevem o canal de comunicação sejam conhecidas pode-se usando o Teorema da Probabilidade Total obter a probabilidade incondicional que determinada mensagem  $A$  seja recebida. Porém geralmente, o que se busca é saber que dado uma certa mensagem foi recebida (efeito)  $A$  qual a probabilidade de cada uma das mensagens  $B_i$  terem sido as mensagens enviadas. Podem-se obter estas probabilidades utilizando a forma geral da famosa *fórmula de Bayes*:

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

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

**Exemplo 2.1.4:** Considere uma imagem formada por  $n \times m$  pixels com a  $k$ -ésima linha contendo  $d_k (\leq m)$  pixels defeituosos. No primeiro estágio do experimento uma linha é escolhida ao acaso e nós não sabemos qual foi a escolha. Nós então examinamos um pixel selecionada ao acaso nesta linha e descobrimos que o pixel é defeituoso (chamamos este evento de  $D$ ). Qual a probabilidade de que este pixel defeituoso esteja na linha  $k$ ? Seja  $R = k$  o evento que este pixel pertencia a  $k$ -ésima linha da imagem. A fórmula de Bayes nos permite determinar que dado que

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

nós temos que

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

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

**Exemplo 2.1.5:** Uma urna contém 4 bolas brancas e 6 bolas pretas. Sacam-se, sucessivamente e sem reposição, duas bolas dessa urna. Determine a probabilidade da primeira bola ser branca sabendo que a segunda bola é branca.

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

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

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

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

**Exemplo 2.1.6:** Se  $P(C|D) = 0,4$  e  $P(D|C) = 0,5$ , que evento é mais provável  $C$  ou  $D$ ?

**Solução:**

**Exemplo 2.1.7:** Se  $P(E) = 0,4$  e  $P(F) = 0,7$ , o que pode-se concluir sobre  $P(E|F)$ ?

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

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

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

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

**Exemplo 2.1.8: (Paradoxo de Monty Hall)** Monty Hall foi um popular apresentador de programa de jogos em TV cujo jogo começava mostrando ao participante 3 portas fechadas  $d_1, d_2, d_3$ , e atrás de apenas uma delas havia um prêmio valioso. O participante selecionava uma porta, por exemplo,  $d_1$ , mas antes que a porta fosse aberta, Monty Hall, que sabia em que porta estava o prêmio, por exemplo,  $d_2$ , abria a porta restante  $d_3$ , que não continha o prêmio. O participante tinha então permissão para ficar com sua porta original,  $d_1$ , ou escolher a outra porta fechada. A pergunta é se é melhor ficar com a porta original ou trocar de porta. Vamos agora utilizar a fórmula de Bayes para analisar este problema. Seja  $G$  uma porta escolhida aleatoriamente para conter o prêmio;  $Y$  a porta que o participante escolhe primeiro; e  $M$  a porta que Monty Hall abre. O participante não tem nenhum conhecimento

a priori sobre a localização do prêmio, ou seja ele considera todas as portas equiprováveis, e isto pode ser modelado por:

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

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

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

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

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

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

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

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

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

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



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

**Exemplo 2.1.9:** Seja  $D$  o evento que um indivíduo selecionado ao acaso de uma população tem uma doença particular,  $D^c$  seu complemento. A probabilidade que um indivíduo selecionado ao acaso nesta população tenha determinada doença é  $p_d$ . Existe um teste para diagnóstico desta doença que sempre acusa presença da doença quando o indivíduo tem a doença. Contudo, quando o indivíduo não tem a doença, o teste reporta falsamente que o indivíduo tem a doença com probabilidade  $p_t$ . Seja  $TP$  o evento que o teste reporta positivamente que o indivíduo tem a doença. Formalmente, temos:

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

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

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

Embora probabilidade condicional seja bastante útil, ela sofre de alguns problemas, em particular quando se quer tratar de eventos de probabilidade zero. Tradicionalmente, se  $P(B) = 0$ , então  $P(A|B)$  não é definida. Isto leva a um número de dificuldades filosóficas em relação a eventos com probabilidade zero. São eles realmente impossíveis? Caso contrário, quão improvável um evento precisa ser antes de ele ser atribuído probabilidade zero? Deve um evento em algum caso ser atribuído probabilidade zero? Se existem eventos com probabilidade zero que não são realmente impossíveis, então o que significa condicionar em eventos de probabilidade zero? Por exemplo, considere o espaço de probabilidade  $([0, 1], \mathcal{B}, \mu)$  onde  $\mathcal{B}$  é a  $\sigma$ -álgebra de Borel restrita a eventos contidos em  $[0, 1]$  e  $\mu$  é uma medida de probabilidade na qual todo intervalo em  $[0, 1]$  possui probabilidade igual ao seu comprimento. Seja  $B = \{1/4, 3/4\}$  e  $A = \{1/4\}$ . Como  $\mu(B) = 0$ ,  $\mu(A|B)$  não é definida. Porém parece razoável assumir que neste caso  $\mu(A|B) = 1/2$  já que  $\mu$  intuitivamente implica que todos os estados são equiprováveis, mas a definição formal de probabilidade condicional não nos permite obter esta conclusão.

Uma maneira de contornar alguns destes problemas é utilizar probabilidades não-padrão, que envolve conceitos de análise matemática não-padrão, que utiliza noções de infinitesimais. Outro modo é considerar probabilidades condicionais (e não incondicionais) como a noção fundamental. Uma medida de probabilidade condicional tem pares de eventos  $A, B$  como argumentos. Formalmente, a medida de probabilidade condicional é definida em uma álgebra de Popper.

**Definição 2.1.10:** Uma álgebra de Popper sobre  $\Omega$  é um conjunto  $\mathcal{A} \times \mathcal{A}'$  de subconjuntos de  $\Omega \times \Omega'$  tal que (a)  $\mathcal{A}$  é uma álgebra sobre  $\Omega$ , (b)  $\mathcal{A}'$  é um subconjunto não-vazio de  $\mathcal{A}$ , e (c)  $\mathcal{A}'$  é fechado em relação a superconjuntos em  $\mathcal{A}$ , ou seja, se  $B \in \mathcal{A}'$ ,  $B \subseteq B'$ ,  $B' \in \mathcal{A}$ , então  $B' \in \mathcal{A}'$ .

Pode-se então definir uma medida de probabilidade condicional da seguinte maneira:

**Definição 2.1.11:** Uma espaço de probabilidade condicional é uma tupla  $(\Omega, \mathcal{A}, \mathcal{A}', \mu)$  tal que  $\mathcal{A} \times \mathcal{A}'$  é uma álgebra de Popper sobre  $\Omega$  e  $\mu : \mathcal{A} \times \mathcal{A}' \rightarrow [0, 1]$  satisfaz as seguintes condições:

CP1.  $\mu(A|A) = 1$  se  $A \in \mathcal{A}'$ .

CP2.  $\mu(A_1 \cup A_2|B) = \mu(A_1|B) + \mu(A_2|B)$  se  $A_1 \cap A_2 = \emptyset$ ,  $A_1, A_2 \in \mathcal{A}$  e  $B \in \mathcal{A}'$ .

CP3.  $\mu(A_1 \cap A_2|A_3) = \mu(A_1|A_2 \cap A_3) \times \mu(A_2|A_3)$  se  $A_2 \cap A_3 \in \mathcal{A}'$  e  $A_1 \in \mathcal{A}$ .

## 2.2 Independência

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

**Definição 2.2.1:** O evento  $A$  é independente do evento  $B$  se  $P(A \cap B) = P(A)P(B)$ .

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

**Teorema 2.2.2:**  $A$  é independente dele mesmo se e somente se  $P(A) = 0$  ou  $P(A) = 1$ .

**Prova:**

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

■

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

**Teorema 2.2.3:** Se  $A$  e  $B$  são eventos independentes,  $A$  e  $B^c$  (resp.,  $A^c$  e  $B$ ,  $A^c$  e  $B^c$ ) também o são.

**Prova:** Note que

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

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

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

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

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

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

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

**Definição 2.2.4:** Uma coleção de eventos  $\{A_i\}_{i \in \mathcal{I}}$  é *independente par a par* se para todo  $i \neq j \in \mathcal{I}$ ,  $A_i$  e  $A_j$  são eventos independentes.

**Definição 2.2.5:** Uma seqüência finita de eventos  $A_1, A_2, \dots, A_n$ ,  $n \geq 1$ , é *mutuamente independente* se para todo  $I \subseteq \{1, \dots, n\}$ ,

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

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

Considere os seguintes exemplos que ilustram o conceito de independência.

**Exemplo 2.2.6:** Se  $\Omega = \{1, 2, 3, 4\}$  e  $P(\{w\}) = 1/4$ , então  $A = \{1, 2\}$ ,  $B = \{1, 3\}$ , e  $C = \{2, 3\}$  são eventos independentes par a par. Pode-se verificar isto pelo fato que

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

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

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

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

**Exemplo 2.2.7:** Certo experimento consiste em lançar um dado equilibrado duas vezes, independentemente. Dado que os dois números sejam diferentes, qual é a probabilidade condicional de

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

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

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

**Exemplo 2.2.8:** Suponha que um determinado experimento é realizado repetidas vezes de forma independente e observa-se a ocorrência de determinado evento  $A$  que tem probabilidade  $p$ . Qual é a probabilidade que  $A$  ocorra  $n$  vezes antes de  $A^c$  ocorrer  $m$  vezes?

**Solução:** Note que o evento  $A$  ocorra  $n$  vezes antes de  $A^c$  ocorrer  $m$  vezes é equivalente ao evento  $A$  ocorrer pelo menos  $n$  vezes nas primeiras  $n + m - 1$  repetições do experimento. Como a ordem de ocorrência do evento  $A$  nas repetições não é importante e as repetições são independentes, temos que o evento  $A$  ocorre  $k$  vezes em  $n + m - 1$  repetições do experimento tem probabilidade igual a:

$$P(k \text{ ocorrências de } A \text{ em } n + m - 1 \text{ repetições}) = \binom{n + m - 1}{k} p^k (1 - p)^{n + m - 1 - k}.$$

e, então,

$$P(n \text{ ocorrências de } A \text{ antes de } m \text{ ocorrências de } A^c) = \sum_{k=n}^{n+m-1} \binom{n + m - 1}{k} p^k (1 - p)^{n + m - 1 - k}.$$

**Exemplo 2.2.9:** Assuma que  $A_1, \dots, A_n$  são eventos mutuamente independentes e que  $P(A_i) = p_i$ . Nós calculamos as probabilidades dos seguintes eventos:

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

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

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

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

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

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

**Exemplo 2.2.10:** João e José disputam um jogo com uma moeda equilibrada. Cada jogador lança a moeda duas vezes e vence o jogo aquele que primeiro obtiver dois resultados iguais. João começa jogando e se não vencer passa a moeda para José e continuam alternando jogadas. Qual a probabilidade de João vencer o Jogo?

**Solução:** Seja  $A_k$  o evento dois resultados iguais são obtidos na  $k$ -ésima tentativa. Note que  $P(A_k) = \frac{1}{2}$ . Seja  $B_k$  o evento João ganha na sua  $k$ -ésima jogada. Então,

$$B_1 = A_1; \quad B_2 = A_1^c \cap A_2^c \cap A_3; \quad B_3 = A_1^c \cap A_2^c \cap A_3^c \cap A_4^c \cap A_5,$$

em geral,

$$B_k = A_1^c \cap A_2^c \cap \cdots \cap A_{2k-2}^c \cap A_{2k-1}.$$

Portanto,

$$P(B_k) = P(A_1^c \cap A_2^c \cap \cdots \cap A_{2k-2}^c \cap A_{2k-1}) = P(A_1^c)P(A_2^c) \cdots P(A_{2k-2}^c)P(A_{2k-1}) = \left(\frac{1}{2}\right)^{2k-1},$$

onde a penúltima igualdade se deve ao fato dos lançamentos serem independentes. Logo,

$$P(\text{João vencer}) = P(\cup_{k=1}^{\infty} B_k) = \sum_{k=1}^{\infty} P(B_k) = \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^{2k-1} = \frac{2}{3}.$$

# Capítulo 3

## Variável Aleatória

### 3.1 Introdução

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

**Definição 3.1.1:** Seja  $(\Omega, \mathcal{A}, P)$  um espaço de probabilidade. Uma função  $X : \Omega \rightarrow R$  é chamada de variável aleatória se para todo evento Boreliano  $B$ ,  $X^{-1}(B) \in \mathcal{A}$ .

Por definição, temos que  $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$  é o conjunto de elementos do espaço amostral cuja imagem segundo  $X$  está em  $B$ . Nós recordamos que um evento Boreliano é qualquer evento pertencente à  $\sigma$ -álgebra de Borel, onde a  $\sigma$ -álgebra de Borel é a menor  $\sigma$ -álgebra contendo todos os intervalos.

Para determinar se uma dada função  $X$  de  $\Omega$  para os reais é uma variável aleatória usando a definição, precisa-se checar se para todo evento Boreliano  $B$ , a imagem inversa de  $B$  de acordo com  $X$  faz parte da  $\sigma$ -álgebra  $\mathcal{A}$ . O próximo teorema prova que na verdade, só precisamos checar que a imagem inversa de intervalos da forma  $(-\infty, x]$  pertence à  $\sigma$ -álgebra  $\mathcal{A}$ .

**Teorema 3.1.2:** *Seja  $(\Omega, \mathcal{A})$  um espaço mensurável. Uma função real  $X : \Omega \rightarrow R$  é uma variável aleatória se e somente se*

$$X^{-1}((-\infty, \lambda]) = \{\omega : X(\omega) \leq \lambda\} \in \mathcal{A}, \forall \lambda \in R.$$

**Prova:**

Para provar este teorema, nós precisamos de uma série de Lemas.

**Lema 3.1.3:** *Seja  $\mathcal{B}$  a  $\sigma$ -álgebra de Borel, então  $X^{-1}(\mathcal{B}) = \{X^{-1}(B) : B \in \mathcal{B}\}$  é uma  $\sigma$ -álgebra de eventos de  $\Omega$ .*

**Prova:** Nós verificamos os três postulados para uma  $\sigma$ -álgebra:

(i)  $\Omega \in X^{-1}(\mathcal{B})$ .

Como  $R \in \mathcal{B}$ , nós temos  $X^{-1}(R) = \Omega \in X^{-1}(\mathcal{B})$ .

(ii) Se  $A \in X^{-1}(\mathcal{B})$ , então  $A^c \in X^{-1}(\mathcal{B})$ .

Suponha que  $A \in X^{-1}(\mathcal{B})$ , então existe  $A' \in \mathcal{B}$  tal que  $A = X^{-1}(A')$ . Como  $\mathcal{B}$  é uma  $\sigma$ -álgebra, temos que  $(A')^c \in \mathcal{B}$ . Logo,  $X^{-1}((A')^c) \in X^{-1}(\mathcal{B})$ . Como

$$X^{-1}((A')^c) = (X^{-1}(A'))^c,$$

temos que  $A^c \in X^{-1}(\mathcal{B})$ .

(iii) Se  $A_1, A_2, \dots \in X^{-1}(\mathcal{B})$ , então  $\cup_{i=1}^{\infty} A_i \in X^{-1}(\mathcal{B})$ .

Suponha que  $A_1, A_2, \dots \in X^{-1}(\mathcal{B})$ , então existem  $A'_1, A'_2, \dots \in \mathcal{B}$  tais que  $A_i = X^{-1}(A'_i)$  para  $i \geq 1$ . Como  $\mathcal{B}$  é uma  $\sigma$ -álgebra, temos que  $\cup_{i=1}^{\infty} A'_i \in \mathcal{B}$ . Logo,  $X^{-1}(\cup_{i=1}^{\infty} A'_i) \in X^{-1}(\mathcal{B})$ . Como

$$\cup_{i=1}^{\infty} X^{-1}(A'_i) = X^{-1}(\cup_{i=1}^{\infty} A'_i),$$

temos que  $\cup_{i=1}^{\infty} A_i \in X^{-1}(\mathcal{B})$ .

■

Dado qualquer classe de conjuntos  $\mathcal{C}$ , denotamos por  $\sigma(\mathcal{C})$  a menor  $\sigma$ -álgebra contendo  $\mathcal{C}$ . Desta forma se  $\mathcal{B}' = \{(-\infty, \lambda] : \lambda \in R\}$ , então  $\mathcal{B} = \sigma(\mathcal{B}')$ . O próximo lema prova um resultado semelhante ao do lema anterior, porém mais forte.

**Lema 3.1.4:**  $X^{-1}(\mathcal{B}) = \sigma(X^{-1}(\mathcal{B}'))$ , isto é, a imagem inversa de eventos Borelianos é igual a menor  $\sigma$ -álgebra contendo as imagens inversas dos eventos Borelianos.

**Prova:** De acordo com Lema 3.1.3,  $X^{-1}(\mathcal{B})$  é uma  $\sigma$ -álgebra. Como  $\mathcal{B}' \subseteq \mathcal{B}$ , temos que  $X^{-1}(\mathcal{B}') \subseteq X^{-1}(\mathcal{B})$ . Então, por definição de menor  $\sigma$ -álgebra, temos que  $\sigma(X^{-1}(\mathcal{B}')) \subseteq X^{-1}(\mathcal{B})$ .

Para provar igualdade, definimos

$$\mathcal{F} = \{B' \subseteq R : X^{-1}(B') \in \sigma(X^{-1}(\mathcal{B}'))\}.$$

É fácil provar que  $\mathcal{F}$  é uma  $\sigma$ -álgebra; nós omitimos os detalhes. Por definição, temos que  $X^{-1}(\mathcal{F}) \subseteq \sigma(X^{-1}(\mathcal{B}'))$  e  $\mathcal{B}' \subseteq \mathcal{F}$ . Como  $\mathcal{F}$  é uma  $\sigma$ -álgebra,  $\mathcal{B} = \sigma(\mathcal{B}') \subseteq \mathcal{F}$ . Portanto,

$$X^{-1}(\mathcal{B}) \subseteq X^{-1}(\mathcal{F}) \subseteq \sigma(X^{-1}(\mathcal{B}')).$$

■

Agora nós podemos provar o teorema. Suponha que  $X^{-1}(\mathcal{B}') \subseteq \mathcal{A}$ . Por definição de menor  $\sigma$ -álgebra,  $\sigma(X^{-1}(\mathcal{B}')) \subseteq \mathcal{A}$ . Então, pelo Lema 3.1.4,  $X^{-1}(\mathcal{B}) \subseteq \mathcal{A}$ , o que implica que  $X$  é uma variável aleatória.

■

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

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

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

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

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

## 3.2 Função de Distribuição Acumulada

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

**Definição 3.2.1:** A função de distribuição acumulada de uma variável aleatória  $X$ , representada por  $F_X$ , é definida por

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

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

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

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

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

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

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

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



**Teorema 3.2.2:** *Uma função real  $G$  satisfaz F1-F3 se e somente se  $G$  é uma distribuição de probabilidade acumulada.*

**Prova:** A prova de que se  $G$  for uma distribuição de probabilidade acumulada, então  $G$  satisfaz F1-F3 foi dada acima. A prova de que toda função real que satisfaz F1-F3 é uma função de probabilidade acumulada é complexa envolvendo o Teorema da Extensão de Carathéodory. Nós apresentamos aqui um esquema de como a prova é feita. Primeiro define-se  $P_X((-\infty, x]) = F_X(x)$ ,  $P_X((x, \infty)) = 1 - F_X(x)$ , e  $P_X((a, b]) = F_X(b) - F_X(a)$ . Com esta definição, considera-se a álgebra formada por união finita de intervalos e prova-se que  $P_X$  é  $\sigma$ -aditiva nesta álgebra. Finalmente, aplica-se o Teorema da Extensão de Carathéodory para provar que  $P_X$  pode ser estendida para todo evento Boreliano. ■

**Observação 3.2.3:** Uma função de distribuição pode corresponder a várias variáveis aleatórias no mesmo espaço de probabilidade  $(\Omega, \mathcal{A}, P)$ . Por exemplo, se  $X$  tem uma distribuição normal com parâmetros 0 e 1, então por simetria é fácil ver que  $-X$  também distribuição normal com parâmetros 0 e 1. Consequentemente,  $F_X = F_{-X}$ . No entanto,  $P(X = -X) = P(X = 0) = 0$ .

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

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

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

$$P_X(x) = F_X(x) - F_X(x^-).$$

Ou seja, a probabilidade da variável aleatória  $X$  assumir o valor  $x$  é igual ao salto da função de distribuição acumulada  $F_X$  no ponto  $x$ . O próximo teorema indica que o conjunto de pontos de descontinuidade de  $F$  é enumerável.

**Teorema 3.2.4:** *Seja  $D$  o conjunto de pontos de descontinuidade da função de distribuição  $F$ . Então,  $D$  é enumerável.*

**Prova:** Pela monotonicidade, temos que para todo  $x \in \mathbb{R}$ ,  $F(x^-) \leq F(x) \leq F(x^+)$ . Logo,  $x \in D$  se, e somente se,  $F(x^+) > F(x^-)$ . Para  $n = 1, 2, 3, \dots$  seja

$$A_n = \{x : F(x^+) - F(x^-) > \frac{1}{n}\}.$$

Então,  $D = \cup_{n=1}^{\infty} A_n$ . Vamos verificar que todo  $A_n$  contém menos que  $n$  pontos e, portanto, é finito. Dessa forma,  $D$  será enumerável.

Por absurdo, suponha que exista  $A_n$  que contém  $n$  pontos. Assim,  $A_n = \{x_1, x_2, \dots, x_n\}$ , onde  $x_1 < x_2 < \dots < x_n$  e

$$0 \leq F(x_1^-) \leq F(x_1^+) \leq F(x_2^-) \leq F(x_2^+) \leq \dots \leq F(x_n^-) \leq F(x_n^+) \leq 1.$$

Então, temos  $\sum_{k=1}^n [F(x_k^+) - F(x_k^-)] \leq 1$ . Mas por definição do conjunto  $A_n$ , temos que  $F(x_i^+) - F(x_i^-) > \frac{1}{n}$  para todo  $x_i \in A_n$ . Portanto,  $\sum_{k=1}^n [F(x_k^+) - F(x_k^-)] > n \times \frac{1}{n} > 1$ , absurdo. Logo,  $A_n$  contém menos que  $n$  pontos. ■

### 3.3 Tipos de Variável Aleatória

**Definição 3.3.1:** Existem três tipos de variáveis aleatórias:

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

$$F_X(x) = \sum_{i: x_i \leq x} p(x_i).$$

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

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

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

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

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

#### 3.3.1 Variável Aleatória Discreta

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

assume valores entre 0 e 1, sendo positiva para um número enumerável de pontos e satisfaz a seguinte propriedade  $\sum_i p(x_i) = 1$ .

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

### 3.3.2 Variável Aleatória Contínua

Vamos considerar agora o caso das variáveis aleatórias contínuas. Nós vimos na seção anterior que se uma variável aleatória é (absolutamente) contínua, então existe uma função  $f_X(x) \geq 0$  tal que  $F_X(x) = \int_{-\infty}^x f_X(t)dt$ . Deste modo,  $F_X$  é contínua e  $f_X(x) = F_X'(x)$ , exceto num conjunto de medida de Lebesgue nula. Uma função  $f(x) \geq 0$  é densidade de alguma variável aleatória se e somente se,  $\int_{-\infty}^{\infty} f(x)dx = 1$ , já que neste caso é fácil provar que a função  $F$  definida por  $\int_{-\infty}^x f(t)dt$  satisfaz as condições F1, F2, e F3. Portanto, pelo Teorema 3.2.2  $F$  é uma função de distribuição acumulada. Logo, a distribuição de uma variável aleatória contínua  $X$  pode ser determinada tanto pela função de distribuição acumulada  $F_X$  ou pela sua função de densidade  $f_X$ .

Uma variável aleatória  $X$  tem densidade se  $F_X$  é a integral (de Lebesgue) de sua derivada; sendo neste caso a derivada de  $F_X$  uma função densidade para  $X$ . Este fato pode ser provado utilizando argumentos de Teoria da Medida, mas omitimos os detalhes aqui. Sem recorrer a argumentos envolvendo Teoria da Medida, em quase todos os casos encontrados na prática, uma variável aleatória  $X$  tem densidade se  $F_X$  é (i) contínua e (ii) derivável por partes, ou seja, se  $F_X$  é derivável no interior de um número finito ou enumerável de intervalos fechados cuja união é a reta  $R$ .

Por exemplo, considere

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

Então  $X$  tem densidade pois  $F_X$  é contínua e derivável em todos os pontos da reta exceto em  $\{0, 1\}$ .

### 3.3.3 Variável Aleatória Singular

Vamos nesta seção dar o exemplo de uma função de distribuição de uma variável aleatória singular conhecida como *função de Cantor*. Esta função é contínua, derivável em todo ponto exceto em um conjunto de medida de Lebesgue nula, mas não é absolutamente contínua. Seja  $F(x) = 0$  se  $x < 0$  e  $F(x) = 1$  se  $x > 1$ . Continuemos por etapas:

**Etapa 1:** Seja  $F(x) = \frac{1}{2}$  para  $x \in (1/3, 2/3)$ . Então, o valor de  $F$  neste intervalo é igual a média dos valores de  $F$  nos intervalos vizinhos em que  $F$  já está definida:  $(-\infty, 0)$  e

$(1, \infty)$ .  $F$  continua sem definição em dois intervalos:  $[0, 1/3]$  e  $[2/3, 1]$  de comprimento total  $2/3$ .

**Etapa  $n + 1$ :** No terço central de cada um dos  $2^n$  intervalos restantes após a etapa  $n$ , seja  $F(x)$  igual à média dos valores nos dois intervalos vizinhos onde  $F$  já está definida. Por exemplo, na etapa 2 defina  $F(x) = 1/4$  para  $x \in (1/9, 2/9)$  e  $F(x) = 3/4$  para  $x \in \cup(7/9, 8/9)$ . Restarão então  $2^{n+1}$  intervalos (o dobro do número restante após a etapa  $n$ ), de comprimento total  $(2/3)^{n+1}$ , em que  $F$  ainda não estará definida.

Então definimos  $F$  por indução em um número enumerável de intervalos abertos, cujo complementar (ou seja, o conjunto onde  $F$  ainda não está definida) é o conjunto de Cantor, um conjunto de comprimento 0. Podemos estender a definição de  $F$  até o conjunto de Cantor  $C$  por continuidade: se  $x \in C$ , a diferença entre os valores de  $F$  nos dois intervalos vizinhos após a etapa  $n$  é  $1/2^n$ . Note que  $F$  é monótona não decrescente em  $C^c$ . Se  $a_n$  é o valor de  $F$  no intervalo vizinho esquerdo após a etapa  $n$ , e  $b_n$  é o valor no intervalo vizinho direito após a etapa  $n$ , então,  $a_n \uparrow$ ,  $b_n \downarrow$  e  $b_n - a_n \downarrow 0$ . Seja  $F(x)$  o limite comum de  $a_n$  e  $b_n$ . Deste modo  $F$  está definida em toda reta e é de fato uma função de distribuição (verifique!).

Seja  $X$  uma variável aleatória cuja função de distribuição é  $F$ , a função de Cantor. Então  $X$  não é discreta e nem contínua pois  $X$  não tem densidade  $F'(x) = 0$  em  $C^c$  e  $\int_{-\infty}^x F'(t)dt = 0$ , ou seja,  $F$  não é a integral de sua derivada, ou melhor, não é absolutamente contínua. Como  $F$  é contínua e  $F'(x) = 0$  para  $x \in C^c$  e  $C$  tem comprimento nulo, temos que  $X$  é uma variável aleatória singular.

### 3.3.4 Decomposição de uma Variável Aleatória

Vamos ver agora que toda variável aleatória é uma mistura dos três tipos: discreto, contínuo e singular. Seja  $X$  uma variável aleatória qualquer e seja  $F$  sua função de distribuição. Se  $J = \{x_1, x_2, \dots\}$  é o conjunto dos pontos de salto de  $F$  (se  $F$  for contínua  $J = \emptyset$ ), indiquemos com  $p_i$  o salto no ponto  $x_i$ , ou seja,

$$p_i = F(x_i) - F(x_i^-).$$

Definimos  $F_d(x) = \sum_{i: x_i \leq x} p_i$ .  $F_d$  é uma função degrau não-decrescente: a *parte discreta* de  $F$ . Como uma função monótona possui derivada em quase toda parte, seja

$$f(x) = \begin{cases} F'(x) & \text{se } F \text{ é diferenciável em } x, \\ 0 & \text{se } F \text{ não é diferenciável em } x. \end{cases}$$

Seja  $F_{ac}(x) = \int_{-\infty}^x f(t)dt$ .  $F_{ac}$  é não-decrescente, pois a integral indefinida de uma função não-negativa ( $f \geq 0$  porque  $F$  é não-decrescente). A sua derivada é igual a  $f$  em quase toda parte, de modo que  $F_{ac}$  é absolutamente contínua:  $F_{ac}$  é a *parte absolutamente contínua* de  $F$ .

Seja  $F_s(x) = F(x) - F_d(x) - F_{ac}(x)$ .  $F_s$  é contínua pois é a diferença de duas funções contínuas. A derivada de  $F_s$  é igual a zero em quase toda parte, porque  $F$  e  $F_{ac}$  têm a mesma derivada  $f$ , e  $F_d$  possui derivada zero em quase toda parte. Pode-se provar que  $F_s$  também é não-decrescente, mas está fora do escopo deste curso.  $F_s$  é a *parte singular* de  $F$ .

Esta discussão nos dá um método de decompor  $F$  em suas partes discreta, absolutamente contínua e singular. Considere o seguinte exemplo:

**Exemplo 3.3.2:** Suponha que  $X \sim U[0, 1]$  e  $Y = \min(X, 1/2)$ . Note que

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

$F_Y$  tem apenas um salto em  $x = 1/2$  e  $p_1 = 1/2$ . Logo,  $F_d(x) = 0$  se  $x < 1/2$  e  $F_d(x) = 1/2$  se  $x \geq 1/2$ . Diferenciando  $F_Y$ , temos

$$F'_Y(x) = \begin{cases} 0 & \text{se } x < 0 \text{ ou } x > 1/2, \\ 1 & \text{se } 0 < x < 1/2. \end{cases}$$

Logo, por definição,

$$f(x) = \begin{cases} 0 & \text{se } x \leq 0 \text{ ou } x \geq 1/2, \\ 1 & \text{se } 0 < x < 1/2. \end{cases}$$

Portanto,

$$F_{ac}(x) = \int_{-\infty}^x f(t)dt = \begin{cases} 0 & \text{se } x < 0, \\ x & \text{se } 0 \leq x \leq 1/2, \\ 1/2 & \text{se } x > 1/2. \end{cases}$$

Como  $F_d + F_{ac} = F_Y$ , temos que  $F_s(x) = 0, \forall x \in \mathbb{R}$  e não há parte singular. *Uma variável aleatória que possui apenas partes discreta e absolutamente contínua é conhecida como uma variável aleatória mista.* Na prática, é pouco provável que surja uma variável aleatória singular. Portanto, quase todas as variáveis aleatórias são discretas, contínuas ou mistas. A seguir veremos os principais tipos de distribuições.

## 3.4 Principais Distribuições de Probabilidade

Vamos primeiro explorar alguns exemplos importantes de variáveis aleatórias discretas.

### Aleatória.

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

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

**Bernoulli.**

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

A função de probabilidade Bernoulli pode ser utilizada para modelar a probabilidade de sucesso em uma única realização de um experimento. Em geral, qualquer variável aleatória dicotômica, ou seja que assume somente dois valores, pode ser modelada por uma distribuição Bernoulli.

**Binomial.**

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

Note que utilizando o Teorema Binomial, temos que

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

Logo, esta é uma legítima função probabilidade de massa.

A função de probabilidade Binomial pode ser utilizada para modelar a quantidade de erros em um texto de  $n$  símbolos quando os erros entre símbolos são assumidos independentes e a probabilidade de erro em um símbolo do texto é igual a  $p$ . Também pode ser utilizada para modelar o número de caras em  $n$  lançamentos de uma moeda que possui probabilidade  $p$  de cair cara em cada lançamento. Se  $p = 1/2$ , temos um modelo para o número de 1's em uma sequência binária de comprimento  $n$  escolhida aleatoriamente ou o número de caras em  $n$  lançamentos de uma moeda justa.

**Geométrica.**

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

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

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

Logo, esta é uma legítima função probabilidade de massa.

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

**Binomial Negativa ou Pascal.**

Esta distribuição é uma generalização óbvia da distribuição geométrica. Suponha que ao invés de estarmos interessados no tempo de espera até a primeira ocorrência de um evento,

estejamos interessados em calcular o tempo de espera até a  $r$ -ésima ocorrência de um evento. Seja  $Y$  o tempo de espera necessário a fim de que um evento  $A$  possa ocorrer exatamente  $r$  vezes. Temos que  $Y = k$  se, e somente se,  $A$  ocorrer na  $(k + 1)$ -ésima repetição e  $A$  tiver ocorrido  $r - 1$  vezes nas  $k$  repetições anteriores. Assumindo independência entre os experimentos, esta probabilidade é igual  $p \binom{k}{r-1} p^{r-1} (1-p)^{k-r+1}$ . Portanto,

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

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

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

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

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

### Zeta ou Zipf.

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

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

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

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

### Hipergeométrica.

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

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

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

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

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

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

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

### Poisson.

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

Por definição, temos que para todo  $x$  real,

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

Utilizando este fato, temos que

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

Logo, esta é uma legítima função probabilidade de massa.

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

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

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



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

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

Consideremos a expressão geral da probabilidade binomial,

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

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

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

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

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

ou seja obtemos a expressão de Poisson.

Mais geralmente, pode-se provar o seguinte teorema:

**Teorema 3.4.1:** *Se  $\lim_{n \rightarrow \infty} np_n = \alpha > 0$ , então*

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

**Prova:** Nós utilizamos os seguintes fatos:

1.  $\lim_{n \rightarrow \infty} \binom{n}{k} = \lim_{n \rightarrow \infty} \frac{n^k}{k!}$ .

2.  $\lim_{n \rightarrow \infty} np_n^2 = 0$ .
3.  $(1 - x)^n \leq e^{-nx}$ , para  $x \geq 0$ .
4.  $(1 - x)^n \geq e^{-nx - nx^2}$ , para  $0 \leq x \leq \frac{1}{2}$ .

Usando fatos 2, 3, e 4, nós obtemos  $\lim_{n \rightarrow \infty} (1 - p_n)^{n-k} = \lim_{n \rightarrow \infty} e^{-(n-k)p_n}$ . Logo, usando fato 1,

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

■

Vamos agora explorar alguns exemplos importantes de variáveis aleatórias contínuas.

### Uniforme.

Dizemos que  $X$  tem uma distribuição *uniforme* com parâmetros  $a$  e  $b$ , onde  $a$  e  $b$  são números reais e  $a < b$ , se a função densidade de  $X$  é igual a

$$f_X(x) = \frac{1}{b-a} U(x-a) U(b-x).$$

Este modelo é frequentemente usado impropriamente para representar “completa ignorância” sobre valores de um parâmetro aleatório sobre o qual apenas sabe-se estar no intervalo finito  $[a, b]$ . Esta distribuição também é frequentemente utilizada a fase de osciladores e fase de sinais recebidos em comunicações incoerentes.

### Exponencial.

Dizemos que  $X$  tem uma distribuição *Exponencial* com parâmetro  $\lambda$ , onde  $\lambda > 0$  é um número real, se a função densidade de  $X$  é igual a

$$f_X(x) = \lambda e^{-\lambda x} U(x).$$

A densidade exponencial pode ser utilizada para modelar os seguintes fenômenos: tempo de vida de componentes que falham sem efeito de idade; tempo de espera entre sucessivas chegadas de fótons, emissões de elétrons de um cátodo, ou chegadas de consumidores; e duração de chamadas telefônicas.

### Qui-quadrado.

Dizemos que  $X$  tem uma distribuição *Qui-quadrado* com parâmetro  $n$ , onde  $n$  é número natural, se a função densidade de  $X$  é igual a

$$f_X(x) = \frac{x^{n/2-1} e^{-x/2}}{2^{n/2} \Gamma(n/2)} U(x),$$

onde  $\Gamma(p) = \int_0^\infty x^{p-1}e^{-x}dx$  para  $p > 0$  é a função gama.  $n$  é conhecido como número de *graus de liberdade* da distribuição Qui-quadrado.

Pode-se provar que a soma dos quadrados de  $n$  variáveis aleatórias independentes com distribuição normal padrão possui uma distribuição Qui-quadrado com  $n$  graus de liberdade. A distribuição Qui-quadrado tem inúmeras aplicações em inferência estatística. Por exemplo, em testes qui-quadrados e na estimação de variâncias.

### Gama.

Dizemos que  $X$  tem uma distribuição *Gama* com parâmetros  $\alpha$  e  $\beta$ , onde  $\alpha > 0$  e  $\beta > 0$  são números reais, se a função densidade de  $X$  é igual a

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} U(x).$$

Pode-se provar que a soma de  $\alpha$  variáveis aleatórias exponenciais com média  $1/\beta$  tem uma distribuição Gama. É fácil ver que se  $\alpha = 1$ , temos uma distribuição exponencial com parâmetro  $\beta$ , e se  $\alpha = n/2$  e  $\beta = 1/2$  temos uma distribuição Qui-quadrado com  $n$  graus de liberdade.

### Beta.

Dizemos que  $X$  tem uma distribuição *Beta* com parâmetros  $\alpha$  e  $\beta$ , onde  $\alpha > 0$  e  $\beta > 0$  são números reais, se a função densidade de  $X$  é igual a

$$f_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1} du} U(x)U(1-x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1} U(x)U(1-x),$$

onde  $B(\alpha, \beta)$ , para  $\alpha > 0, \beta > 0$ , é a função beta que é o fator de normalização que garante que  $f_X$  é uma densidade.

Distribuições Beta são usadas exaustivamente em Estatística Bayesiana, pois elas são uma família de distribuições *a priori* conjugadas para distribuições binomiais e geométricas. A distribuição beta pode ser utilizada para modelar eventos que tem restrição de estar em um intervalo finito.

### t de Student.

Dizemos que  $X$  tem uma distribuição *t de Student* com parâmetro  $n$ , onde  $n$  é número natural, se a função densidade de  $X$  é igual a

$$f_X(x) = \frac{\Gamma[(n+1)/2]}{\Gamma[n/2]\sqrt{\pi n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{(n+1)}{2}},$$

onde  $n$  é conhecido como número de *graus de liberdade* da distribuição t de Student.

Pode-se provar que se  $Z$  tem uma distribuição normal padrão,  $V$  tem uma distribuição qui-quadrado com  $n$  graus de liberdade e  $Z$  e  $V$  forem independentes, então  $\frac{Z}{\sqrt{V/n}}$  tem uma distribuição t de Student com  $n$  graus de liberdade. A distribuição t de Student é bastante utilizada em inferência estatística. Por exemplo, pode-se utilizá-la para calcular intervalos de confiança para a média de uma amostra quando a variância da população não é conhecida.

**Pareto.**

Dizemos que  $X$  tem uma distribuição *Pareto* com parâmetros  $\alpha$  e  $\tau$ , onde  $\alpha$  e  $\tau$  são números reais positivos, se a função densidade de  $X$  é igual a

$$f_X(x) = \alpha\tau^\alpha x^{-\alpha-1}U(x - \tau).$$

A distribuição de Pareto é o exemplo mais fundamental de uma distribuição de cauda pesada. Ela pode ser utilizada para modelar distribuição de riquezas; atrasos em transmissão de pacotes; e duração sessões de Internet.

**Normal ou Gaussiana.**

Dizemos que  $X$  tem uma distribuição *Normal* (ou *Gaussiana*) com parâmetros  $m$  e  $\sigma$ , onde  $m$  e  $\sigma > 0$  são números reais, se a função densidade de  $X$  é igual a

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Historicamente, esta distribuição foi chamada de “normal” porque ela era amplamente aplicada em fenômenos biológicos e sociais que era sempre tida como a distribuição antecipada ou normal. Se  $m = 0$  e  $\sigma = 1$ , diz-se que  $X$  tem uma distribuição *normal padrão* ou *normal reduzida*. Aplicações da distribuição normal incluem ruído térmico em resistores e em outros sistemas físicos que possuem um componente dissipativo; ruídos de baixa-frequência como os encontrados em amplificadores de baixa frequência; e variabilidade em parâmetros de componentes manufaturados e de organismos biológicos (por exemplo, altura, peso, inteligência).

**Cauchy.**

Dizemos que  $X$  tem uma distribuição *Cauchy* com parâmetro  $a > 0$ , se a função densidade de  $X$  é igual a

$$f_X(x) = \frac{1}{\pi} \cdot \frac{a}{a^2 + x^2}.$$

A razão entre duas variáveis aleatórias com distribuição Normal padrão independentes tem uma distribuição Cauchy com parâmetro 1.

## 3.5 Variáveis Aleatórias Multidimensionais

Muitas vezes estamos interessados na descrição probabilística de mais de um característica numérico de um experimento aleatório. Por exemplo, podemos estar interessados na distribuição de alturas e pesos de indivíduos de uma certa classe. Para tanto precisamos estender a definição de variável aleatória para o caso multidimensional.

**Definição 3.5.1:** Seja  $(\Omega, \mathcal{A}, P)$  um espaço de probabilidade. Uma função  $\vec{X} : \Omega \rightarrow \mathbb{R}^n$  é chamada de um vetor aleatório se para todo evento  $B$  Boreliano de  $\mathbb{R}^n$ ,  $\vec{X}^{-1}(B) \in \mathcal{A}$ .

Onde um evento é Boreliano em  $\mathbb{R}^n$  se pertence a menor  $\sigma$ -álgebra que contem todas regiões da seguinte forma:  $C_{\vec{a}} = \{(X_1, X_2, \dots, X_n) : X_i \leq a_i, 1 \leq i \leq n\}$ .

Dado um vetor aleatório  $\vec{X}$ , pode-se definir uma probabilidade induzida  $P_{\vec{X}}$  no espaço mensurável  $(\mathbb{R}^n, \mathcal{B}^n)$  da seguinte maneira: para todo  $A \in \mathcal{B}^n$ , definimos  $P_{\vec{X}}(A) = P(\vec{X}^{-1}(A))$ . Por definição de vetor aleatório, tem-se que  $\vec{X}^{-1}(A) \in \mathcal{A}$ , então  $P_{\vec{X}}$  está bem definida.

### 3.5.1 Função de Distribuição Acumulada Conjunta

Para um vetor aleatório  $\vec{X}$ , uma maneira simples e básica de descrever a probabilidade induzida  $P_{\vec{X}}$  é utilizando sua *função de distribuição acumulada conjunta*.

**Definição 3.5.2:** A função de distribuição acumulada conjunta de um vetor aleatório  $\vec{X}$ , representada por  $F_{\vec{X}}$  ou simplesmente por  $F$ , é definida por

$$F_{\vec{X}}(\vec{x}) = P(C_{\vec{x}}) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n), \forall \vec{x} \in \mathbb{R}^n.$$

A função de distribuição acumulada  $F_{\vec{X}}$  satisfaz as seguintes propriedades:

F1. Se  $x_i \leq y_i, \forall i \leq n$ , então  $F_{\vec{X}}(\vec{x}) \leq F_{\vec{X}}(\vec{y})$ .

$$x_i \leq y_i \forall i \leq n \Rightarrow C_{\vec{x}} \subseteq C_{\vec{y}} \Rightarrow P(C_{\vec{x}}) \leq P(C_{\vec{y}}) \Rightarrow F_{\vec{X}}(\vec{x}) \leq F_{\vec{X}}(\vec{y}).$$

F2.  $F(x_1, x_2, \dots, x_n)$  é contínua a direita em cada uma das variáveis. Por exemplo, se  $y_m \downarrow x_1$ , então

$$F(y_m, x_2, \dots, x_n) \downarrow F(x_1, x_2, \dots, x_n), \text{ quando } m \rightarrow \infty.$$

F3a. Se para algum  $i \leq n$   $x_i \rightarrow -\infty$ , então  $C_{\vec{x}}$  decresce monotonicamente para o conjunto vazio  $\emptyset$ . Logo, pela continuidade monotônica de probabilidade, temos que

$$\lim_{x_i \rightarrow -\infty} F_{\vec{X}}(\vec{x}) = 0.$$

F3b. Se  $x_i \rightarrow \infty$ , então  $C_{\vec{x}}$  cresce monotonicamente para o conjunto  $\{X_1 \leq x_1, \dots, X_{i-1} \leq x_{i-1}, X_{i+1} \leq x_{i+1}, \dots, X_n \leq x_n\}$ , ou seja a restrição em  $X_i$  é removida. Então, podemos escrever

$$\lim_{x_i \rightarrow \infty} F_{\vec{X}}(\vec{x}) = F_{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n).$$

Portanto, a função de distribuição acumulada conjunta de  $X_1, \dots, X_{n-1}$  pode ser facilmente determinada da função de distribuição acumulada conjunta de  $X_1, \dots, X_n$  fazendo  $x_n \rightarrow \infty$ . Observe que *funções de distribuição acumuladas conjuntas de ordem maiores determinam as de ordem menores, mas o contrário não é verdadeiro*. Em particular, temos que

$$\lim_{\vec{x} \rightarrow \infty} F_{\vec{X}}(\vec{x}) = 1.$$

A função de distribuição acumulada de  $X_i$  que se obtém a partir da função acumulada conjunta de  $X_1, \dots, X_n$  fazendo  $x_j \rightarrow \infty$  para  $j \neq i$  é conhecida como *função de distribuição marginal* de  $X_i$ .

O próximo exemplo mostra que para  $n \geq 2$  as propriedades F1, F2, e F3 não são suficientes para que  $F$  seja uma função de distribuição.

**Exemplo 3.5.3:** Seja  $F_0 : \mathbb{R}^2 \rightarrow \mathbb{R}$  uma função definida no plano tal que  $F_0(x, y) = 1$  se  $x \geq 0$ ,  $y \geq 0$ , e  $x + y \geq 1$ , e  $F_0(x, y) = 0$ , caso contrário. É claro que F1, F2, e F3 são satisfeitas, mas  $F_0$  não é função de distribuição de nenhum vetor aleatório  $(X, Y)$ . Se fosse, teríamos uma contradição

$$\begin{aligned} 0 &\leq P(0 < X \leq 1, 0 < Y \leq 1) \\ &= F_0(1, 1) - F_0(1, 0) - F_0(0, 1) + F_0(0, 0) = 1 - 1 - 1 + 0 = -1 \end{aligned}$$

Os tipos discretos e contínuos de variáveis aleatórias têm os seguintes análogos no caso multivariado. (a) Se  $\vec{X}$  for um vetor aleatório discreto, ou seja assumir um número enumerável de valores  $\{\vec{x}_1, \vec{x}_2, \dots\}$ , podemos definir uma função de probabilidade de massa conjunta,  $p$  tal que

- $p(\vec{x}_i) \geq 0$ .
- $\sum_{i=1}^{\infty} p(\vec{x}_i) = 1$ .

Neste caso, pode-se definir a *função probabilidade de massa marginal* de  $X_i$  como sendo

$$p_{X_i}(x_i) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} p(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n).$$

(b) Seja  $\vec{X} = (X_1, \dots, X_n)$  um vetor aleatório e  $F$  sua função de distribuição. Se existe uma função  $f(x_1, \dots, x_n) \geq 0$  tal que

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(t_1, \dots, t_n) dt_1 \dots dt_n, \forall (x_1, \dots, x_n) \in \mathbb{R}^n,$$

então  $f$  é chamada de densidade conjunta das variáveis aleatórias  $X_1, \dots, X_n$ , e neste caso, dizemos que  $\vec{X}$  é (absolutamente) contínuo. Neste caso, define-se a *densidade marginal* de  $X_i$  como sendo

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n.$$

### 3.5.2 Independência entre Variáveis Aleatórias.

Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias definidas no mesmo espaço de probabilidade  $(\Omega, \mathcal{A}, P)$ . Informalmente, as variáveis aleatórias  $X_i$ 's são independentes se, e somente se, quaisquer eventos determinados por qualquer grupo de variáveis aleatórias distintas são independentes. Por exemplo,  $[X_1 < 5]$ ,  $[X_2 > 9]$ , e  $0 < X_5 \leq 3$  são independentes. Formalmente,

**Definição 3.5.4:** Dizemos que um conjunto de variáveis aleatórias  $\{X_1, \dots, X_n\}$  é mutuamente independente se, e somente se, para quaisquer eventos borelianos  $A_1, \dots, A_n$ ,

$$P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i).$$

O próximo teorema estabelece três critérios para provar que um conjunto de variáveis aleatórias é mutuamente independente.

**Teorema 3.5.5:** *As seguintes condições são necessárias e suficientes para testar se um conjunto  $\{X_1, \dots, X_n\}$  de variáveis aleatórias é mutuamente independente:*

(a)  $F_{\vec{X}}(\vec{x}) = \prod_{i=1}^n F_{X_i}(x_i)$ .

(b) Se  $\vec{X}$  for um vetor aleatório discreto,

$$p_{\vec{X}}(\vec{x}) = \prod_{i=1}^n p_{X_i}(x_i).$$

(c) Se  $\vec{X}$  for um vetor aleatório contínuo,

$$f_{\vec{X}}(\vec{x}) = \prod_{i=1}^n f_{X_i}(x_i), \forall (x_1, \dots, x_n) \in \mathbb{R}^n.$$

**Prova:** Para parte (a), note que se  $\{X_1, \dots, X_n\}$  são variáveis aleatórias mutuamente independentes, então

$$\begin{aligned} F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) &= P(X_1 \leq x_1, \dots, X_n \leq x_n) \\ &= \prod_{i=1}^n P(X_i \leq x_i) = \prod_{i=1}^n F_{X_i}(x_i), \forall (x_1, \dots, x_n) \end{aligned}$$

A prova da suficiência da parte (a) será omitida pois envolve argumentos de teoria da medida.

Para parte (b), se  $\{X_1, \dots, X_n\}$  são variáveis aleatórias mutuamente independentes, então

$$\begin{aligned} p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) &= P(X_1 = x_1, \dots, X_n = x_n) \\ &= \prod_{i=1}^n P(X_i = x_i) = \prod_{i=1}^n p_{X_i}(x_i), \forall (x_1, \dots, x_n) \end{aligned}$$

Reciprocamente, se a função de probabilidade de massa conjunta fatora e se  $\{x_{i1}, x_{i2}, \dots, x_{in}, \dots\}$  são os possíveis valores assumidos pela variável aleatória  $X_i$ , temos que

$$\begin{aligned} P(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) &= \sum_{i: x_{1i} \in B_1} \cdots \sum_{i: x_{ni} \in B_n} P(X_1 = x_{1i}, \dots, X_n = x_{ni}) \\ &= \sum_{i: x_{1i} \in B_1} \cdots \sum_{i: x_{ni} \in B_n} p_{X_1, \dots, X_n}(x_{1i}, \dots, x_{ni}) \\ &= \sum_{i: x_{1i} \in B_1} \cdots \sum_{i: x_{ni} \in B_n} \prod_{j=1}^n p_{X_j}(x_{ji}) = \prod_{j=1}^n P(X_j \in B_j) \end{aligned}$$

A parte (c) é uma consequência direta da parte (a) e da definição de função de densidade. Omitimos os detalhes. ■

É fácil observar que utilizando, a definição de probabilidade condicional que se  $X$  e  $Y$  são independentes, então para todo  $A$  e  $B$  boreliano tal que  $P(Y \in B) > 0$ :

$$P(X \in A|Y \in B) = P(X \in A),$$

ou seja, se  $X$  e  $Y$  são independentes o conhecimento do valor de  $Y$  não altera a descrição probabilística de  $X$ .

### 3.5.3 Exemplos de Distribuições Multivariadas

#### A Distribuição Multinomial

Vamos dar o exemplo de uma distribuição conjunta de variáveis aleatórias, que pode ser considerada como uma generalização da distribuição binomial. Considere um experimento aleatório qualquer e suponha que o espaço amostral deste experimento é particionado em  $k$  eventos  $\{A_1, A_2, \dots, A_k\}$ , onde o evento  $A_i$  tem probabilidade  $p_i$ . Suponha que se repita este experimento  $n$  vezes de maneira independente e seja  $X_i$  o número de vezes que o evento  $A_i$  ocorreu nestas  $n$  repetições. Então,

$$P(X_1 = n_1, X_2 = n_2, \dots, X_k = n_k) = \frac{n!}{n_1!n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k},$$

onde  $\sum_{i=1}^k n_i = n$ . (Relembre que o número de maneiras de arranjar  $n$  objetos,  $n_1$  dos quais é de uma espécie,  $n_2$  dos quais é de uma segunda espécie,  $\dots$ ,  $n_k$  dos quais são de uma  $k$ -ésima espécie é dado pelo coeficiente multinomial  $\frac{n!}{n_1!n_2! \dots n_k!}$ .)

#### A Distribuição Normal Bivariada

Dizemos que o vetor aleatório  $(X, Y)$  possui distribuição normal bivariada quando tem densidade dada por

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\},$$

onde  $\sigma_1 > 0, \sigma_2 > 0, -1 < \rho < 1, \mu_1 \in \mathbb{R}, \mu_2 \in \mathbb{R}$ .

Se  $\rho = 0$ , esta densidade fatora e temos que  $X$  e  $Y$  são independentes. Se  $\rho \neq 0$ , esta densidade não fatora e  $X$  e  $Y$  não são independentes.

## 3.6 Funções de Variáveis Aleatórias

Muitas vezes sabemos a distribuição de probabilidade que descreve o comportamento de uma variável aleatória  $X$  definida no espaço mensurável  $(\Omega, \mathcal{A})$ , mas estamos interessados na descrição de uma função  $Y = H(X)$ . Por exemplo,  $X$  pode ser uma mensagem enviada em um canal de telecomunicações e  $Y$  ser a mensagem recebida. Nosso problema é



determinar  $P(Y \in A)$ , onde  $A$  é um evento Boreliano, dado  $P_X$ . Para determinarmos esta probabilidade, estaremos interessados na imagem inversa a função  $H$ , ou seja, a probabilidade do evento  $\{Y \in A\}$  será por definição igual a probabilidade do evento  $\{X \in H^{-1}(A)\}$ , onde  $H^{-1}(A) = \{x \in \mathbb{R} : H(x) \in A\}$ . Para que esta probabilidade esteja bem definida, precisamos restringir  $H$  tal que  $H^{-1}(A)$  seja um evento boreliano para todo  $A$  boreliano, caso contrário não poderemos determinar  $P(\{X \in H^{-1}(A)\})$ ; uma função que satisfaz esta condição é conhecida como *mensurável com respeito a  $\mathcal{A}$  e  $\mathcal{B}$* . Note que  $Y$  também pode ser vista como uma função do espaço amostral  $\Omega$ ,  $Y(\omega) = H(X(\omega))$  para todo  $\omega \in \Omega$ . Visto dessa maneira  $Y$  é uma variável aleatória definida em  $(\Omega, \mathcal{A})$ , pois para todo boreliano  $A$   $Y^{-1}(A) = X^{-1}(H^{-1}(A))$  e como por suposição  $H^{-1}(A)$  é boreliano e  $X$  é uma variável aleatória, temos que  $X^{-1}(H^{-1}(A)) \in \mathcal{A}$  e portanto satisfaz a definição de uma variável aleatória. *Nesses problemas é sempre útil fazer um esboço do gráfico da transformação  $H$  para determinarmos quais são as regiões inversas  $H^{-1}(A)$ .*

Vamos primeiro tratar este problema no caso de variáveis aleatórias discretas. Neste caso para qualquer função  $H$ , temos que  $Y = H(X)$  é uma variável aleatória discreta.

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

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

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

**Exemplo 3.6.1:** Admita-se que  $X$  tenha os valores possíveis  $1, 2, 3, \dots$  e suponha que  $P(X = n) = (1/2)^n$ . Seja  $Y = 1$  se  $X$  for par e  $Y = -1$  se  $X$  for ímpar. Então, temos que

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

Consequentemente,

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

Podemos estender este resultado para uma função de um vetor aleatório  $\vec{X}$  de forma análoga. Neste caso se  $\vec{Y} = H(\vec{X})$ , denotemos por  $\vec{x}_{i1}, \vec{x}_{i2}, \vec{x}_{i3}, \dots$  os valores de  $\vec{X}$  tal que  $H(\vec{x}_{ij}) = \vec{y}_i$  para todo  $j$ . Então, temos que

$$P(\vec{Y} = \vec{y}_i) = P(\vec{X} \in \{\vec{x}_{i1}, \vec{x}_{i2}, \vec{x}_{i3}, \dots\}) = \sum_{j=1}^{\infty} P(\vec{X} = \vec{x}_{ij}) = \sum_{j=1}^{\infty} p_{\vec{X}}(\vec{x}_{ij}),$$

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

Vamos ver agora um exemplo no caso em que  $\vec{X}$  é contínuo.

**Exemplo 3.6.2:** Se  $X \sim U[0, 1]$ , qual a distribuição de  $Y = -\log(X)$ ? Como

$$0 < Y < \infty \Leftrightarrow 0 < X < 1$$

e  $P(0 < X < 1) = 1$ , temos  $F_Y(y) = 0, y \leq 0$ . Se  $y > 0$ , então

$$P(Y \leq y) = P(-\log(X) \leq y) = P(X \geq e^{-y}) = 1 - e^{-y},$$

ou seja,  $Y \sim Exp(1)$ .

Antes de prosseguirmos ao caso geral, vamos abrir um parênteses relembrar o que é o Jacobiano de uma função.

### Jacobiano de uma Função

Dado um conjunto de  $n$  equações em  $n$  variáveis  $x_1, \dots, x_n$ ,

$$y_1 = f_1(x_1, \dots, x_n), \dots, y_n = f_n(x_1, \dots, x_n),$$

a matriz Jacobiana é definida por

$$J = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \dots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \dots & \frac{\partial y_n}{\partial x_n} \end{pmatrix}$$

O determinante de  $J$  é chamado de *Jacobiano*. Pode-se provar que o módulo Jacobiano nos dá a razão entre volumes  $n$ -dimensionais em  $\vec{y}$  e  $\vec{x}$  quando a maior dimensão  $\Delta x_i$  tende a zero. Deste modo, temos que o módulo do Jacobiano aparece quando queremos mudar as variáveis de integração em integrais múltiplas, ou seja, existe um teorema do cálculo que afirma que se  $f : G_0 \rightarrow G$  for uma bijeção entre  $G_0$  e  $G$ ,  $f$  e as derivadas parciais que aparecem na matriz Jacobiana forem funções contínuas em  $G_0$ , e o Jacobiano for diferente de zero para todo  $x \in G_0$

$$\int_A \dots \int g(y_1, \dots, y_n) dy_1 \dots dy_n = \int_{f^{-1}(A)} \dots \int g(f_1(x_1, \dots, x_n), \dots, f_n(x_1, \dots, x_n)) |J| dx_1 \dots dx_n,$$

para qualquer função  $g$  integrável em  $A \subseteq G$ .

Vamos agora utilizar mudança de variáveis para resolver o seguinte exemplo da soma de duas variáveis aleatórias.

**Exemplo 3.6.3:** Suponha que  $(X, Y)$  tenha densidade conjunta  $f(x, y)$  e seja  $Z = X + Y$ . Neste caso,

$$F_Z(z) = P(Z \leq z) = P(X + Y \leq z) = P((X, Y) \in B_z),$$

onde  $B_z = \{(x, y) : x + y \leq z\}$ . Portanto,

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f(x, y) dx dy.$$

Fazendo a mudança de variáveis  $s = x + y$ ,  $t = y$ , que tem jacobiano igual a 1, temos

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^z f(s-t, t) ds dt = \int_{-\infty}^z \int_{-\infty}^{\infty} f(s-t, t) dt ds.$$

Logo,  $\int_{-\infty}^{\infty} f(s-t, t) dt$  é a densidade da soma  $Z = X + Y$ , ou seja,

$$f_Z(z) = \int_{-\infty}^{\infty} f(z-t, t) dt = \int_{-\infty}^{\infty} f(s, z-s) ds,$$

onde fizemos a troca de variáveis  $s = z - t$  para obter a última expressão.

Se  $X$  e  $Y$  forem variáveis aleatórias independentes com densidades  $f_X$  e  $f_Y$ , temos que  $f(x, y) = f_X(x)f_Y(y)$ , então,

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-t)f_Y(t) dt = \int_{-\infty}^{\infty} f_X(t)f_Y(z-t) dt = f_X * f_Y,$$

onde  $f_X * f_Y$  é conhecida como a *convolução das densidades*  $f_X$  e  $f_Y$ .

Vamos agora descrever o método do Jacobiano para funções mais gerais  $H$ . Suponha que  $G_0 \subseteq \mathbb{R}^n$ ,  $G \subseteq \mathbb{R}^n$  sejam regiões abertas, e que  $H : G_0 \rightarrow G$  seja uma bijeção entre  $G_0$  e  $G$ . Logo, existe a função inversa  $H^{-1}$  em  $G$ , de modo que  $\vec{X} = H^{-1}\vec{Y}$ . Suponha ainda que  $f$  é a densidade conjunta de  $\vec{X}$  e que  $P(\vec{X} \in G_0) = 1$ . Se as derivadas parciais de  $H^{-1}$  existirem e o Jacobiano  $J$  de  $H^{-1}$  for diferente de zero para todo  $\vec{y} \in G$ , podemos utilizar o teorema da mudança de variáveis e obter que para  $B \subseteq G$ ,  $B$  boreliano, temos

$$\begin{aligned} P(\vec{Y} \in B) &= P(\vec{X} \in H^{-1}(B)) = \int \cdots \int_{H^{-1}(B)} f(x_1, \dots, x_n) dx_1 \cdots dx_n \\ &= \int \cdots \int_B f(H_1^{-1}(y_1, \dots, y_n), \dots, H_n^{-1}(y_1, \dots, y_n)) |J| dy_1 \cdots dy_n. \end{aligned}$$

Como  $P(\vec{Y} \in G) = P(\vec{X} \in H^{-1}(G)) = P(\vec{X} \in G_0) = 1$ , temos que para todo boreliano  $B$  no  $\mathbb{R}^n$ ,

$$P(\vec{Y} \in B) = P(\vec{Y} \in B \cap G) = \int \cdots \int_{B \cap G} f(H_1^{-1}(y_1, \dots, y_n), \dots, H_n^{-1}(y_1, \dots, y_n)) |J| dy_1 \cdots dy_n.$$

Esta última integral é igual a integral sobre o conjunto  $B$  da função que toma o valor  $f(H_1^{-1}(y_1, \dots, y_n), \dots, H_n^{-1}(y_1, \dots, y_n)) |J|$  para  $\vec{y} \in G$ , e zero no caso contrário. Portanto, pela definição de densidade temos que

$$f_{\vec{Y}}(y_1, \dots, y_n) = \begin{cases} f(H_1^{-1}(y_1, \dots, y_n), \dots, H_n^{-1}(y_1, \dots, y_n)) |J|, & \text{se } \vec{y} \in G, \\ 0, & \text{caso contrário.} \end{cases}$$

**Observação 3.6.4:**

- (a) Note que  $J$  é o Jacobiano da função inversa  $H^{-1}$ , em alguns casos pode ser útil obter  $J$  a partir do Jacobiano  $J'$  da função  $H$  através da relação  $J = \frac{1}{J'}|_{\vec{x}=H^{-1}(\vec{y})}$ .
- (b) Para obter a distribuição de  $\vec{Y} = H(\vec{X})$  quando a dimensão de  $\vec{Y}$  é menor que a dimensão de  $\vec{X}$  muitas vezes é possível definir outras variáveis aleatórias  $Y'_1, \dots, Y'_m$ , utilizar o método do Jacobiano para determinar a densidade conjunta de  $\vec{Y}, Y'_1, \dots, Y'_m$  e, finalmente, obter a densidade marginal conjunta de  $\vec{Y}$ . Considere o seguinte exemplo:

**Exemplo 3.6.5:** Suponha que  $X_1, X_2$  tem densidade conjunta dada por  $f(x, y)$  e que estamos interessados na distribuição de  $Y_1 = X_1^2 + X_2$ . Como esta não é uma transformação 1-1, ela não possui inversa. Vamos definir uma nova variável  $Y_2 = X_1$  de modo que a função  $(Y_1, Y_2) = H(X_1, X_2) = (X_1^2 + X_2, X_1)$  possua uma função inversa diferenciável,  $(X_1, X_2) = H^{-1}(Y_1, Y_2) = (Y_2, Y_1 - Y_2^2)$ . Deste modo temos que

$$J = \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -2y_2 \end{pmatrix} = -1$$

Então temos que,  $f_{Y_1, Y_2}(y_1, y_2) = f(y_2, y_1 - y_2^2)$ . Finalmente, para encontrarmos  $f_{Y_1}$  integramos sobre todos os possíveis valores da variável  $Y_2$  que introduzimos:

$$f_{Y_1} = \int_{-\infty}^{\infty} f(y_2, y_1 - y_2^2) dy_2.$$

- (c) Podemos utilizar o método do Jacobiano em outros casos em que a função  $H$  não é 1-1. Para tanto, suponha que  $G, G_1, \dots, G_k$  sejam subregiões abertas do  $\mathbb{R}^n$  tais que  $G_1, \dots, G_k$  sejam disjuntas e  $P(\vec{X} \in \cup_{i=1}^k G_i) = 1$ , tais que a função  $H|_{G_i}$ , a restrição de  $H$  a  $G_i$ , seja um correspondência 1-1 entre  $G_i$  e  $G$ , para  $i = 1, \dots, k$ . Suponha que para todo  $i$ , a função inversa de  $H|_{G_i}$  satisfaça as hipóteses do caso anterior, e seja  $J_i$  o Jacobiano da inversa de  $H|_{G_i}$ . Pode-se provar que

$$f_{\vec{Y}}(y_1, \dots, y_n) = \begin{cases} \sum_{i=1}^k f(H|_{G_i}^{-1}(y_1, \dots, y_n)) |J_i|, & \text{se } \vec{y} \in G, \\ 0, & \text{caso contrário.} \end{cases}$$

# Capítulo 4

## Esperança e Momentos de Variáveis Aleatórias

### 4.1 O Conceito de Esperança

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

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

### 4.2 Definição da Esperança - Caso Discreto

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

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

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

**Definição 4.2.1:** Se  $X$  é uma variável aleatória discreta assumindo valores  $\{x_1, x_2, x_3, \dots\}$  com probabilidade  $\{p_1, p_2, p_3, \dots\}$ , respectivamente, então sua esperança é dada pela fórmula

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

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

**Exemplo 4.2.2:** Considere uma variável aleatória  $X$  tal que:  $P(X = -1) = 0.25$ ,  $P(X = 0) = 0.5$  e  $P(X = 2) = 0.25$ . Então,

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

**Exemplo 4.2.3:** Considere uma variável aleatória  $X$  tal que:  $P(X = -a) = P(X = a) = 1/2$ . Então,

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

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

**Exemplo 4.2.4: Aleatória.** Se  $X \in \{1, 2, \dots, n\}$  for uma variável aleatória com distribuição de probabilidade aleatória com parâmetro  $n$ , temos que sua esperança é dada por:

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

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

**Exemplo 4.2.5: Bernoulli.** Se  $X \in \{0, 1\}$  for uma variável aleatória com distribuição de probabilidade Bernoulli com parâmetro  $p$ , temos que sua esperança é dada por:

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

**Exemplo 4.2.6: Binomial.** Se  $X$  for uma variável aleatória com distribuição de probabilidade Binomial com parâmetros  $n$  e  $p$ , temos que sua esperança é dada por:

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

Onde utilizamos o Teorema Binomial na última igualdade.

**Exemplo 4.2.7: Geométrica.** Se  $X$  for uma variável aleatória com distribuição de probabilidade Geométrica com parâmetro  $\beta$ , temos que sua esperança é dada por:

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

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

**Exemplo 4.2.8: Binomial Negativa.** Se  $X$  for uma variável aleatória com distribuição de probabilidade Binomial Negativa com parâmetros  $r$  e  $p$ , temos que sua esperança é dada por:

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

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

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

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

**Exemplo 4.2.9: Poisson.** Se  $X$  for uma variável aleatória com distribuição de probabilidade Poisson com parâmetros  $\lambda$ , temos que sua esperança é dada por:

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

**Exemplo 4.2.10: Zeta.** Se  $X$  for uma variável aleatória com distribuição de probabilidade Zeta com parâmetro  $\alpha > 2$ , temos que sua esperança é dada por:

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

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

**Exemplo 4.2.11: Hipergeométrica.** Se  $X$  for uma variável aleatória com distribuição de probabilidade Hipergeométrica com parâmetro  $N, D, n$ , temos que sua esperança é dada por:

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

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

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

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

Antes de introduzirmos a definição geral da Esperança de uma variável aleatória qualquer, vamos estudar um pouco sobre as integrais de Riemann-Stieltjes e de Lebesgue-Stieltjes.

### 4.3 As integrais de Riemman-Stieltjes e de Lebesgue-Stieltjes

Antes de darmos as definições das integrais de Riemman-Stieltjes e Lebesgue-Stieltjes, vamos relembrar a definição da integral de Riemann. Uma partição  $P$  do intervalo  $[a, b]$  é uma seqüência de pontos  $\{x_1, \dots, x_n\}$  tal que  $a = x_1 < x_2 < \dots < x_n = b$ ; a norma da partição  $P$  é definida como sendo  $\max_{1 \leq i \leq n-1} x_{i+1} - x_i$ . Suponha que  $\varphi$  seja uma função real qualquer definida no intervalo  $[a, b]$ . Diz-se que esta função é Riemann integrável se as somas de Riemann

$$\sum_{i=1}^{n-1} \varphi(y_i)(x_{i+1} - x_i),$$

onde  $y_i \in [x_i, x_{i+1}]$ , convergem quando a norma de  $P$  tende a zero e este limite é independente da escolha dos  $y_i$ 's e da partição  $P$ . Se esta integral existe denota-se o limite por  $\int_a^b \varphi(x)dx$ .

A integral de Riemann-Stieltjes é uma generalização de integral de Riemann. Se  $\varphi$  é uma função contínua definida no intervalo  $[a, b]$  e  $F$  é uma função de distribuição, define-se a integral de Riemann-Stieltjes de  $\varphi$  em  $[a, b]$ , em relação a  $F$ , como o limite de somas de Riemann da forma

$$\sum_{i=1}^{n-1} \varphi(y_i)[F(x_{i+1}) - F(x_i)],$$



onde  $a = x_1 < x_2 < \dots < x_n = b$ ,  $y_i$  é um ponto arbitrário de  $[x_i, x_{i+1}]$ , e toma-se o limite quando a norma de partição  $P$  tende a zero. Tal limite existe e é finito sob as condições descritas, e é representado por

$$\int_a^b \varphi(x) dF(x).$$

A função  $\varphi$  é chamada de integrando e  $F$  de integrador. O limite acima existe mesmo que  $F$  não seja uma função de distribuição basta que ela seja de variação limitada.

**Definição 4.3.1:** Define-se variação total de uma função  $f$  em  $[a, b]$  pelo funcional:

$$V(f, [a, b]) = \sup \sum_{i=1}^n |f(x_{i+1}) - f(x_i)|,$$

onde o supremo é tomado sobre todas as possíveis partições do intervalo fechado  $[a, b]$ . Uma função é de variação limitada se  $V(f, [a, b]) < \infty$ . ■

A integral de Riemann-Stieltjes sobre a reta é uma integral imprópria definida da mesma maneira que a integral imprópria de Riemann:

$$\int_{-\infty}^{\infty} \varphi(x) dF(x) = \lim_{a \rightarrow -\infty, b \rightarrow \infty} \int_a^b \varphi(x) dF(x),$$

se o limite existe. Esta definição da integral de Riemann-Stieltjes pode ser estendida a outras funções  $\varphi$  além das contínuas. Para uma função qualquer  $\varphi$ , define-se  $\int_a^b \varphi(x) dF(x)$  como sendo o limite das somas de Riemann descritas acima quando a norma da partição tende a zero, se este limite existe e é independente das escolhas dos  $y_i$ 's e da partição  $P$ . O problema é que mesmo para funções bem simples este limite pode não existir como mostra o próximo exemplo:

**Exemplo 4.3.2:** Seja  $F_0(x) = 1$  se  $x \geq 0$ , e  $F_0(x) = 0$ , caso contrário. Consideremos a integral de Riemann-Stieltjes de  $F_0$  em  $[-1, 1]$  em relação a  $F_0$ . Note que se zero não é um dos pontos da partição, de modo que  $x_i < 0 < x_{i+1}$  para algum  $i$ , com  $F_0(x_{i+1}) - F_0(x_i) = 1$ , então o somatório assume como valor escolhido para  $y_i$  ser maior que 0, ou não.

Uma integral mais robusta que não sofre desta deficiência é a integral de Lebesgue-Stieltjes. A idéia da integral de Lebesgue-Stieltjes é particionar a imagem da função  $\varphi$  ao invés de particionar o seu domínio. Diz-se que uma partição  $P'$  é um refinamento de  $P$  se  $P \subseteq P'$ , ou seja, quando os intervalos da partição  $P$  são particionados na partição  $P'$ . Suponha que  $\varphi$  seja não negativa e mensurável em relação a  $\sigma$ -álgebra de Borel. Seja  $\mu$  for uma medida nos reais, ou seja, uma função cujo domínio é a  $\sigma$ -álgebra de Borel que tem como imagem do conjunto vazio zero, é não-negativa e  $\sigma$ -aditiva. Dada uma seqüência  $\{P_1, P_2, \dots\}$  de partições de  $[0, \infty)$  onde  $P_n = \{y_1, y_2, \dots, y_n\}$ ,  $y_n \rightarrow \infty$ ,  $P_{i+i}$  é um refinamento de  $P_i$ , e a norma de  $P_n$  tende a zero quando  $n \rightarrow \infty$ , define-se a soma de Lebesgue em relação a partição  $P_n$  como sendo,

$$\sum_{i=1}^{n-1} y_i \mu(\{x : y_i \leq \varphi(x) < y_{i+1}\}) + y_n \mu(\{x : \varphi(x) \geq y_n\}).$$

A integral de Lebesgue-Stieltjes de  $\varphi$  em relação a  $\mu$  é definida como sendo igual ao limite das somas de Lebesgue, quando  $n \rightarrow \infty$ . Dadas as condições acima este limite sempre existe (pode ser  $+\infty$ ) e é denotado por  $\int \varphi d\mu$ .

Para uma função mensurável  $\varphi$  qualquer, podemos escrever  $\varphi = \varphi^+ - \varphi^-$ , onde  $\varphi^+ = \max(\varphi, 0)$ , a parte positiva de  $\varphi$ , e  $\varphi^- = -\min(\varphi, 0)$ , o módulo da parte negativa de  $\varphi$ , são funções não-negativas e portanto possuem integral de Lebesgue-Stieltjes. Se  $\varphi^+$  ou  $\varphi^-$  possui integral de Lebesgue-Stieltjes finita em relação a  $\mu$ , define-se a integral de Lebesgue-Stieltjes de  $\varphi$  em relação a  $\mu$  como sendo

$$\int \varphi d\mu = \int \varphi^+ d\mu - \int \varphi^- d\mu.$$

Se  $\mu$  for uma medida de probabilidade em  $(\mathbb{R}, \mathcal{B})$  e  $F$  for a distribuição de probabilidade acumulada associada a variável aleatória  $X(\omega) = \omega$ , então escreve-se  $\int \varphi(x) dF(x)$  (ou simplesmente,  $\int \varphi dF$ ) para denotar  $\int \varphi d\mu$ . Em geral, usa-se a notação  $\int \varphi(x) dF(x)$  não somente para funções de distribuição, mas para qualquer função  $F$  que pode ser escrita como a diferença de duas funções monótonas não-decrescentes, limitadas e contínuas à direita. Se  $G$  for uma função monótona não-decrescente, limitada e contínua à direita, então dado um intervalo qualquer  $I = [x_1, x_2]$ , defina  $\nu(I) = G(x_2) - G(x_1)$ , então usa-se a notação  $\int \varphi(x) dG(x)$  para denotar a integral  $\int \varphi(x) d\nu$ , onde  $\nu$  é a única medida que satisfaz  $\nu(I) = G(x_2) - G(x_1)$  para todo intervalo  $I$ . Desta forma, se  $F = G_1 - G_2$ , onde  $G_1$  e  $G_2$  são funções monótonas não-decrescentes, limitadas e contínuas à direita, então  $\int \varphi(x) dF(x)$  é utilizado para denotar  $\int \varphi(x) dG_1(x) - \int \varphi(x) dG_2(x)$ .

Dada um intervalo qualquer  $[a, b]$ , define-se a integral de Lebesgue-Stieltjes de  $\varphi$  em relação a  $\mu$  no intervalo  $[a, b]$  como sendo  $\int \varphi I_{[a,b]} d\mu$  e denota-se por  $\int_a^b \varphi d\mu$ .

### 4.3.1 Propriedades da Integral de Lebesgue-Stieltjes

P1. Quando o integrando é contínuo, a integral de Lebesgue-Stieltjes torna-se uma integral de Riemman-Stieltjes.

P2.  $\int_a^b dF = F(b) - F(a)$ . Análoga ao teorema fundamental do cálculo:  $\int_a^b \varphi'(x) dx = \varphi(b) - \varphi(a)$ , onde  $\varphi(x)$  é a derivada de  $\varphi$ .

P3. Linearidade no integrando e no integrador. Se  $\varphi(x) = \alpha f(x) + \beta g(x)$ , temos

$$\int \varphi dF = \alpha \int f dF + \beta \int g dF,$$

e para  $H(x) = \alpha F(x) + \beta G(x)$ , temos

$$\int \varphi dH = \alpha \int \varphi dF + \beta \int \varphi dG.$$

P4. Aditividade. Se  $-\infty \leq a < b < c \leq \infty$ , então

$$\int_a^c \varphi dF = \int_a^b \varphi dF + \int_b^c \varphi dF.$$

P5. Se  $F$  for a função de distribuição de uma variável aleatória discreta, ou seja, se

$$F(x) = \sum_{i=1}^{\infty} p_i U(x - x_i),$$

onde  $P(X = x_i) = p_i$  e  $\sum_{i=1}^{\infty} p_i = 1$ , então

$$\int \varphi dF = \sum_{i=1}^{\infty} p_i \varphi(x_i).$$

P6. Se  $F$  for a função de distribuição de uma variável aleatória contínua, tendo densidade  $f$ , temos  $\frac{dF(x)}{dx} = f(x)$  em quase toda parte, e conseqüentemente,

$$\int \varphi(x) dF(x) = \int \varphi(x) f(x) dx.$$

P7. No caso de uma distribuição geral  $F$ , vimos que  $F$  pode ser decomposta em suas partes discreta, contínua e singular da seguinte forma  $F = F_d + F_{ac} + F_s$ , então por linearidade do integrador:

$$\int \varphi(x) dF(x) = \int \varphi(x) dF_d(x) + \int \varphi(x) dF_{ac}(x) + \int \varphi(x) dF_s(x).$$

Se a parte singular for nula,  $F_s(x) = 0, \forall x$ , então:

$$\int \varphi(x) dF(x) = \sum_i \varphi(x_i) p_i + \int \varphi(x) f(x) dx,$$

onde  $p_i$  é o salto de  $F$  em  $x_i$  e  $f$  é a derivada de  $F$ .

## 4.4 Definição da Esperança - Caso Geral

Vamos agora motivar a definição da Esperança no caso geral. Consideremos uma seqüência  $\{P_1, P_2, \dots\}$  de partições de  $[0, \infty)$  onde  $P_n = \{y_1, y_2, \dots, y_n\}$ ,  $y_n \rightarrow \infty$ ,  $P_{i+i}$  é um refinamento de  $P_i$ , e a norma de  $P_n$  tende a zero quando  $n \rightarrow \infty$ . Dada uma variável aleatória não-negativa qualquer  $X$  e uma partição  $P_n$  desta seqüência, definamos uma outra variável aleatória  $Y$  discreta que aproxima  $X$  assumindo o valor  $y_i$  quando  $y_i \leq X < y_{i+1}$  e  $Y = y_n$  se  $X \geq y_n$ , ou seja,  $Y = \sum_{i=1}^{n-1} y_i I_{[y_i \leq X < y_{i+1}]} + y_n I_{[X \geq y_n]}$ . Como  $Y$  é discreta temos que sua esperança é dada por

$$EY = \sum_{i=1}^n y_i P(Y = y_i) = \sum_{i=1}^{n-1} y_i P(y_i \leq X < y_{i+1}) + y_n P(X \geq y_n).$$

Note que esta esperança é uma soma de Lebesgue em relação a partição  $P_n$  com integrando  $X$  e função integradora dada pela medida de probabilidade  $P$ . Note que a medida que pegamos

partições mais refinadas na seqüência,  $Y$  se torna cada vez uma melhor aproximação para  $X$ . Já que os valores de  $X$  e  $Y$  ficam cada vez mais próximos é intuitivo requerer que nossa definição de esperança (média)  $EX$  seja igual ao limite de  $EY$  quando  $n \rightarrow \infty$ , ou seja

$$EX = \lim_{n \rightarrow \infty} \sum_{i=1}^n y_i P(Y = y_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^{n-1} y_i P(y_i \leq X < y_{i+1}) + y_n P(X \geq y_n) = \int X dP.$$

Logo,  $EX$  é definida como sendo a integral de Lebesgue-Stieltjes de  $X$  em relação a medida de probabilidade  $P$ , ou similarmente,  $EX = \int X dF$ , onde  $F$  é a função de distribuição acumulada de  $X$ . No caso geral, temos a seguinte definição

**Definição 4.4.1:** Se  $X$  é uma variável aleatória com função de distribuição  $F$ , então sua esperança é dada pela fórmula

$$EX = \int X dF = \int_{-\infty}^0 X dF + \int_0^{\infty} X dF,$$

desde que pelo menos uma das integrais seja finita. Em caso as duas integrais não sejam finitas, a esperança não existe. Caso  $EX$  seja finita, diz-se que  $X$  é integrável.

Pela Propriedade P7 da integral de Lebesgue-Stieltjes, temos que se  $F = F_d + F_{ac} + F_s$ , então

$$EX = \int X dF = \sum_i x_i p_i + \int x f(x) dx + \int x dF_s(x),$$

onde  $p_i$  é o salto de  $F$  em  $x_i$  e  $f$  é a derivada de  $F$ . Como a parte singular costuma ser nula, na prática a esperança reduz-se a uma série e/ou uma integral imprópria, usualmente de Riemann se  $f$  for integrável a Riemann.

**Exemplo 4.4.2:** Considere uma variável aleatória  $Y$  com função de distribuição  $F$ , tal que

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

Decompondo em parte discreta e contínua tem-se

$$F_d(x) = \begin{cases} 0 & , \text{ se } x < 1/2 \\ 1/2 & , \text{ se } x \geq 1/2, \end{cases}$$

e

$$F_{ac}(x) = \begin{cases} 0 & , \text{ se } x < 0 \\ x & , \text{ se } 0 \leq x < 1/2 \\ 1/2 & , \text{ se } x \geq 1/2. \end{cases}$$

Portanto,

$$EY = \frac{1}{2}P(Y = \frac{1}{2}) + \int_0^{1/2} y dy = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}.$$

**Exemplo 4.4.3: Uniforme.** Se  $X \sim U(a, b)$ , então  $X$  possui densidade igual a  $f(x) = \frac{1}{b-a}$  se  $x \in (a, b)$ , e  $f(x) = 0$ , caso contrário. Logo, temos que sua esperança é dada por:

$$EX = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}.$$

**Exemplo 4.4.4: Exponencial.** Se  $X \sim Exp(\lambda)$ , então  $X$  possui densidade igual a  $f_X(x) = \lambda e^{-\lambda x} U(x)$ . Logo, temos que sua esperança é dada por:

$$EX = \int_0^{\infty} x \lambda e^{-\lambda x} dx = -x e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = \frac{-e^{-\lambda x}}{\lambda} \Big|_0^{\infty} = \frac{1}{\lambda}.$$

**Exemplo 4.4.5: Normal.** Se  $X \sim \mathcal{N}(m, \sigma)$ , então  $X$  possui densidade igual a  $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$ . Logo, temos que sua esperança é dada por:

$$EX = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx.$$

Fazendo a mudança de variável  $y = \frac{x-m}{\sigma}$ , temos

$$EX = \int_{-\infty}^{\infty} \frac{\sigma y + m}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \int_{-\infty}^{\infty} \frac{\sigma y}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy + \int_{-\infty}^{\infty} \frac{m}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 0 + m = m.$$

**Exemplo 4.4.6: Cauchy.** Se  $X \sim Cauchy(a)$ , então  $X$  possui densidade igual a  $f_X(x) = \frac{1}{\pi} \cdot \frac{a}{a^2+x^2}$ . Neste caso  $X$  não é integrável, ou seja  $EX$  não está definida, pois:

$$\int_{-\infty}^0 \frac{x}{\pi} \cdot \frac{a}{a^2+x^2} dx = -\infty, \text{ e}$$

$$\int_0^{\infty} \frac{x}{\pi} \cdot \frac{a}{a^2+x^2} dx = \infty.$$

#### 4.4.1 Interpretação Geométrica da Esperança

Por definição,  $EX = \int x dF(x)$ , ou seja,  $EX$  é a integral da diferencial  $x dF$ . Mas  $x dF$  é uma diferencial de área. Para  $x > 0$ ,  $x dF$  é uma diferencial da área da região compreendida entre as curvas  $x = 0$ ,  $y = 1$ , e  $y = F(x)$  no plano Euclideano, cuja área total é dada por  $\int_0^{\infty} (1 - F(x)) dx$ . Para  $x < 0$ ,  $-x dF$  é uma diferencial da área da região compreendida entre as curvas  $x = 0$ ,  $y = 0$ , e  $y = F(x)$  no plano Euclideano, cuja área total é dada por  $\int_{-\infty}^0 F(x) dx$ . Logo, temos que  $EX = \int_0^{\infty} (1 - F(x)) dx - \int_{-\infty}^0 F(x) dx$ . Formalmente, podemos provar isso da seguinte maneira. A prova será dividida em duas etapas: (a)  $\int_0^{\infty} x dF(x) = \int_0^{\infty} (1 - F(x)) dx$  e (b)  $\int_{-\infty}^0 x dF(x) = -\int_{-\infty}^0 F(x) dx$ . Começemos provando (b). Utilizando integração por partes, temos que  $\forall a < 0$ ,

$$\int_a^0 x dF(x) = -aF(a) - \int_a^0 F(x) dx = \int_a^0 [F(a) - F(x)] dx.$$

Como  $F(a) \geq 0$  e  $a < 0$ , temos

$$\int_a^0 x dF(x) \geq - \int_a^0 F(x) dx.$$

Como a desigualdade é válida para todo  $a < 0$ , temos que tomando o limite quando  $a \rightarrow -\infty$

$$\int_{-\infty}^0 x dF(x) \geq - \int_{-\infty}^0 F(x) dx.$$

Por outro lado, seja  $\lambda < 0$ . Se  $a < \lambda$ , então

$$\int_a^0 [F(a) - F(x)] dx \leq \int_\lambda^0 [F(a) - F(x)] dx = F(a)(-\lambda) - \int_\lambda^0 F(x) dx,$$

e portanto, tomando o limite quando  $a \rightarrow -\infty$ , temos que

$$\int_{-\infty}^0 x dF(x) \leq - \int_\lambda^0 F(x) dx.$$

Como isto é válido para todo  $\lambda < 0$ , tomando o limite quando  $\lambda \rightarrow -\infty$ , temos

$$\int_{-\infty}^0 x dF(x) \leq - \int_{-\infty}^0 F(x) dx,$$

como queríamos demonstrar.

Para parte (a), utilizando integração por partes, temos que  $\forall b > 0$ ,

$$\int_0^b x dF(x) = bF(b) - \int_0^b F(x) dx = \int_0^b [F(b) - F(x)] dx.$$

Como  $F(b) \leq 1$  e  $1 - F(x) \geq 0$ , temos

$$\int_0^b x dF(x) = \int_0^b [F(b) - F(x)] dx \leq \int_0^\infty [1 - F(x)] dx.$$

Como a desigualdade é válida para todo  $b > 0$ , temos que tomando o limite quando  $b \rightarrow \infty$

$$\int_0^\infty x dF(x) \leq \int_0^\infty [1 - F(x)] dx.$$

Por outro lado, seja  $\lambda > 0$ . Se  $b > \lambda$ , então

$$\begin{aligned} \int_0^b [F(b) - F(x)] dx &\geq \int_0^\lambda [F(b) - F(x)] dx = \int_0^\lambda [F(b) - 1] dx + \int_0^\lambda [1 - F(x)] dx \\ &= \lambda[F(b) - 1] + \int_0^\lambda [1 - F(x)] dx, \end{aligned}$$

e portanto, tomando o limite quando  $b \rightarrow \infty$ , temos que

$$\int_0^\infty x dF(x) \geq \int_0^\lambda [1 - F(x)] dx.$$

Como isto é válido para todo  $\lambda > 0$ , tomando o limite quando  $\lambda \rightarrow \infty$ , temos

$$\int_0^\infty x dF(x) \geq \int_0^\infty [1 - F(x)] dx,$$

como queríamos demonstrar.

## 4.5 Esperança de Funções de Variáveis Aleatórias

Vamos iniciar considerando o caso discreto.

### 4.5.1 Caso Discreto

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

**Definição 4.5.1:** Seja  $X$  uma variável aleatória discreta e seja  $Y = H(X)$ . Se  $Y$  assumir os seguintes valores  $y_1, y_2, \dots$  e se  $p(y_i) = P(Y = y_i)$ , definimos:

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

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

**Teorema 4.5.2:** *Seja  $X$  uma variável aleatória discreta assumindo os valores  $x_1, x_2, \dots$  e seja  $Y = H(X)$ . Se  $p(x_i) = P(X = x_i)$ , temos*

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

**Prova:** Vamos re-ordenar o somatório  $\sum_{i=1}^{\infty} H(x_i) p(x_i)$ , agrupando os termos onde  $x_i$  tem a mesma imagem de acordo com a função  $H$ , ou seja, sejam  $x_{i_1}, x_{i_2}, \dots$ , todos os valores  $x_i$  tal que  $H(x_{ij}) = y_i$  para  $j \geq 1$ , onde  $y_1, y_2, \dots$  são os possíveis valores de  $Y$ . Desse modo podemos reescrever

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

■

**Exemplo 4.5.3:** Suponha que  $X$  é uma variável aleatória Poisson com parâmetro  $\lambda$ . Seja  $Y = X^2$ , vamos calcular  $EY$ . Utilizando o Teorema 4.5.2, temos

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

Também podemos estender este resultado para o caso de uma função real de um vetor aleatório. Neste caso, se  $Y = H(\vec{X})$ , temos que  $EY = \sum_i H(\vec{x}_i) p_{\vec{X}}(\vec{x}_i)$ , onde os  $\vec{x}_i$  são os valores assumidos pelo vetor aleatório  $\vec{X}$ .

### 4.5.2 Caso Geral

No caso de uma variável aleatória qualquer  $X$  também podemos calcular a esperança de uma função  $Y = \varphi(X)$  de forma similar.

**Teorema 4.5.4:** *Seja  $X$  uma variável aleatória qualquer,  $Y = \varphi(X)$  uma outra variável aleatória, então*

$$EY = \int y dF_Y(y) = \int \varphi(x) dF_X(x),$$

desde que estas integrais existam.

**Prova:** A prova no caso geral envolve Teoria da Medida e será omitida. ■

Uma fórmula análoga também é válida quando consideramos funções de vetores aleatórios.

**Teorema 4.5.5:** *Seja  $\vec{X}$  um vetor aleatório e  $Y = \varphi(\vec{X})$  uma variável aleatória. Então,*

$$EY = \int y dF_Y(y) = \int \varphi dF_{\vec{X}}.$$

## 4.6 Propriedades da Esperança

As seguintes propriedades são aplicações imediatas da definição de esperança:

1.  $P(X = c) = 1 \Rightarrow EX = c$ .
2.  $P(X \geq 0) = 1 \Rightarrow EX \geq 0$ .
3.  $E(aX) = aEX$ , onde  $a$  um número real qualquer. Esta propriedade segue facilmente da expressão da esperança de uma função de variável aleatória.
4.  $E(X + Y) = EX + EY$ . No caso discreto, note que

$$\begin{aligned} E(X + Y) &= \sum_i \sum_j (x_i + y_j) p(x_i, y_j) = \sum_i x_i \sum_j p(x_i, y_j) + \sum_i \sum_j y_j p(x_i, y_j) \\ &= \sum_i x_i p(x_i) + \sum_j y_j \sum_i p(x_i, y_j) = EX + \sum_j y_j p(y_j) = EX + EY. \end{aligned}$$

No caso geral, temos que

$$E(X + Y) = E(\varphi(X, Y)) = \int \int (x + y) dF_{X,Y}(x, y),$$

e pela linearidade da integral obtemos

$$E(X + Y) = \int \int x dF_{X,Y}(x, y) + \int \int y dF_{X,Y}(x, y) = EX + EY.$$



**Corolário 4.6.1:**  $E(\sum_i^n a_i X_i) = \sum_i^n a_i EX_i$ .

**Prova:** Aplicação das duas propriedades anteriores e indução matemática. ■

5.  $P(X \geq Y) = 1 \Rightarrow EX \geq EY$ . Propriedade 5 segue da propriedades 2 e do corolário anterior, pois

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

o que, pela propriedade 2, implica que  $E(X - Y) \geq 0$ . Pelo corolário, temos que  $E(X - Y) = EX - EY$ , ou seja podemos concluir que  $EX - EY \geq 0$ .

6. Se  $\{X_1, \dots, X_n\}$  são variáveis aleatórias mutuamente independentes, então  $E(\prod_{i=1}^n X_i) = \prod_{i=1}^n EX_i$ . Provaremos esta propriedade nos casos discreto e contínuo. No caso discreto, note que

$$\begin{aligned} E\left(\prod_{i=1}^n X_i\right) &= \sum_{i_1} \cdots \sum_{i_n} x_{i_1} \cdots x_{i_n} p(x_{i_1}, \dots, x_{i_n}) \\ &= \sum_{i_1} \cdots \sum_{i_n} x_{i_1} \cdots x_{i_n} \prod_{j=1}^n p(x_{i_j}) = \sum_{i_1} x_{i_1} p(x_{i_1}) \cdots \sum_{i_n} x_{i_n} p(x_{i_n}) = \prod_{i=1}^n EX_i. \end{aligned}$$

No caso contínuo, temos que  $f_{\vec{X}}(\vec{x}) = \prod_{i=1}^n f_{X_i}(x_i)$ , logo

$$\begin{aligned} E\left(\prod_{i=1}^n X_i\right) &= \int \cdots \int x_1 \cdots x_n f_{\vec{X}}(\vec{x}) dx_1 \cdots dx_n \\ &= \int \cdots \int \prod_{i=1}^n x_i f_{X_i}(x_i) dx_1 \cdots dx_n = \prod_{i=1}^n \int x_i f_{X_i}(x_i) dx_i = \prod_{i=1}^n EX_i. \end{aligned}$$

De maneira análoga, pode-se provar a seguinte generalização deste resultado: se  $\{X_1, \dots, X_n\}$  são variáveis aleatórias mutuamente independentes, então  $E(\prod_{i=1}^n G(X_i)) = \prod_{i=1}^n EG(X_i)$ .

7. Se  $Y$  for uma variável aleatória que assume valores inteiros não-negativos, temos que

$$EY = \sum_{k=1}^{\infty} kP(Y = k) = \sum_{k=1}^{\infty} \sum_{j=1}^k P(Y = k),$$

trocando a ordem dos somatórios:

$$EY = \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} P(Y = k) = \sum_{j=1}^{\infty} P(Y \geq j).$$

8. (Desigualdade de Jensen) Seja  $\varphi$  uma função mensurável e convexa definida na reta. Se  $X$  é integrável, então  $E\varphi(X) \geq \varphi(EX)$ .

**Prova:** Pela convexidade de  $\varphi$ , dado algum ponto  $(x_0, \varphi(x_0))$  do gráfico de  $\varphi$ , existe uma reta que passa por esse ponto e fica sempre abaixo do gráfico de  $\varphi$ , ou seja, existe algum  $\lambda$  tal que

$$\varphi(x) \geq \varphi(x_0) + \lambda(x - x_0), \forall x.$$

Logo, pela monotonicidade e linearidade da esperança, temos

$$E\varphi(X) \geq \varphi(x_0) + \lambda(EX - x_0).$$

Em particular, para  $x_0 = EX$ , temos  $E\varphi(X) \geq \varphi(EX)$ . ■

O próximo Lema estabelece um critério para integrabilidade de variáveis aleatórias.

**Lema 4.6.2:** *Seja  $X$  uma variável aleatória qualquer. Então,*

$$\sum_{n=1}^{\infty} P(|X| \geq n) \leq E|X| \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n),$$

e, portanto,  $X$  é integrável se, e somente se,  $\sum_{n=1}^{\infty} P(|X| \geq n) < \infty$ .

**Prova:** Se  $x \geq 0$ , seja  $[x]$  a parte inteira de  $x$ . Então, a variável aleatória  $[|X|]$  assume o valor  $k$  quando  $k \leq |X| < k+1$  e  $0 \leq [|X|] \leq |X| \leq [|X|] + 1$ , então pela monotonicidade e linearidade da esperança temos:

$$0 \leq E[|X|] \leq E|X| \leq 1 + E[|X|].$$

Como  $[|X|]$  é uma variável aleatória que só assume valores inteiros não-negativos, temos

$$E[|X|] = \sum_{n=1}^{\infty} P([|X|] \geq n) = \sum_{n=1}^{\infty} P(|X| \geq n),$$

logo

$$\sum_{n=1}^{\infty} P(|X| \geq n) \leq E|X| \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n).$$

■

Se  $X^+ = \max(X, 0)$  e  $X^- = -\min(X, 0)$ , temos que  $X = X^+ - X^-$  e  $|X| = X^+ + X^-$ . Por definição, temos que  $EX < \infty$  se, e somente se,  $EX^+ < \infty$  e  $EX^- < \infty$ . Portanto, vemos que  $EX < \infty$  se, e somente se,  $E|X| < \infty$ . De forma análoga, pode-se concluir que  $E\varphi(X) < \infty$  se, e somente se,  $E|\varphi(X)| < \infty$  para qualquer função mensurável  $\varphi$ . O próximo teorema nos dá um outro critério para integrabilidade de uma variável aleatória.

**Teorema 4.6.3:** *Sejam  $X$  e  $Y$  variáveis aleatórias tais que  $Y \geq 0$ ,  $Y$  é integrável, e  $|X| < Y$ . Então,  $X$  é integrável.*

**Prova:** Note que  $0 \leq |X| \leq Y$  implica que  $0 \leq E|X| \leq EY$ . Portanto, se  $EY < \infty$ , temos que  $E|X| < \infty$ , o que por sua vez implica que  $EX < \infty$ . ■

## 4.7 Momentos

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

**Definição 4.7.1:** Para qualquer inteiro não-negativo  $n$ , o  $n$ -ésimo momento da variável aleatória  $X$  é  $EX^n$ , se esta esperança existe.

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

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

**Teorema 4.7.2:** Se o  $k$ -ésimo momento de uma variável aleatória existir, então todos os momentos de ordem menores do que  $k$  também existem.

**Prova:** Por hipótese, temos que  $E|X^k| < \infty$ , logo  $E(1 + |X^k|) < \infty$ . Como para qualquer  $j$  tal que  $0 < j < k$ ,  $|X^j| \leq 1 + |X^k|$ , e  $1 + |X^k|$  é integrável, temos que  $|X^j|$  também é integrável. ■

Vamos agora enunciar dois teoremas importantes que tratam da convergência de esperanças de variáveis aleatórias. Neste caso, estaremos tratando de convergência pontual de variáveis aleatórias, ou seja,  $X_n \rightarrow X$  se, e somente se,  $X_n(\omega) \rightarrow X(\omega)$  para todo  $\omega \in \Omega$ . No próximo capítulo, veremos outras noções de convergência de variáveis aleatórias.

**Teorema 4.7.3: Teorema da Convergência Monótona.** Sejam  $X, X_1, X_2, \dots$  variáveis aleatórias. Se  $0 \leq X_n \uparrow X$ , então,  $EX_n \uparrow EX$ .

**Teorema 4.7.4: Teorema da Convergência Dominada.** Sejam  $Y, X, X_1, X_2, \dots$  variáveis aleatórias. Considere que  $Y$  seja integrável,  $|X_n| \leq Y$  e  $X_n \rightarrow X$ . Assim  $X$  e  $X_n$  são integráveis e  $EX_n \rightarrow EX$ .

O próximo exemplo mostra que nem sempre  $X_n \rightarrow X \Rightarrow EX_n \rightarrow EX$ .

**Exemplo 4.7.5:** Seja  $Y \sim U(0, 1)$ . Considere a seguinte seqüência  $\{X_1, X_2, \dots\}$  de variáveis aleatórias:  $X_n(\omega) = n$  se  $Y(\omega) \in (0, 1/n)$  e  $X_n(\omega) = 0$  em caso contrário. Então, temos que  $X_n(\omega) \rightarrow 0, \forall \omega$ . Mas,  $EX_n = 1 \neq 0 = E0$ , ou seja,  $EX_n \not\rightarrow 0$ .

### 4.7.1 Momentos Centrais

**Definição 4.7.6:** Se  $X$  é uma variável aleatória seu  $n$ -ésimo momento central é:  $E(X - EX)^n$ , se esta esperança existir.

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

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

Do Teorema Binomial e da linearidade da esperança, temos

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

e

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

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

**Exemplo 4.7.7:** Considere uma variável aleatória  $X$  tal que

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

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

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

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

### Propriedades da Variância e de outros Momentos

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

1.  $VarX \geq 0$ .
2. Se  $X = c$ ,  $Var(X) = 0$ .

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

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

**Prova:**

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

■

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

**Prova:**

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

■

5. Se  $X$  e  $Y$  forem variáveis aleatórias mutuamente independentes, então  $Var(X + Y) = VarX + VarY$ .

**Prova:**

$$\begin{aligned} Var(X + Y) &= E(X + Y)^2 - [E(X + Y)]^2 \\ &= E(X^2 + 2XY + Y^2) - (EX)^2 - 2EXEY - (EY)^2 \\ &= EX^2 + EY^2 - (EX)^2 - (EY)^2 + 2E(XY) - 2EXEY = VarX + VarY \end{aligned}$$

■

6. Se  $X_1, \dots, X_n$  são variáveis aleatórias independentes, então  $Var(X_1 + \dots + X_n) = VarX_1 + \dots + VarX_n$ . Esta propriedade segue da propriedade anterior e de uma aplicação de indução matemática.

7. **Desigualdade de Chebyshev Generalizada.** Dado um conjunto  $A$  e uma função  $g(x)$  tal que  $\forall x g(x) \geq I_A(x)$ , tem-se que  $P(X \in A) \leq \min(1, Eg(X))$ .

**Prova:** Pela monotonicidade da Esperança, temos que  $Eg(X) \geq EI_A(X) = P(X \in A)$ . Mas, como a cota superior pode exceder 1, temos que  $\min(1, Eg(X)) \geq P(X \in A)$ .

■

**Corolário 4.7.8:** *Seja  $X$  uma variável aleatória, então para todo  $\epsilon > 0$ ,  $P(|X| \geq \epsilon) \leq \frac{E|X|}{\epsilon}$ .*

**Prova:** Escolha  $A = \{x : |x| \geq \epsilon\}$  e  $g(x) = \frac{|x|}{\epsilon}$ . Note que  $g(x) \geq I_A(x)$ , então  $P(|X| \geq \epsilon) \leq \frac{E|X|}{\epsilon}$ . ■

**Corolário 4.7.9:** *Se  $Z \geq 0$  e  $EZ = 0$ , então  $P(Z = 0) = 1$ .*

**Prova:**  $P(Z \geq \frac{1}{n}) \leq nEZ = 0$ . Como  $[Z > 0] = \cup_n [Z \geq \frac{1}{n}]$ , temos que

$$P(Z > 0) = P(\cup_n [Z \geq \frac{1}{n}]) \leq \sum_n P(Z \geq \frac{1}{n}) = 0.$$

Portanto,  $P(Z = 0) = 1 - P(Z > 0) = 1$ . ■

Note que este último corolário implica que, quando  $Var(X) = 0$ , ou seja  $E(X - EX)^2 = 0$ , temos que  $P(X = EX) = 1$ , ou seja  $X$  é constante com probabilidade 1.

**Corolário 4.7.10: Desigualdade (Original) de Chebyshev.** *Seja  $X$  uma variável aleatória, então  $P(|X - EX| \geq \epsilon) \leq \frac{VarX}{\epsilon^2}$ .*

**Prova:** Escolha  $A = \{x : |x| \geq \epsilon\}$  e  $g(x) = \frac{x^2}{\epsilon^2}$ . Note que  $g(x) \geq I_A(x)$ , então pelo teorema anterior,  $P(X \in A) = P(|X| \geq \epsilon) \leq \frac{EX^2}{\epsilon^2}$ . Substituindo  $X$  por  $X - EX$ , temos  $P(|X - EX| \geq \epsilon) \leq \frac{VarX}{\epsilon^2}$ . ■

Note que a desigualdade de Chebyshev converte conhecimento sobre um momento de segunda ordem ou uma variância numa cota superior para a probabilidade da cauda de uma variável aleatória.

8. Se  $X$  e  $Y$  são variáveis aleatórias em  $(\Omega, \mathcal{A}, P)$  tais que  $E|X|^t < \infty$  e  $E|Y|^t < \infty$ , então  $E|X + Y|^t < \infty$ . **Prova:**  $|X + Y| \leq |X| + |Y| \leq 2 \max(|X|, |Y|)$ . Portanto,  $|X + Y|^t \leq 2^t \max(|X|^t, |Y|^t) \leq 2^t(|X|^t + |Y|^t)$ . Logo,  $E|X + Y|^t \leq 2^t(E|X|^t + E|Y|^t) < \infty$ . ■

Como  $E|X|^t < \infty$  obviamente implica  $E|aX|^t < \infty, \forall a \in \mathbb{R}$ , esta propriedade diz que a classe de variáveis aleatórias em  $(\Omega, \mathcal{A}, P)$  possuidoras do  $t$ -ésimo momento finito é um espaço vetorial ou espaço linear.

9.  $VarX = E(X - \mu)^2 = \min_{c \in \mathbb{R}} E(X - c)^2$ .

**Prova:**

$$(X - c)^2 = (X - \mu + \mu - c)^2 = (X - \mu)^2 + 2(\mu - c)(X - \mu) + (\mu - c)^2,$$

logo

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

Portanto,  $E(X - c)^2 \geq E(X - \mu)^2, \forall c \in \mathbb{R}$ . ■

## 4.8 Momentos Conjuntos

Podemos definir a noção de momento quando lidamos com vetores aleatórios.

**Definição 4.8.1:** Seja  $\vec{X} = (X_1, X_2, \dots, X_k)$  um vetor aleatório  $k$ -dimensional. Então, os *momentos conjuntos* de  $\vec{X}$  são da forma  $E(\prod_{i=1}^k X_i^{j_i})$ , onde  $j_i$ 's são inteiros positivos, se esta esperança existir. De forma análoga ao caso unidimensional pode-se definir também *momentos conjuntos centrais*. ■

No caso bidimensional, temos que a correlação e a covariância são momentos conjuntos que são medidas do grau de dependência linear entre duas variáveis.

**Definição 4.8.2:** A correlação entre duas variáveis aleatórias  $X$  e  $Y$  é dada por  $EXY$  se esta esperança existe. A covariância entre elas é dada por  $Cov(X, Y) = E[(X - EX)(Y - EY)] = EXY - (EX)(EY)$ .

Note que  $Cov(X, X) = VarX$ . Pela prova da propriedade 5 de variância, vemos que a seguinte relação é válida:

$$Var(X + Y) = VarX + VarY + 2Cov(X, Y).$$

Diz-se que duas variáveis são *não-correlacionadas* se  $Cov(X, Y) = 0$ . Como já provamos que se  $X$  e  $Y$  são independentes, então  $EXY = EXEY$ . Temos que se  $X$  e  $Y$  são independentes, elas necessariamente são não-correlacionadas. O contrário nem sempre é verdadeiro como o próximo exemplo ilustra.

**Exemplo 4.8.3:** Se  $X$  é uma variável aleatória tal que  $P(X = -a) = P(X = a) = 1/2$  e  $Y = X^2$ , temos que  $EXY = -a^3(1/2) + a^3(1/2) = 0$  e  $EX = -a(1/2) + a(1/2) = 0$ . Logo,  $EXY = EXEY = 0$ , ou seja,  $Cov(X, Y) = 0$ . Porém,  $X$  e  $Y$  não são independentes, pois  $Y$  é uma função de  $X$ .

Vejam agora uma expressão para a variância da soma de  $n$  variáveis aleatórias.

**Teorema 4.8.4:** Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias tais que  $Var(X_i) < \infty$ , então

$$Var(X_1 + \dots + X_n) = \sum_{i=1}^n VarX_i + 2 \sum_{i < j} Cov(X_i, X_j).$$

**Prova:**

$$\begin{aligned} Var(X_1 + \dots + X_n) &= E(X_1 + \dots + X_n - E(X_1 + \dots + X_n))^2 \\ &= E\left(\sum_{i=1}^n (X_i - EX_i)\right)^2 = E\left[\sum_{i=1}^n (X_i - EX_i)^2 + 2 \sum_{i < j} (X_i - EX_i)(X_j - EX_j)\right] \\ &= \sum_{i=1}^n Var(X_i) + 2 \sum_{i < j} Cov(X_i, X_j). \end{aligned}$$

■

**Corolário 4.8.5:** *Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias tais que  $Var(X_i) < \infty$  e  $Cov(X_i, X_j) = 0$  para  $i \neq j$ , então*

$$Var(X_1 + \dots + X_n) = \sum_{i=1}^n Var X_i.$$

O próximo teorema trata de uma importante desigualdade em teoria da probabilidade:

**Teorema 4.8.6:**  $(E(XY))^2 \leq EX^2EY^2$  e  $(Cov(X, Y))^2 \leq VarXVarY$ .

**Prova:**  $(aX + Y)^2 \geq 0 \Rightarrow E(aX + Y)^2 \geq 0 \Rightarrow a^2EX^2 + 2aEXY + EY^2 \geq 0$ . Observe que esta equação do segundo grau em  $a$  não pode ter duas raízes reais diferentes, pois caso contrário essa expressão seria negativa para os valores entre as raízes. Então, utilizando a regra do discriminante, temos que

$$4(EXY)^2 - 4EX^2EY^2 \leq 0,$$

e temos a primeira desigualdade. A segunda desigualdade segue da primeira trocando  $X$  por  $X - EX$  e  $Y$  por  $Y - EY$  na expressão da primeira desigualdade. ■

O coeficiente de correlação entre duas variáveis aleatórias  $X$  e  $Y$  é dado por

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}.$$

O teorema anterior provou que  $|\rho(X, Y)| \leq 1$ . O próximo teorema mostra que o módulo do coeficiente de correlação entre duas variáveis é igual a 1 se, e somente se, as variáveis são linearmente dependentes.

**Teorema 4.8.7:** *Sejam  $X$  e  $Y$  variáveis aleatórias com variâncias finitas e positivas. Então,*

(a)  $\rho(X, Y) = 1$  se, e somente se,  $P(Y = aX + b) = 1$  para algum  $a > 0$  e  $b \in \mathbb{R}$ .

(b)  $\rho(X, Y) = -1$  se, e somente se,  $P(Y = aX + b) = 1$  para algum  $a < 0$  e  $b \in \mathbb{R}$ .

**Prova:** Parte (a). Como  $(\frac{X-EX}{\sqrt{Var(X)}} - \frac{Y-EY}{\sqrt{Var(Y)}})^2 \geq 0$ , temos que

$$\begin{aligned} 0 &\leq E\left(\frac{X-EX}{\sqrt{Var(X)}} - \frac{Y-EY}{\sqrt{Var(Y)}}\right)^2 \\ &= E\left(\frac{X-EX}{\sqrt{Var(X)}}\right)^2 + E\left(\frac{Y-EY}{\sqrt{Var(Y)}}\right)^2 - \frac{2}{\sqrt{Var(X)Var(Y)}}E((X-EX)(Y-EY)) \\ &= \frac{VarX}{VarX} + \frac{VarY}{VarY} - \frac{2Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} = 2 - 2\rho(X, Y). \end{aligned}$$

Se  $\rho(X, Y) = 1$ , então

$$E\left(\frac{X-EX}{\sqrt{Var(X)}} - \frac{Y-EY}{\sqrt{Var(Y)}}\right)^2 = 0,$$



o que por sua vez implica que

$$P\left(\frac{X - EX}{\sqrt{\text{Var}(X)}} = \frac{Y - EY}{\sqrt{\text{Var}(Y)}}\right) = 1,$$

em outras palavras,

$$P\left(Y = EY + \frac{\sqrt{\text{Var}Y}}{\sqrt{\text{Var}X}}(X - EX)\right) = 1.$$

A prova da parte (b) é análoga, substituindo o sinal “+” por “-” na expressão acima. Deixamos para o leitor verificar os detalhes. ■

O próximo teorema apresenta uma nova relação entre momentos conjuntos de variáveis aleatórias. Ele é conhecido como *Desigualdade de Hölder*.

**Teorema 4.8.8:** *Suponha que  $p$  e  $q$  satisfazem:  $p > 1$ ,  $q > 1$ , e  $\frac{1}{p} + \frac{1}{q} = 1$ . Então, se  $E(|X|^p) < \infty$  e  $E(|Y|^q) < \infty$ , temos que*

$$E(|XY|) \leq (E|X|^p)^{1/p} (E|Y|^q)^{1/q}.$$

**Prova:** A prova da desigualdade de Hölder utiliza um argumento de convexidade. Como  $|X|^p \geq 0$  (resp.,  $|X|^q \geq 0$ ), já vimos que se  $E|X|^p = 0$ , então  $P(X = 0) = 1$ . Portanto, em ambos os casos  $E(|XY|) = 0$  e a desigualdade de Hölder é válida. Considere então o caso em que o lado direito da desigualdade de Hölder é estritamente positivo.

Note que para  $a > 0$  e  $b > 0$ , existe  $s, t \in \mathbb{R}$  tal que

$$a = \exp\left(\frac{s}{p}\right) \text{ e } b = \exp\left(\frac{t}{q}\right).$$

Como a função exponencial é convexa e  $p^{-1} + q^{-1} = 1$ , temos por convexidade que

$$\exp\left(\frac{s}{p} + \frac{t}{q}\right) \leq p^{-1} \exp(s) + q^{-1} \exp(t),$$

ou pela definição de  $s, t$

$$ab \leq p^{-1} a^p + q^{-1} b^q.$$

Agora substituindo  $a$  por  $\frac{|X|}{(E(|X|^p))^{1/p}}$  e  $b$  por  $\frac{|Y|}{(E(|Y|^q))^{1/q}}$ , temos

$$\frac{|XY|}{(E(|X|^p))^{1/p} (E(|Y|^q))^{1/q}} \leq p^{-1} \left(\frac{|X|}{(E(|X|^p))^{1/p}}\right)^p + q^{-1} \left(\frac{|Y|}{(E(|Y|^q))^{1/q}}\right)^q.$$

Finalmente, tomando o valor esperado, temos

$$\begin{aligned} & \frac{E|XY|}{(E(|X|^p))^{1/p} (E(|Y|^q))^{1/q}} \\ & \leq p^{-1} \left(\frac{E|X|^p}{(E(|X|^p))}\right)^p + q^{-1} \left(\frac{E|Y|^q}{(E(|Y|^q))}\right)^q \\ & = p^{-1} + q^{-1} = 1. \end{aligned}$$

■

■

# Capítulo 5

## Distribuição e Esperança Condicionais

### 5.1 Distribuição condicional de $X$ dada $Y$ discreta

Seja  $X$  uma variável aleatória no espaço de probabilidade  $(\Omega, \mathcal{A}, P)$ , e seja  $A$  um evento aleatório tal que  $P(A) > 0$ . Usando o conceito de probabilidade condicional, podemos definir a distribuição condicional de  $X$  dado o evento  $A$  por

$$P(X \in B|A) = \frac{P([X \in B] \cap A)}{P(A)},$$

para  $B$  boreliano. Pode-se verificar facilmente que isto define uma probabilidade nos borelianos verificando-se os axiomas. Podemos interpretar a distribuição condicional de  $X$  dado  $A$  como a nova distribuição que se atribui a  $X$  quando sabe-se da ocorrência do evento  $A$ . A função de distribuição associada à distribuição condicional é chamada função distribuição condicional de  $X$  dado  $A$ :

$$F_X(x|A) = P(X \leq x|A).$$

A esperança condicional de  $X$  dado  $A$  é a esperança da distribuição condicional, definida por

$$E(X|A) = \int x dF_X(x|A),$$

se esta esperança existe.

Agora suponhamos que os eventos aleatórios  $A_1, A_2, \dots$  formem uma partição (finita ou enumerável) de  $\Omega$ . Pelo Teorema da Probabilidade Total, temos

$$P(X \in B) = \sum_n P(A_n)P(X \in B|A_n), \forall B \in \mathcal{B},$$

e

$$\begin{aligned} F_X(x) &= P(X \leq x) = \sum_n P(A_n)P(X \leq x|A_n) \\ &= \sum_n P(A_n)F_X(x|A_n), \forall x, \end{aligned}$$

e se a esperança de  $X$  existe,

$$\begin{aligned} EX &= \int x dF_X(x) = \int x d\left(\sum_n P(A_n)F_X(x|A_n)\right) \\ &= \sum_n P(A_n) \int x dF_X(x|A_n) = \sum_n P(A_n)E(X|A_n). \end{aligned}$$

Em outras palavras, a distribuição de  $X$  (resp., função de distribuição, esperança de  $X$ ) é uma média ponderada da distribuição condicional (resp., função de distribuição condicional, esperança condicional de  $X$ ) dado  $A_n$ , onde os pesos são as probabilidades dos membros  $A_n$  da partição.

Consideremos agora o caso em que a partição do espaço amostral é gerada por uma variável aleatória discreta. Para tanto, seja  $Y$  uma variável aleatória discreta em  $(\Omega, \mathcal{A}, P)$ , tomando somente os valores  $y_1, y_2, \dots$ . Então, os eventos  $A_n = [Y = y_n]$  formam uma partição de  $\Omega$ . Neste caso, a distribuição

$$P(X \in B|Y = y_n) = P(X \in B|A_n),$$

para  $B$  boreliano, é chamada de distribuição condicional de  $X$  dado que  $Y = y_n$ , e valem as fórmulas

$$\begin{aligned} P(X \in B) &= \sum_n P(Y = y_n)P(X \in B|Y = y_n), \quad B \text{ boreliano} \\ F_X(x) &= \sum_n P(Y = y_n)F_X(x|Y = y_n) \\ EX &= \sum_n P(Y = y_n)E(X|Y = y_n), \end{aligned}$$

onde vale a última fórmula se  $EX$  existe; em particular, se  $X$  é integrável.

Notemos que para  $B$  fixo,  $P(X \in B|Y = y_n)$  é função de  $y_n$ , digamos  $g(y_n)$ . Se definirmos  $g(y) = P(X \in B|Y = y)$  arbitrariamente para  $y \notin \{y_n : n \geq 1\}$ , por exemplo,  $g(y) = P(X \in B)$ , então teremos

$$P(X \in B) = \int P(X \in B|Y = y)dF_Y(y) = \int g(y)dF_Y(y),$$

pelas propriedades da integral de Lebesgue no caso de  $Y$  discreto. As outras fórmulas possuem interpretações análogas, logo teremos

$$\begin{aligned} P(X \in B) &= \int P(X \in B|Y = y)dF_Y(y), \quad B \text{ boreliano} \\ F_X(x) &= \int F_X(x|Y = y)dF_Y(y) \\ EX &= \int E(X|Y = y)dF_Y(y). \end{aligned}$$

Essas fórmulas valem também no caso geral, como veremos adiante. Salientamos que a esperança precisa existir para que a última fórmula valha. De fato, quando  $X$  for integrável,  $\varphi(y) = E(X|Y = y)$  será finito. Nesse caso, a variável aleatória  $\varphi(Y)$  será chamada de esperança condicional de  $X$  dada  $Y$  e será indicada por  $\varphi(Y) = E(X|Y)$ . Notemos que  $E(X|Y = y)$  é um valor particular da variável aleatória  $E(X|Y)$ : é o valor quando  $Y = y$ . Portanto, a última fórmula pode ser reescrita assim

$$EX = E\varphi(Y) = E(E(X|Y)).$$

Em outras palavras, a esperança de  $X$  é igual à esperança da esperança condicional de  $X$  dada  $Y$ .

**Exemplo 5.1.1:** Consideremos o seguinte experimento em que participam dois jogadores, I e II. Suponhamos que o jogador I lance uma moeda honesta  $n$  vezes, obtendo  $k$  caras, onde  $0 \leq k \leq n$ , e que depois disso o jogador II lance a mesma moeda  $k$  vezes. Seja  $X$  o número de caras obtidas pelo jogador II. Qual a esperança de  $X$  supondo independência de todos os lançamentos?

Seja  $Y$  o número de caras nos  $n$  lançamentos do jogador I. Decorre das condições do experimento que  $Y \sim b(n, \frac{1}{2})$  e que  $X|Y = k \sim b(k, \frac{1}{2})$ . Por isso, a esperança condicional de  $X$  dado que  $Y = k$  é a esperança da distribuição  $b(k, \frac{1}{2})$ :  $E(X|Y = k) = \frac{k}{2}$ , ou seja,  $E(X|Y) = \frac{Y}{2}$ . Utilizando a fórmula, temos

$$EX = E(E(X|Y)) = E\left(\frac{Y}{2}\right) = \frac{n}{4}.$$

**Exemplo 5.1.2:** Consideremos outro jogo que conta com a participação de dois jogadores I e II. Neste jogo, o jogador I vai fazer uma seqüência de lançamentos independentes de uma moeda que tem probabilidade  $p$  de dar cara, onde  $0 < p < 1$ . Antes do jogador I começar, o jogador II observa uma variável aleatória  $N$  tendo distribuição *Poisson*( $\lambda$ ), onde  $\lambda > 0$ . Supomos que  $N$  seja independente da seqüência de lançamentos do jogador I. Se o jogador II observar  $N = n$ , ele vai parar o jogador I depois de ter feito  $n$  lançamentos (se  $N = 0$ , o jogador II não permite nenhum lançamento). Se  $S$  for o número de caras observadas até o jogador I parar, qual é a esperança de  $S$ ?

**Solução:** Como a seqüência de lançamentos é independente de  $N$ , a distribuição condicional de  $S$  dado que  $N = n$  é *binomial*( $n, p$ ). Portanto,  $E(S|N = n) = np$ , ou seja,  $E(S|N) = Np$ . Logo,

$$ES = E(Np) = pEN = p\lambda.$$

## 5.2 Distribuição condicional de $X$ dada $Y$ : caso geral

Nosso objetivo nesta seção é definir a distribuição condicional de  $X$  dado que  $Y = y$  para todo  $y \in R$  e todo par de variáveis aleatórias  $X$  e  $Y$  definidas no mesmo espaço de probabilidade  $(\Omega, \mathcal{A}, P)$ . Na seção anterior definimos a distribuição condicional dado que  $Y = y$  quando  $P(Y = y) > 0$ ; portanto nosso problema agora é como definir distribuição condicional quando  $P(Y = y) = 0$ . No caso discreto essa definição era arbitrária, pois o conjunto

$B_0 = \{y_n : n = 1, 2, \dots\}^c$  também tinha probabilidade zero. Mas é evidente que essa solução não serve no caso geral, já que no caso contínuo  $P(Y = y) = 0$  para todo  $y \in R$ .

Para termos uma intuição sobre a definição formal da distribuição condicional no caso geral, consideremos novamente o caso discreto. Pelas fórmulas obtidas na seção anterior a distribuição (resp., função de distribuição, esperança) de  $X$  é determinada pela distribuição  $Y$  e a distribuição (resp., função de distribuição, esperança) condicional de  $X$  dada  $Y$ . De fato, o Teorema da Probabilidade Total nos dá um resultado muito mais forte: a distribuição conjunta de  $X$  e  $Y$  é determinada pela distribuição de  $Y$  e a distribuição condicional de  $X$  dada  $Y$ . Para ver isto, basta notar que para todo  $x$  e  $y$ ,

$$\begin{aligned} F_{X,Y}(x, y) &= P(X \leq x, Y \leq y) = \sum_{n: y_n \leq y} P(X \leq x, Y = y_n) \\ &= \sum_{n: y_n \leq y} P(Y = y_n)P(X \leq x|Y = y_n) = \sum_{n: y_n \leq y} P(Y = y_n)F_X(x|Y = y_n) \\ &= \int_{-\infty}^y F_X(x|Y = t)dF_Y(t). \end{aligned}$$

Vemos então que no caso discreto a função de distribuição conjunta é uma espécie de composta da função de distribuição marginal de  $Y$  com a função de distribuição condicional de  $X$  dada  $Y$ . E pode-se provar que para todo par de variáveis aleatórias  $X$  e  $Y$ , definidas no mesmo espaço de probabilidade, existe uma, e somente uma, família de funções de distribuição condicional satisfazendo a condição acima. Isto justifica a seguinte definição formal para a distribuição condicional de  $X$  dada  $Y$ :

**Definição 5.2.1:** Sejam  $X$  e  $Y$  variáveis aleatórias definidas no mesmo espaço de probabilidade  $(\Omega, \mathcal{A}, P)$ . Uma função  $P(X \in B|Y = y)$ , definida para  $B$  boreliano e  $y \in R$ , será chamada uma distribuição condicional (regular) para  $X$  dada  $Y$  se

- (i) para todo  $y \in R$  fixo,  $P(X \in B|Y = y)$  define uma probabilidade na  $\sigma$ -álgebra de Borel; e
- (ii) para todo  $B$  boreliano fixo,  $P(X \in B|Y = y)$  é função mensurável de  $y$  e para todo  $(x, y) \in R^2$ ,

$$\int_{-\infty}^y F_X(x|Y = t)dF_Y(t) = F_{X,Y}(x, y).$$

O próximo teorema prova que esta definição determina uma única distribuição condicional quase certamente.

**Teorema 5.2.2:** Sejam  $X$  e  $Y$  variáveis aleatórias em  $(\Omega, \mathcal{A}, P)$ . Então existe uma distribuição condicional regular para  $X$  dada  $Y$ . Existe apenas uma, no sentido de que duas distribuições condicionais são iguais quase certamente: se  $P_1(X \in B|Y = y)$  e  $P_2(X \in B|Y = y)$  são ambas distribuições condicionais para  $X$  dada  $Y$ , então existe um boreliano  $B_0$  tal que  $P(Y \in B_0) = 1$  e  $P_1(X \in B|Y = y) = P_2(X \in B|Y = y)$ , para todo  $B$  boreliano e  $y \in B_0$ .

**Prova:** Omitida. ■

Existe uma outra alternativa para se calcular a distribuição condicional de  $X$  dada  $Y$  que utiliza uma aproximação da definição do caso discreto. Para tanto, seja  $I$  um intervalo pequeno de comprimento  $\Delta y$  e que contém o ponto  $y$ . Tomemos como aproximação para a probabilidade condicional de  $X$  pertencer a  $B$  dado que  $Y = y$ , a probabilidade condicional do mesmo evento dado que  $Y \in I$ , ou seja,

$$P(X \in B|Y = y) \approx P(X \in B|Y \in I) = \frac{P(X \in B, Y \in I)}{P(Y \in I)}.$$

O seguinte teorema prova que esta maneira alternativa de calcular a distribuição condicional de  $X$  dado  $Y$  quase sempre coincide com a Definição 5.2.1.

**Teorema 5.2.3:** *Para cada  $B$  boreliano fixo, o limite na definição 4.2 existe quase certamente, i.e.,  $P(Y \in \{y : \text{limite existe em } y\}) = 1$ . Além disso, para cada  $B$  fixo, o limite é igual a  $P(X \in B|Y = y)$  como definido na Definição 5.2.1, quase certamente, ou seja, o conjunto dos  $y$ 's para os quais o limite converge para  $P(X \in B|Y = y)$  conforme a Definição 5.2.1 tem probabilidade 1.*

Tanto a Definição 5.2.1 quanto o método da aproximação por limites não são úteis para encontrar a distribuição condicional. Para tanto deve-se tentar adivinhar um candidato. Consideremos alguns casos simples em que a solução vem de imediato:

**Caso I:  $Y$  discreta.** Considere a solução que obtivemos quando analisamos o caso discreto. Portanto, se  $Y$  assume os valores  $y_1, y_2, \dots$  tais que  $P(Y = y_n) > 0$ , então

$$P(X \in B|Y = y_n) = \frac{P(X \in B, Y = y_n)}{P(Y = y_n)}, \forall B \in \mathcal{B},$$

e  $P(X \in B|Y = y) = P(X \in B)$  se  $P(Y = y) = 0$ . Note que esta distribuição satisfaz as duas condições da Definição 5.2.1 e portanto é uma distribuição condicional de acordo com a definição do caso geral.

**Caso II:  $X$  e  $Y$  independentes.** Intuitivamente, a distribuição condicional de  $X$  dado que  $Y = y$  não deveria depender de  $y$ . Portanto, nosso candidato é:

$$P(X \in B|Y = y) = P(X \in B), \forall B \in \mathcal{B}, \forall y \in \mathcal{R}.$$

Portanto, a primeira condição da Definição 5.2.1 é satisfeita e nosso candidato para  $F_X(x|Y = y)$  é  $F_X(x)$ , logo

$$\int_{-\infty}^y F_X(x) dF_Y(t) = F_X(x) \int_{-\infty}^y dF_Y(t) = F_X(x) F_Y(y) = F_{X,Y}(x, y),$$

ou seja, a segunda condição da definição também é satisfeita.

**Caso III:  $X$  e  $Y$  possuem densidade conjunta  $f(x, y)$ .** Neste caso nosso candidato será

$$f(x|y) = \frac{f(x, y)}{f(y)}, x \in \mathcal{R},$$

se  $f(y) > 0$ , e  $f(x|y) = f(x)$  se  $f(y) = 0$ . Esta função é chamada de densidade condicional de  $X$  dado que  $Y = y$ . Note que  $f(x|y)$  preserva as chances relativas e realmente é uma densidade. Agora, vamos mostrar que ela satisfaz a Definição 5.2.1. Parte (i), segue do fato que  $f(x|y)$  é uma densidade de probabilidade e portanto  $P(X \in B|Y = y) = \int_{X \in B} f(x|y)dx$  é uma probabilidade para todo boreliano  $B$ . Para verificar (ii), note que a função de distribuição condicional é  $F_X(x|Y = t) = \int_{-\infty}^x f(s|t)ds$ . Logo

$$\begin{aligned} \int_{-\infty}^y \left( \int_{-\infty}^x f(s|t)ds \right) dF_Y(t) &= \int_{-\infty}^y \left( \int_{-\infty}^x \frac{f(s,t)}{f_Y(t)} ds \right) f_Y(t) dt \\ &= \int_{-\infty}^y \int_{-\infty}^x f(s,t) ds dt = F_{X,Y}(x,y). \end{aligned} \quad (5.1)$$

**Caso IV: X discreta e Y com densidade  $f_Y$ .** De acordo com a definição de distribuição condicional, ela deve satisfazer neste caso:

$$\int_{-\infty}^y P(X = x_i|Y = t) f_Y(t) dt = P(X = x_i, Y \leq y).$$

Note que se definirmos

$$\begin{aligned} P(X = x_i|Y = t) &= \frac{1}{f_Y(t)} \frac{\partial P(X = x_i, Y \leq t)}{\partial t} \\ &= \frac{1}{f_Y(t)} \frac{\partial P(Y \leq t|X = x_i) P(X = x_i)}{\partial t} \\ &= \frac{P(X = x_i)}{f_Y(t)} f_{Y|X}(t|x_i), \end{aligned}$$

obtemos o resultado desejado.

Em casos mais complexos, no processo de escolha da distribuição condicional, ajuda observar os seguintes princípios:

- **Princípio da preservação das chances relativas.** Este princípio diz que condicionalmente, dada a ocorrência de um evento  $A$ , os resultados possíveis (ou seja,  $w \in A$ ) mantêm as mesmas chances relativas que tinham antes da realização do experimento.
- **Princípio da substituição.** Este princípio diz que condicionalmente, dado que  $Y = y$ , a variável aleatória  $Y$  pode ser substituída pelo valor  $y$  sempre que  $Y$  aparecer em uma probabilidade (ou esperança) condicional. Mais geralmente, diz que para obter a distribuição condicional de  $\varphi(X, Y)$  dado que  $Y = y$ , basta substituir  $Y$  pelo valor  $y$ .

**Exemplo 5.2.4:** Seja  $X$  uma variável aleatória simétrica em torno de zero, de modo que  $P(X \leq x) = P(X \geq -x), \forall x \in \mathbb{R}$ . Qual a distribuição condicional de  $X$  dado  $|X|$ ?

Utilizando o princípio da preservação das chances relativas e a simetria da variável  $X$ , temos que nosso candidato para distribuição condicional deve ser:  $P(X = y||X| = y) = P(X = -y||X| = y) = 1/2$  se  $y > 0$  e  $P(X = 0||X| = 0) = 1$ .

Como

$$\begin{aligned}
& \int_0^y P(X \leq x | |X| = t) dF_{|X|}(t) \\
&= \begin{cases} \int_0^y 0 dF_{|X|}(t) & , \text{ se } x < -y \\ \int_0^{|x|^-} 0 dF_{|X|}(t) + \int_{|x|^-}^y \frac{1}{2} dF_{|X|}(t) & , \text{ se } -y \leq x < 0 \\ \int_0^x 1 dF_{|X|}(t) + \int_x^y \frac{1}{2} dF_{|X|}(t) & , \text{ se } 0 \leq x < y \\ \int_0^y 1 dF_{|X|}(t) & , \text{ se } x \geq y \geq 0 \end{cases} \\
&= \begin{cases} 0 & , \text{ se } x < -y \\ 1/2(F_{|X|}(y) - F_{|X|}(|x|^-)) & , \text{ se } -y \leq x < 0 \\ 1/2(F_{|X|}(y) + F_{|X|}(x)) & , \text{ se } 0 \leq x < y \\ F_{|X|}(y) & , \text{ se } x \geq y \geq 0 \end{cases} \\
&= \begin{cases} 0 & , \text{ se } x < -y \\ F_X(x) - F_X(-y^-) & , \text{ se } -y \leq x < 0 \\ F_X(x) - F_X(-y^-) & , \text{ se } 0 \leq x < y \\ F_{|X|}(y) & , \text{ se } x \geq y \geq 0 \end{cases}
\end{aligned}$$

Mas esta última expressão é igual a  $F_{X,|X|}(x, y)$ . Portanto, nosso candidato satisfaz a definição de distribuição condicional.

**Exemplo 5.2.5:** Se  $f_{Y|X}(y|x) = |x + 1|e^{-|x+1|y}U(y)$  e  $X \sim \text{Binomial}(2, 1/2)$ , qual a densidade de  $Y$ ? Dado que  $Y = y$ , qual a distribuição de  $X$  para  $y > 0$ ?

**Solução:**

$$\begin{aligned}
f_Y(y) &= \sum_{i=0}^2 |i + 1|e^{-|i+1|y}U(y) \binom{2}{i} (1/2)^2 \\
&= \frac{1}{4}U(y)(e^{-y} + 4e^{-2y} + 3e^{-3y})
\end{aligned}$$

Utilizando o resultado do Caso IV acima temos que

$$\begin{aligned}
P(X = i | Y = y) &= \frac{P(X = i)}{f_Y(y)} f_{Y|X}(y|i) \\
&= \frac{\binom{2}{i}|i + 1|e^{-|i+1|y}}{(e^{-y} + 4e^{-2y} + 3e^{-3y})}, i = 0, 1, 2.
\end{aligned}$$

## 5.3 Esperança Condicional

**Definição 5.3.1:** Sejam  $X$  e  $Y$  variáveis aleatórias em  $(\Omega, \mathcal{A}, P)$ . A esperança condicional de  $X$  dado que  $Y = y$ , é a esperança da distribuição condicional de  $X$  dado que  $Y = y$ , se esta esperança existir. Ou seja,

$$E(X|Y = y) = \int x dF_X(x|Y = y).$$



Pode-se provar que:

**Teorema 5.3.2:** *Se  $X$  é integrável, então  $E(X|Y = y)$  existe e é finita quase certamente, i.e., existe um boreliano  $B_0$  tal que  $P(Y \in B_0) = 1$  e  $E(X|Y = y)$  é finita para todo  $y \in B_0$ .*

Se definirmos  $\varphi(y) = E(X|Y = y)$ , a variável aleatória  $\varphi(Y) = E(X|Y)$  chama-se esperança condicional de  $X$  dada  $Y$ . A esperança condicional, sendo a esperança da distribuição condicional, possui todas as propriedades da esperança ordinária (por exemplo, linearidade, desigualdade de Jensen, convergência monótona, convergência dominada), mais a propriedade importante de que  $E(E(X|Y)) = EX$ , ou seja

$$EX = \int E(X|Y = y)dF_Y(y).$$

Já demonstramos esta equação no caso discreto, vamos verificá-las quando  $X$  e  $Y$  têm densidade conjunta  $f(x, y)$ :

$$E(X|Y = y) = \int x dF_X(x|Y = y) = \int_{-\infty}^{\infty} x f(x|y) dx = \int_{-\infty}^{\infty} x \frac{f(x, y)}{f_Y(y)} dx,$$

se  $f_Y(y) > 0$ . Logo, quando  $X$  é integrável,

$$\begin{aligned} E(E(X|Y)) &= \int E(X|Y = y)dF_Y(y) = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} x \frac{f(x, y)}{f_Y(y)} dx \right) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx dy = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} f(x, y) dy \right) x dx \\ &= \int_{-\infty}^{\infty} x f_X(x) dx = EX. \end{aligned}$$

Como  $A = [I_A = 1]$ , temos

$$\begin{aligned} E(I_A|Y = y) &= 1 \cdot P(I_A = 1|Y = y) + 0 \cdot P(I_A = 0|Y = y) \\ &= P(I_A = 1|Y = y) = P(A|Y = y). \end{aligned}$$

De fato, como  $I_A$  é integrável, nós temos

$$P(A) = E(I_A) = E(E(I_A|Y)) = E(P(A|Y)),$$

ou seja, a probabilidade de um evento é a esperança de sua probabilidade condicional dada  $Y$ , para qualquer  $Y$ .

A seguir enumeramos algumas propriedades da esperança condicional, que são generalizações de propriedades da esperança incondicional.

EC1.  $E(E(X|Y)) = EX$ .

EC2. Se  $X = c$ , para alguma constante  $c$ , então  $E(X|Y) = c$ .

EC3. Se  $X_1 \leq X_2$ , então  $E(X_1|Y) \leq E(X_2|Y)$ .

EC4.  $E(aX_1 + bX_2|Y) = aE(X_1|Y) + bE(X_2|Y)$ .

EC5. Seja  $\varphi$  uma função convexa. Então,  $\varphi(E(X|Y)) \leq E(\varphi(X)|Y)$ .

EC6. Se  $X_n \geq 0$  e  $X_n \uparrow X$ , então  $E(X_n|Y) \uparrow E(X|Y)$ .

EC7. Se  $X_n \rightarrow X$  e se existe  $X_0$  integrável tal que  $|X_n| \leq X_0$ , então  $\lim_n E(X_n|Y) = E(X|Y)$ .

EC8. Se  $\varphi(X, Y)$  é integrável, então

$$E(\varphi(X, Y)|Y = y) = E(\varphi(X, y)|Y = y) = \int \varphi(x, y) dF_X(x|Y = y).$$

Assim como no caso incondicional podemos definir momentos condicionais de ordem mais elevada de maneira análoga. O  $k$ -ésimo momento de  $X$  dado  $Y$  é dado por  $E(X^k|Y)$ . E o  $k$ -ésimo momento central é dado por  $E((X - E(X|Y))^k|Y)$ . Em particular, o segundo momento central é conhecido como variância condicional de  $X$  dado  $Y$  e pode ser reescrito como:

$$\text{Var}(X|Y) = E((X - E(X|Y))^2|Y) = E(X^2|Y) - (E(X|Y))^2.$$

**Exemplo 5.3.3:** Sejam  $X$  e  $Y$  variáveis aleatórias independentes e identicamente distribuídas, com  $X \sim U[0, 1]$ , e sejam  $U = \min(X, Y)$  e  $V = \max(X, Y)$ . Encontre  $E(U|V)$ .

**Solução:**

$$\begin{aligned} F_{U,V}(x, y) &= P(U \leq x, V \leq y) = P(V \leq y) - P(U > x, V \leq y) \\ &= \begin{cases} P(X \leq y, Y \leq y) - P(x < X \leq y, x < Y \leq y) & , \text{ se } x < y \\ P(X \leq y, Y \leq y) & , \text{ se } x \geq y. \end{cases} \end{aligned}$$

Portanto, como  $X$  e  $Y$  são independentes, temos

$$F_{U,V}(x, y) = \begin{cases} 0 & , \text{ se } x \leq 0 \text{ ou } y \leq 0 \\ y^2 - (y - x)^2 & , \text{ se } 0 < x < y < 1 \\ y^2 & , \text{ se } 0 < y \leq x \text{ e } y < 1 \\ 1 - (1 - x)^2 & , \text{ se } y \geq 1 \text{ e } 0 < x < 1 \\ 1 & , \text{ se } y \geq 1 \text{ e } x \geq 1. \end{cases}$$

Logo,

$$f_{U,V}(x, y) = \frac{\partial^2 F_{U,V}(x, y)}{\partial x \partial y} = \begin{cases} 2 & , \text{ se } 0 < x < y < 1 \\ 0 & , \text{ caso contrário.} \end{cases}$$

Como  $f_V(y) = \int f_{U,V}(x, y) dx = \int_0^y 2 dx = 2y$ , se  $0 < y < 1$ , e  $f_V(y) = 0$  caso contrário, temos que

$$f_{U|V}(x|y) = \frac{f_{U,V}(x, y)}{f_V(y)} = \begin{cases} \frac{1}{y} & , \text{ se } 0 < x < y < 1 \\ 0 & , \text{ caso contrário.} \end{cases}$$

Então,

$$E(U|V = y) = \int x f_{U|V}(x|y) dx = \int_0^y \frac{x}{y} dx = \frac{y}{2},$$

se  $0 < y < 1$ , e  $E(U|V = y) = 0$ , caso contrário. Portanto,

$$E(U|V) = \begin{cases} \frac{V}{2} & , \text{ se } 0 < V < 1 \\ 0 & , \text{ caso contrário.} \end{cases}$$

**Exemplo 5.3.4:** Sejam  $X_1, \dots, X_n$  independentes, identicamente distribuídas e integráveis, e seja  $S = X_1 + \dots + X_n$ . Demonstre que  $E(X_i|S) = \frac{S}{n}$ , para  $i = 1, 2, \dots, n$ .

**Solução:** Note que os vetores  $(X_1, \dots, X_n)$  e  $(X_i, X_2, \dots, X_{i-1}, X_1, X_{i+1}, \dots, X_n)$  têm a mesma distribuição. Isto implica que  $(X_1, S)$  e  $(X_i, S)$  possuem a mesma distribuição. Como a distribuição conjunta determina a distribuição condicional, temos que  $X_1$  e  $X_i$  têm a mesma distribuição condicional dado que  $S = s$ , e conseqüentemente tem a mesma esperança condicional dado  $S = s$ . Portanto,

$$E(X_1|S = s) = E(X_2|S = s) = \dots = E(X_n|S = s).$$

Utilizando a linearidade da esperança, temos

$$\begin{aligned} nE(X_i|S = s) &= \sum_{i=1}^n E(X_i|S = s) \\ &= E\left(\sum_{i=1}^n X_i|S = s\right) = E(S|S = s) = s. \end{aligned}$$

Então, podemos concluir que  $E(X_i|S = s) = \frac{s}{n}$ , ou seja,  $E(X_i|S) = \frac{S}{n}$ .

**Exemplo 5.3.5:** Sejam  $X$  e  $Y$  duas variáveis aleatórias. Calculemos a distribuição de  $Z = X + Y$ . Temos

$$\begin{aligned} P(X + Y \leq z) &= E(P(X + Y \leq z|Y)) = \int P(X + Y \leq z|Y = y) dF_Y(y) \\ &= \int P(X \leq z - y|Y = y) dF_Y(y) = \int F_X(z - y|Y = y) dF_Y(y). \end{aligned} \quad (5.2)$$

Se  $X$  e  $Y$  são independentes, então  $F_X(z - y|Y = y) = F_X(z - y)$  e temos

$$F_Z(z) = P(X + Y \leq z) = \int F_X(z - y) dF_Y(y). \quad (5.3)$$

Esta distribuição é a convolução das distribuições de  $X$  e  $Y$ .

# Capítulo 6

## Convergência Estocástica

### 6.1 Seqüência de Eventos

A definição de conceitos de convergência de variáveis aleatórias depende de manipulações de seqüências de eventos. Seja  $A_n \subseteq \Omega$ , define-se:

$$\begin{aligned}\inf_{k \geq n} A_k &= \bigcap_{k=n}^{\infty} A_k, & \sup_{k \geq n} A_k &= \bigcup_{k=n}^{\infty} A_k \\ \liminf_n A_n &= \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \\ \limsup_n A_n &= \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.\end{aligned}$$

O limite de uma seqüência de eventos é definido da seguinte maneira: se para alguma seqüência  $(B_n)$  de eventos  $\liminf_n B_n = \limsup_n B_n = B$ , então  $B$  é chamado de limite de  $(B_n)$  e nós escrevemos  $\lim_n B_n = B$  ou  $B_n \rightarrow B$ .

**Exemplo 6.1.1:**  $\liminf[0, \frac{n}{n+1}) = \limsup[0, \frac{n}{n+1}) = [0, 1)$

**Teorema 6.1.2:** *Seja  $(A_n)$  uma seqüência de eventos de  $\Omega$ .*

(a)  $\omega \in \limsup A_n$  se, e somente se,  $\omega \in A_k$  para um número infinito de índices  $k$ .

(b)  $\omega \in \liminf A_n$  se, e somente se,  $\omega \notin A_k$  para um número finito de índices  $k$ .

**Prova:** Para parte (a), note que  $\omega \in \limsup A_n$ , se, e somente se, para todo  $n$ ,  $\omega \in \bigcup_{k=n}^{\infty} A_k$ , ou seja, se, e somente se, para todo  $n$  existe  $n' \geq n$  tal que  $\omega \in A_{n'}$ . Como isto é válido para todo  $n$ , temos que isto é equivalente a existência de um número infinito de índices  $k$  tais que  $\omega \in A_k$ .

A prova da parte (b) é similar. ■

A seguir descreveremos algumas propriedades do  $\liminf$  e  $\limsup$  de uma seqüência de eventos.

1.  $\liminf A_n \subseteq \limsup A_n$

Este fato é uma simples consequência do Teorema 6.1.2, pois se  $\omega \in \liminf A_n$ ,  $\omega$  não pertence apenas a um número finito de eventos  $A_k$ 's, e consequentemente pertence a um número infinito deles. Logo,  $\omega \in \limsup A_n$ .

$$2. (\liminf A_n)^c = \limsup A_n^c$$

Este fato decorre aplicando a Lei de De Morgan duas vezes:

$$(\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k)^c = \cap_{n=1}^{\infty} (\cap_{k=n}^{\infty} A_k)^c = \cap_{n=1}^{\infty} (\cup_{k=n}^{\infty} A_k^c).$$

### Seqüências Monotônicas

Uma seqüência de eventos  $(A_n)$  é monotônica não-decrescente (resp., não-crescente) se  $A_1 \subseteq A_2 \subseteq \dots$  (resp.,  $A_1 \supseteq A_2 \supseteq \dots$ ). Denotaremos por  $A_n \uparrow$  (resp.,  $A_n \downarrow$ ) uma seqüência não-decrescente (resp. não-crescente) de eventos.

**Teorema 6.1.3:** *Suponha que  $(A_n)$  é uma seqüência monotônica de eventos. Então,*

$$1. \text{ Se } A_n \uparrow, \text{ então } \lim_n A_n = \cup_{n=1}^{\infty} A_n.$$

$$2. \text{ Se } A_n \downarrow, \text{ então } \lim_n A_n = \cap_{n=1}^{\infty} A_n.$$

Conseqüentemente, como para qualquer seqüência  $B_n$ , temos  $\inf_{k \geq n} B_k \uparrow$  e  $\sup_{k \geq n} B_k \downarrow$ , segue que:

$$\liminf B_n = \lim_n (\inf_{k \geq n} B_k), \quad \limsup B_n = \lim_n (\sup_{k \geq n} B_k)$$

**Prova:** Para provar (1), precisamos mostrar que  $\liminf A_n = \limsup A_n = \cup_{n=1}^{\infty} A_n$ . Como  $A_j \subseteq A_{j+1}$ , temos  $\cap_{k \geq n} A_k = A_n$ , e portanto,

$$\liminf A_n = \cup_{n=1}^{\infty} (\cap_{k \geq n} A_k) = \cup_{n=1}^{\infty} A_n.$$

Por outro lado, temos,

$$\begin{aligned} \limsup A_n &= \cap_{n=1}^{\infty} (\cup_{k \geq n} A_k) \subseteq \cup_{k=1}^{\infty} A_k \\ &= \liminf A_n \subseteq \limsup A_n. \end{aligned}$$

Logo, temos igualdade acima, ou seja,  $\limsup A_n = \cup_{k=1}^{\infty} A_k$ .

A prova de (2) é similar. ■

### Exemplo 6.1.4:

$$1. \lim_n [0, 1 - \frac{1}{n}] = \cup_{n=1}^{\infty} [0, 1 - \frac{1}{n}] = [0, 1).$$

$$2. \lim_n [0, 1 + \frac{1}{n}) = \cap_{n=1}^{\infty} [0, 1 + \frac{1}{n}) = [0, 1].$$

$$3. \lim_n (\frac{n}{n+1}, \frac{n}{n-1}) = \cap_{n=1}^{\infty} (\frac{n}{n+1}, \frac{n}{n-1}) = \{1\}.$$

**Exemplo 6.1.5:** Sejam  $A_n, A, B_n, B$  eventos em  $\Omega$ . Mostre que:

$$1. \text{ se } \lim_n A_n = A, \text{ então } \lim_n A_n^c = A^c.$$

**Solução:**  $\liminf A_n^c = (\limsup A_n)^c = A^c$  e  $\limsup A_n^c = (\liminf A_n)^c = A^c$ .

$$2. \limsup(A_n \cup B_n) = \limsup A_n \cup \limsup B_n.$$

**Solução:** Se  $\omega \in \limsup(A_n \cup B_n)$ , então  $\omega \in (A_k \cup B_k)$  para infinitos índices  $k$ . Logo, temos que  $\omega \in A_k$  para infinitos índices  $k$ , ou  $\omega \in B_k$  para infinitos índices  $k$ . Portanto, temos  $\omega \in \limsup A_n$  ou  $\omega \in \limsup B_n$ , ou seja,  $\omega \in \limsup A_n \cup \limsup B_n$ .

Reciprocamente, se  $\omega \in \limsup A_n \cup \limsup B_n$ , então  $\omega \in \limsup A_n$  ou  $\omega \in \limsup B_n$ . Logo, temos que  $\omega \in A_k$  para infinitos índices  $k$ , ou  $\omega \in B_k$  para infinitos índices  $k$ , ou seja,  $\omega \in (A_k \cup B_k)$  para infinitos índices  $k$ . Portanto,  $\omega \in \limsup(A_n \cup B_n)$ .

$$3. \text{ Não é verdade que } \liminf(A_n \cup B_n) = \liminf A_n \cup \liminf B_n.$$

**Solução:** Vamos construir um contra-exemplo: Suponha que  $A \cap B = \emptyset$ ,  $A_n = A \neq \emptyset$  e  $B_n = B \neq \emptyset$  para  $n$  par; e  $A_n = B$  e  $B_n = A$  para  $n$  ímpar. Como  $A_n \cup B_n = A \cup B$  para todo  $n$ , é fácil ver que  $\liminf(A_n \cup B_n) = A \cup B$ . Também é fácil ver que  $\liminf A_n = \liminf B_n = A \cap B = \emptyset$ , pois somente os  $\omega$ 's em  $A \cap B$  não ocorrem para um número finito de índices  $n$  tanto na seqüência  $A_n$  quanto na seqüência  $B_n$ . Então,  $A \cup B = \liminf(A_n \cup B_n) \neq \emptyset = \liminf A_n \cup \liminf B_n$ .

$$4. \text{ se } A_n \rightarrow A \text{ e } B_n \rightarrow B, \text{ então } A_n \cup B_n \rightarrow A \cup B \text{ e } A_n \cap B_n \rightarrow A \cap B.$$

**Solução:** Pela parte (2), temos que

$$\limsup A_n \cup B_n = \limsup A_n \cup \limsup B_n = A \cup B,$$

e pela propriedade (1) de  $\liminf$  e  $\limsup$ , temos

$$\liminf A_n \cup B_n \subseteq \limsup A_n \cup B_n = A \cup B.$$

Resta-nos provar que  $A \cup B \subseteq \liminf A_n \cup B_n$ . Suponha que  $\omega \in A \cup B$ , então  $\omega \in \liminf A_n$  ou  $\omega \in \liminf B_n$ , ou seja,  $\omega$  não pertence a um número finito de  $A_k$ 's, ou  $\omega$  não pertence a um número finito de  $B_k$ 's. Logo,  $\omega$  não pertence a um número finito de  $A_k \cup B_k$ 's. Portanto,  $\omega \in \liminf A_n \cup B_n$ . Então,  $A_n \cup B_n \rightarrow A \cup B$ .

Utilizando os itens anteriores e a Lei de De Morgan, temos:

$$\begin{aligned} A \cap B &= (A^c \cup B^c)^c = (\lim A_n^c \cup \lim B_n^c)^c = \\ &= (\lim A_n^c \cup B_n^c)^c = \lim(A_n^c \cup B_n^c)^c = \lim A_n \cap B_n. \end{aligned}$$

### 6.1.1 Borel-Canteli

A seguir vamos enunciar e provar um importante Lema, conhecido como Lema de Borel-Cantelli, que trata da probabilidade da ocorrência de um número infinito de eventos.

**Lema 6.1.6:** *Sejam  $A_1, A_2, \dots$  eventos aleatórios em  $(\Omega, \mathcal{A}, P)$ , ou seja,  $A_n \in \mathcal{A}, \forall n$ .*

(a) *Se  $\sum_{n=1}^{\infty} P(A_n) < \infty$ , então  $P(A_n \text{ infinitas vezes}) = 0$ .*

(b) *Se  $\sum_{n=1}^{\infty} P(A_n) = \infty$  e os eventos  $A_n$ 's são independentes, então*

$$P(A_n \text{ infinitas vezes}) = 1.$$

**Observação:** O ítem (b) não vale necessariamente sem independência. Por exemplo, seja  $A_n = A, \forall n$ , onde  $0 < P(A) < 1$ . Então,  $\sum P(A_n) = \infty$  mas o evento  $[A_n \text{ infinitas vezes}] = A$  e  $P(A_n \text{ infinitas vezes}) = P(A) < 1$ .

**Prova:** Para parte (a), se  $\sum P(A_n) < \infty$ , então  $\sum_{k=j}^{\infty} P(A_k) \rightarrow 0$  quando  $j \rightarrow \infty$ . Mas

$$[A_n \text{ infinitas vezes}] \subseteq \cup_{k=j}^{\infty} A_k, \forall j,$$

logo

$$P(A_n \text{ infinitas vezes}) \leq P(\cup_{k=j}^{\infty} A_k) \leq \sum_{k=j}^{\infty} P(A_k) \rightarrow 0.$$

Portanto,  $P(A_n \text{ infinitas vezes}) = 0$ .

Para parte (b), basta provar que

$$P(\cup_{k=n}^{\infty} A_k) = 1, \forall n$$

(pois sendo  $[A_n \text{ infinitas vezes}] = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k$  a intersecção de um número enumerável de eventos de probabilidade 1, é também de probabilidade 1). Para tanto, seja  $B_n = \cup_{k=n}^{\infty} A_k$ . Então  $B_n$  contém  $\cup_{k=n}^{n+m} A_k$  para todo  $m$ , e

$$B_n^c \subseteq (\cup_{k=n}^{n+m} A_k)^c = \cap_{k=n}^{n+m} A_k^c.$$

Logo para todo  $m$ ,

$$1 - P(B_n) = P(B_n^c) \leq P(\cap_{k=n}^{n+m} A_k^c) = \prod_{k=n}^{n+m} P(A_k^c) = \prod_{k=n}^{n+m} (1 - P(A_k)).$$

Como  $1 - p \leq e^{-p}$  para  $0 \leq p \leq 1$ , temos

$$1 - P(B_n) \leq \prod_{k=n}^{n+m} e^{-P(A_k)} = \exp(-\sum_{k=n}^{n+m} P(A_k)) \rightarrow 0$$

quando  $m \rightarrow \infty$ , pois  $\sum_{k=n}^{n+m} P(A_k) \rightarrow \infty$  quando  $m \rightarrow \infty$ . Logo  $P(B_n) = 1, \forall n$ . ■

**Exemplo 6.1.7:** Se sabemos que para uma dada coleção de eventos  $\{A_k\}$ , as suas probabilidades individuais satisfazem  $P(A_k) \leq \frac{1}{k^2}$ , então podemos concluir que infinitos desses eventos ocorrem com probabilidade zero ou, que apenas um número finito deles ocorrem com probabilidade 1. Podemos reescrever isso da seguinte forma: existe um instante aleatório  $N$  tal que, com probabilidade 1, nenhum dos  $A_k$  ocorrem para  $k > N$ . É importante ressaltar que nós podemos chegar a essa conclusão sem saber nada sobre as interações entre esses eventos como as que são expressas por probabilidades de pares de eventos  $P(A_i \cap A_j)$ . Contudo, se apenas sabemos que  $P(A_k) > 1/k$ , então não podemos concluir nada baseados no Lema de Borel-Cantelli. Se soubermos que os eventos são mutuamente independentes, então sabendo que  $P(A_k) > 1/k$ , podemos concluir que infinitos  $A_k$  ocorrem com probabilidade 1.

**Exemplo 6.1.8:** Considere uma seqüência de variáveis aleatórias  $X_1, X_2, X_3, \dots$ . Podemos usar o Lema de Borel-Cantelli para determinar a probabilidade que  $X_k > b_k$  infinitas vezes para qualquer seqüência de números reais  $\{b_k\}$ . Note que  $P(X_k > b_k) = 1 - F_{X_k}(b_k)$ . Logo, se

$$\sum_{k=1}^{\infty} P(X_k > b_k) = \sum_{k=1}^{\infty} 1 - F_{X_k}(b_k) < \infty,$$

então, não importa qual a distribuição conjunta das variáveis aleatórias  $\{X_k\}$ , temos que o evento  $\{X_k > b_k\}$  só ocorrerá para um número finito de índices  $k$ . Por outro lado, se

$$\sum_{k=1}^{\infty} P(X_k > b_k) = \sum_{k=1}^{\infty} 1 - F_{X_k}(b_k) = \infty,$$

então precisaríamos de informação adicional sobre a distribuição conjunta das variáveis aleatórias  $\{X_k\}$  para determinar se os eventos  $\{X_k > b_k\}$  ocorrem um número finito ou infinito de vezes.

**Exemplo 6.1.9:** Considere uma moeda não necessariamente honesta com probabilidade de cara igual a  $p$ , onde  $0 < p < 1$ . Se esta moeda for jogada um número infinito de vezes de maneira independente, qual a probabilidade da seqüência  $(cara, cara, coroa, coroa)$  aparecer um número infinito de vezes? Justifique sua resposta.

**Solução:** Seja  $X_i$  o resultado do  $i$ -ésimo lançamento da moeda. Defina o evento  $A_i = \{X_i = cara, X_{i+1} = cara, X_{i+2} = coroa, X_{i+3} = coroa\}$ , queremos calcular  $P(A_i \text{ infinitas vezes})$ . Note que para todo  $i$ , temos  $P(A_i) = p^2(1-p)^2 > 0$ . Não podemos aplicar diretamente o lema de Borel Cantelli, pois os eventos  $A_i$ 's não são independentes, visto que, por exemplo, ambos  $A_1$  e  $A_2$  dependem de  $X_2, X_3, X_4$ . Considere a seguinte subseqüência da seqüência de eventos  $(A_i)$  tal que  $B_i = A_{4i-3}$ . Como os eventos  $B_i$ 's dependem de famílias disjuntas de variáveis aleatórias independentes, eles são independentes. Além disso temos que  $P(B_i) = p^2(1-p)^2 > 0$ . Logo,  $\sum_i P(B_i) = \infty$ . Portanto, Borel-Cantelli implica que  $P(B_i \text{ infinitas vezes}) = 1$ . Como  $(B_i)$  é uma subseqüência de  $(A_i)$ , temos que

$$[B_i \text{ infinitas vezes}] \subseteq [A_i \text{ infinitas vezes}].$$

Portanto,  $P(A_i \text{ infinitas vezes}) = 1$ .

## 6.2 Convergência de Variáveis Aleatórias

Seguindo uma interpretação freqüentista, probabilidade está relacionada com a freqüência relativa de eventos no longo prazo. A matemática para estudar o longo prazo é a dos limites. Mas quando se trata de funções, existem vários tipos de limites (por exemplo, pontual, uniforme, em quase todo lugar). O mesmo ocorre quando consideramos limites de variáveis aleatórias definidas em um mesmo espaço de probabilidade  $(\Omega, \mathcal{A}, P)$ , visto que variáveis aleatórias são funções reais cujo domínio é  $\Omega$ .

**Relembrando:** Seja  $(\Omega, \mathcal{A})$  um espaço mensurável. Uma função  $X : \Omega \rightarrow R$  é chamada de variável aleatória se para todo evento Boreliano  $B$ ,  $X^{-1}(B) \in \mathcal{A}$ . Nós recordamos que um evento Boreliano é qualquer evento pertencente à  $\sigma$ -álgebra de Borel, onde a  $\sigma$ -álgebra de Borel é a menor  $\sigma$ -álgebra contendo intervalos da forma  $(-\infty, x]$  para todo  $x \in R$ .



### 6.2.1 Tipos de Convergência

Vamos a seguir descrever vários tipos de convergência estocástica, ilustrando com exemplos cada tipo de convergência, e depois provaremos algumas relações entre os vários tipos de convergência. Sejam  $Y, Y_1, Y_2, \dots$  variáveis aleatórias definidas em um mesmo espaço de probabilidade  $(\Omega, \mathcal{A}, P)$ .

#### Convergência Quase Certa

**Definição 6.2.1:** A seqüência de variáveis aleatórias  $Y_1, Y_2, \dots$  converge *quase certamente* (ou *com probabilidade 1*) para a variável aleatória  $Y$  se

$$P(\{w : \lim_{n \rightarrow \infty} Y_n(w) = Y(w)\}) = 1.$$

Notação:  $Y_n \rightarrow Y$  cp1.

Então se uma seqüência de variáveis aleatórias  $Y_1, Y_2, \dots$  converge quase certamente para  $Y$  não significa que para todo  $w \in \Omega$ ,  $Y_n(w) \rightarrow Y(w)$ , apenas o que se sabe é que a probabilidade do evento  $D = \{w : Y_n(w) \not\rightarrow Y(w)\}$  é nula.  $D$  é chamado de conjunto de exceção.

**Exemplo 6.2.2:** Considere uma variável aleatória  $Z$  tal que  $P(\{w : 0 \leq |Z(w)| < 1\}) = 1$ . Seja  $X_n(w) = Z^n(w)$ , então  $X_n(w) \rightarrow 0$  cp1; note que o conjunto de exceção é  $D = \{w \in \Omega : |Z(w)| \geq 1\}$  e que  $P(D) = 0$ .  $\square$

Podemos obter uma definição alternativa para convergência quase-certa, observando que, pela definição de limite de seqüências de números reais, para um dado  $w \in \Omega$  fixo, temos que  $\lim_n Y_n(w) = Y(w)$  se, e somente se, para todo  $k \in \mathbb{N}$ , existir  $N$  tal que para todo  $n \geq N$ , temos  $|Y_n(w) - Y(w)| < \frac{1}{k}$ . Portanto:

$$\{w : \lim_n Y_n(w) = Y(w)\} = \{w : \cap_{k=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} |Y_n(w) - Y(w)| < \frac{1}{k}\}.$$

Então,  $Y_n \rightarrow Y$  cp1 se, e somente se,

$$P(\{w : \cap_{k=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} |Y_n(w) - Y(w)| < \frac{1}{k}\}) = 1.$$

Isto é equivalente a:

$$P(\{w : \cap_{k=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} |Y_n(w) - Y(w)| \geq \frac{1}{k}\}) = 0.$$

Defina  $A_{n,k} = \{w : |Y_n(w) - Y(w)| \geq \frac{1}{k}\}$ . Então para cada  $k$  fixo, temos que

$$\limsup_n A_{n,k} = \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} A_{n,k}.$$

**Logo,  $Y_n \rightarrow Y$  cp1 se, e somente se, para todo  $k \in \mathbb{N}$ ,**

$$P(\limsup_n A_{n,k}) = 0.$$

**Exemplo 6.2.3:** Seja  $\{X_n\}_{n \geq 3}$  uma seqüência de variáveis aleatórias independentes com distribuição de probabilidade dada por:

$$P(X_n = 0) = 1 - \frac{1}{\log n} \text{ e } P(X_n = n) = \frac{1}{\log n}, \forall n \geq 3.$$

Mostre que  $X_n \not\rightarrow 0$  cp1.

**Solução:** Para qualquer  $\epsilon$  tal que  $0 < \epsilon < 1$ , temos que

$$P(|X_n| > \epsilon) = P(X_n = n) = \frac{1}{\log n}.$$

Logo,  $\sum_n P(|X_n| > \epsilon) = \sum_n \frac{1}{\log n} = \infty$ . Então, o Lema de Borel-Cantelli implica que  $P(|X_n| > \epsilon \text{ infinitas vezes}) = 1$ , portanto com probabilidade 1,  $X_n \not\rightarrow 0$ .

**Exemplo 6.2.4:** Considere  $\{X_n : n \geq 1\}$  uma seqüência de variáveis aleatórias i.i.d. com função de distribuição  $F$ . Suponha que  $F(x) < 1$ , para todo  $x < \infty$ . Defina  $Y_n = \max(X_1, X_2, \dots, X_n)$ . Vamos verificar que  $Y_n \rightarrow \infty$  cp1.

Inicialmente, observe que para cada  $\omega \in \Omega$ , as variáveis  $Y_n$  formam uma seqüência não-decrescente de números reais. Seja  $M$  um número real, temos

$$\begin{aligned} P(Y_n \leq M : n = 1, 2, \dots) &\leq P(Y_n \leq M : n = 1, 2, \dots, k) = P(Y_k \leq M) \\ &= P(\max(X_1, X_2, \dots, X_k) \leq M) = P(X_1 \leq M, X_2 \leq M, \dots, X_k \leq M) \\ &= \prod_{n=1}^k P(X_n \leq M) = F^k(M), \forall k \geq 1. \end{aligned}$$

Fazendo  $k \rightarrow \infty$ , temos que para todo  $M$  finito,

$$P(\lim_n Y_n \leq M) = P(Y_n \leq M : n = 1, 2, \dots) = 0;$$

pois  $F^k(M)$  tende a zero, quando  $k \rightarrow \infty$ . Dessa forma, o conjunto dos  $w \in \Omega$ , em que  $\lim_n Y_n(w)$  é finito, tem probabilidade zero e, portanto,  $Y_n \rightarrow \infty$  cp1.

### Convergência na $r$ -ésima Média

**Definição 6.2.5:** A seqüência de variáveis aleatórias  $Y_1, Y_2, \dots$  converge na  $r$ -ésima Média, onde  $r > 0$ , para a variável aleatória  $Y$  se

$$\lim_{n \rightarrow \infty} E|Y_n - Y|^r = 0.$$

Notação:  $Y_n \rightarrow^r Y$ .

Se  $r = 2$  este tipo de convergência é freqüentemente chamado de *convergência em média quadrática*.

**Exemplo 6.2.6:** Sejam  $Z, X_1, X_2, \dots$  variáveis aleatórias tais que

$$X_n = \frac{n}{n+1}Z,$$

então  $X_n \xrightarrow{2} Z$  se  $EZ^2 < \infty$ , mas não em caso contrário.  $\square$

**Exemplo 6.2.7:** Considere a seqüência de variáveis aleatórias definidas no Exemplo 6.2.3. Mostre que  $X_n \not\xrightarrow{r} 0$ , para todo  $r > 0$ .

**Solução:** Temos que

$$E|X_n|^r = n^r P(X_n = n) = \frac{n^r}{\log n} \rightarrow \infty.$$

Logo,  $X_n \not\xrightarrow{r} 0$ .

O próximo teorema afirma que se  $X_n \xrightarrow{r} X$ , então  $X_n \xrightarrow{s} X$  para  $s < r$ .

**Teorema 6.2.8:** Se  $X_n \xrightarrow{r} X$ , então  $X_n \xrightarrow{s} X$  para  $0 < s < r$

**Prova:** Defina  $p = \frac{r}{s} > 1$  e  $q = \frac{r}{r-s}$ . Então,

$$\frac{1}{p} + \frac{1}{q} = \frac{s}{r} + \frac{r-s}{r} = 1.$$

Seja  $Z = |X|^s$  e  $Y = 1$ . Com estas definições, a desigualdade de Hölder implica que

$$E|ZY| \leq (E|Z|^p)^{1/p} (E|Y|^q)^{1/q},$$

ou seja,

$$E(|X|^s) \leq (E|X|^{ps})^{1/p} 1 = (E|X|^r)^{s/r}.$$

Substituindo  $X$  por  $X_n - X$ , temos

$$E(|X_n - X|^s) \leq (E|X_n - X|^r)^{s/r}.$$

Portanto, se  $\lim_n E|X_n - X|^r = 0$ , então  $\lim_n E|X_n - X|^s = 0$ . ■

## Convergência em Probabilidade

**Definição 6.2.9:** A seqüência de variáveis aleatórias  $Y_1, Y_2, \dots$  converge *em probabilidade* para a variável aleatória  $Y$  se  $\forall \epsilon > 0$

$$\lim_{n \rightarrow \infty} P(\{w : |Y_n(w) - Y(w)| > \epsilon\}) = 0.$$

Notação:  $Y_n \xrightarrow{P} Y$ .

A intuição por trás desta definição é que para  $n$  muito grande a probabilidade de que  $Y_n$  e  $Y$  sejam bem próximas é bastante alta.

**Exemplo 6.2.10:** Considere a seqüência de variáveis aleatórias definidas no Exemplo 6.2.3. Mostre que  $X_n \rightarrow^P 0$ . **Solução:** Temos que para  $0 < \epsilon < 1$ ,  $P(|X_n| > \epsilon) = P(X_n = n)$  e para  $\epsilon \geq 1$ ,  $P(|X_n| > \epsilon) \leq P(X_n = n)$ . Como  $P(X_n = n) = \frac{1}{\log n} \rightarrow 0$ , temos que  $\forall \epsilon > 0$ ,  $\lim P(|X_n| > \epsilon) = 0$ . Portanto,  $X_n \rightarrow^P 0$ .

**Exemplo 6.2.11:** Considere  $X, X_1, X_2, \dots$  onde as variáveis aleatórias têm distribuição normal conjunta, todas com média 0 e matriz de covariância parcialmente descrita por

$$COV(X, X) = COV(X_n, X_n) = 1, COV(X, X_n) = 1 - \frac{1}{n}.$$

Seja  $Y_n = X_n - X$ , como  $Y_n$  é uma combinação linear de variáveis aleatórias com distribuição normal, ela também possui distribuição normal. Precisamos determinar então sua média e sua variância. Mas  $EY = E(X_n - X) = EX_n - EX = 0$  e

$$VarY = EY^2 = E(X_n - X)^2 = EX_n^2 - 2EX_nX + EX^2 = 1 - 2(1 - \frac{1}{n}) + 1 = \frac{2}{n}.$$

Portanto,  $Y_n \sim \mathcal{N}(0, \frac{2}{n})$ . Então,

$$P(|X_n - X| > \epsilon) = P(|Y_n| > \epsilon) = 2P(Y_n > \epsilon) = 2 \int_{\epsilon}^{\infty} \frac{\sqrt{n}}{\sqrt{4\pi}} e^{-\frac{ny^2}{4}} dy = 2 \int_{\epsilon\sqrt{\frac{n}{2}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Logo,  $\forall \epsilon > 0$ ,  $\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0$ , ou seja,  $X_n \rightarrow^P X$ .  $\square$

### Convergência em Distribuição

O último tipo de convergência estocástico que mencionamos não é exatamente uma noção de convergência das variáveis aleatórias propriamente ditas, mas uma noção de convergência de suas respectivas funções de distribuição acumuladas.

**Definição 6.2.12:** A seqüência de variáveis aleatórias  $Y_1, Y_2, \dots$ , converge em distribuição para a variável aleatória  $Y$  se para todo ponto  $x$  de continuidade de  $F_Y$

$$\lim_{n \rightarrow \infty} F_{Y_n}(x) = F_Y(x).$$

Notação:  $Y_n \rightarrow^D Y$ .

**Exemplo 6.2.13:** Seja  $\{X_n : n \geq 1\}$  uma seqüência de variáveis aleatórias independentes com distribuição Uniforme em  $(0, b)$ ,  $b > 0$ . Defina  $Y_n = \max(X_1, X_2, \dots, X_n)$  e  $Y = b$ . Vamos verificar que  $Y_n \rightarrow^D Y$ . Temos

$$F_{Y_n}(y) = P(\max(X_1, X_2, \dots, X_n) \leq y) = F_{X_1}^n(y) = \begin{cases} 0 & \text{se } y < 0, \\ (\frac{y}{b})^n & \text{se } 0 \leq y < b, \\ 1 & \text{se } y \geq b. \end{cases}$$

Fazendo  $n$  tender ao infinito, temos que

$$\lim_n F_{Y_n}(y) = \begin{cases} 0 & \text{se } y < b, \\ 1 & \text{se } y \geq b, \end{cases}$$

que corresponde à função de distribuição de  $Y$  e, portanto,  $Y_n \rightarrow^D Y$ .

Deve-se ficar atento que convergência em distribuição não implica nada em relação aos outros tipos de convergência. Uma seqüência convergindo em distribuição para uma variável aleatória  $X$  também converge em distribuição para qualquer outra variável aleatória  $Y$  tal que  $F_Y = F_X$ . O próximo exemplo serve para ilustrar melhor este fato.

**Exemplo 6.2.14:** Se uma seqüência de variáveis aleatórias  $Y_1, Y_2, \dots$  é independente e identicamente distribuída de acordo com  $F$ , então para todo  $n$  tem-se que  $F_{Y_n} = F$ , logo a seqüência converge em distribuição para qualquer variável aleatória  $X$  tal que  $F_X = F$ . Claro, como a seqüência é independente, os valores de termos sucessivos são independentes e não exibem nenhum comportamento usual de convergência.  $\square$

O requisito de continuidade, mencionado na definição acima, se justifica para evitar algumas anomalias. Por exemplo, para  $n \geq 1$  seja  $X_n = \frac{1}{n}$  e  $X = 0$ , para todo  $\Omega$ . Parece aceitável que deveríamos ter convergência de  $X_n$  para  $X$ , qualquer que fosse o modo de convergência. Observe que

$$F_n(x) = \begin{cases} 0 & \text{se } x < \frac{1}{n}, \\ 1 & \text{se } x \geq \frac{1}{n}, \end{cases} \text{ e}$$

$$F(x) = \begin{cases} 0 & \text{se } x < 0, \\ 1 & \text{se } x \geq 0. \end{cases}$$

Portanto, como  $\lim_n F_n(0) = 0 \neq F(0) = 1$ , não temos  $\lim_n F_n(x) = F(x)$  para todo  $x \in \mathbb{R}$ . Desse modo se houvesse a exigência de convergência em todos os pontos, não teríamos convergência em distribuição. Entretanto, note que para  $x \neq 0$ , temos  $\lim_n F_n(x) = F(x)$  e, como o ponto 0 não é de continuidade de  $F$ , concluímos que  $X_n \rightarrow^D X$ .

Um exemplo mais complexo de convergência em distribuição pode ser visto na análise do limite de

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - EX_i),$$

onde  $X_i$ 's são variáveis aleatórias independentes e identicamente distribuídas. Neste, o Teorema Central do Limite afirma que se  $VAR(X_i) = \sigma^2 < \infty$ , então  $S_n$  converge em distribuição para qualquer variável aleatória com distribuição  $\mathcal{N}(0, \sigma^2)$ .

O próximo teorema estabelece duas condições suficientes para que uma seqüência de variáveis aleatórias convirja em distribuição.

**Teorema 6.2.15:** *Seja  $X, X_1, X_2, \dots$  uma seqüência de variáveis aleatórias:*

- (a) *Se  $X, X_1, X_2, \dots$  são variáveis aleatórias discretas com  $P(X_n = x_i) = p_n(i)$  e  $P(X = x_i) = p(i)$ , onde  $p_n(i) \rightarrow p(i)$  quando  $n \rightarrow \infty$  para todo  $i = 0, 1, 2, 3, \dots$ , então  $X_n \rightarrow^D X$ .*
- (b) *Se  $X, X_1, X_2, \dots$  são variáveis aleatórias absolutamente contínuas com densidades dadas respectivamente por  $f, f_1, f_2, f_3, \dots$ , onde  $f_n(x) \rightarrow f(x)$  quando  $n \rightarrow \infty$  em quase todo lugar, então  $X_n \rightarrow^D X$ .*

**Prova:** Se  $p_n(i) \rightarrow p(i)$  para todo  $i$ , então

$$F_{X_n}(x) = \sum_{i:x_i \leq x} p_n(i) \rightarrow \sum_{i:x_i \leq x} p(i) = F_X(x).$$

Onde a convergência acima segue do Teorema da Convergência Dominada, visto que  $F_{X_n}(x) \leq 1, \forall x \in \mathbb{R}$ .

A prova da parte (b) usa conceitos de Teoria da Medida e será omitida. ■

O próximo exemplo mostra que se uma seqüência de variáveis aleatórias discretas converge em distribuição, não necessariamente sua função probabilidade de massa converge.

**Exemplo 6.2.16:** Sejam  $X, X_1, X_2, \dots$  variáveis aleatórias tais que  $P(X = 0) = 1$  e  $P(X_n = 1/n) = 1$ . Então, temos  $F_X(x) = 1$  se  $x \geq 0$ , e  $F_X(x) = 0$  caso contrário; e  $F_{X_n}(x) = 1$  se  $x \geq 1/n$  e  $F_{X_n}(x) = 0$  caso contrário. Logo,  $F_{X_n}(x) \rightarrow F_X(x), \forall x \neq 0$ , ou seja,  $X_n \rightarrow^D X$ . Porém,  $p(0) = 1 \neq 0 = \lim_n p_n(0)$ .

O próximo exemplo mostra que se uma seqüência de variáveis aleatórias absolutamente contínuas converge em distribuição, não necessariamente sua função densidade de probabilidade converge.

**Exemplo 6.2.17:** Considere uma seqüência de variáveis aleatórias  $X, X_1, X_2, \dots$  com função de distribuição acumuladas dadas respectivamente por  $F, F_1, F_2, F_3, \dots$ , onde

$$F_n(x) = \begin{cases} 0 & , \text{ se } x \leq 0 \\ x(1 - \frac{\text{sen}2n\pi x}{2n\pi x}) & , \text{ se } 0 < x \leq 1 \\ 1 & , \text{ se } x > 1; \end{cases}$$

e

$$F(x) = \begin{cases} 0 & , \text{ se } x \leq 0 \\ x & , \text{ se } 0 < x \leq 1 \\ 1 & , \text{ se } x > 1. \end{cases}$$

Então  $F_n$  e  $F$  são absolutamente contínuas com densidade dada por

$$f_n(x) = \begin{cases} 1 - \cos 2n\pi x & , \text{ se } 0 \leq x \leq 1 \\ 0 & , \text{ caso contrário;} \end{cases}$$

e

$$f(x) = \begin{cases} 1 & , \text{ se } 0 < x \leq 1 \\ 0 & , \text{ caso contrário.} \end{cases}$$

É fácil ver que  $F_n(x) \rightarrow F(x), \forall x \in \mathbb{R}$ . Contudo,  $f_n(x) \not\rightarrow f(x)$ .

### 6.2.2 Relação Entre os Tipos de Convergência

A primeira relação que iremos provar é que convergência quase certa implica convergência em probabilidade.

**Teorema 6.2.18:**  $X_n \rightarrow X \text{ cp1} \Rightarrow X_n \rightarrow^P X$ .

**Prova:** Para provar que convergência quase certa implica em convergência em probabilidade, considere a seguinte família de eventos

$$A_{n,\epsilon} = \{w : |X_n(w) - X(w)| \leq \epsilon\}.$$

Logo, pela interpretação de convergência pontual,

$$C = \{w : X_n(w) \rightarrow X(w)\} = \bigcap_{\epsilon > 0} \bigcup_{N=1}^{\infty} \bigcap_{n \geq N} A_{n,\epsilon}.$$

Se  $X_n \rightarrow X \text{ cp1}$ , então  $P(C) = 1$ . Equivalentemente, pela Lei de De Morgan,

$$D = C^c = \bigcup_{\epsilon > 0} D_\epsilon, \text{ onde } D_\epsilon = \bigcap_{N=1}^{\infty} \bigcup_{n \geq N} A_{n,\epsilon}^c,$$

e

$$P(\bigcup_{\epsilon > 0} D_\epsilon) = 0.$$

Portanto, convergência quase certa implica que  $\forall \epsilon > 0$ ,  $P(D_\epsilon) = 0$ . Seja  $F_N = \bigcup_{n \geq N} B_n$ . Note que  $F_N \downarrow$ . Logo,  $\lim_N F_N = \bigcap_{N=1}^{\infty} \bigcup_{n \geq N} B_n$ . Portanto, pelo axioma da continuidade monotônica da probabilidade, tem-se que

$$P(\bigcap_{N=1}^{\infty} \bigcup_{n \geq N} B_n) = \lim_{N \rightarrow \infty} P(\bigcup_{n \geq N} B_n).$$

Então,

$$\begin{aligned} 0 = P(D_\epsilon) &= \lim_{N \rightarrow \infty} P(\bigcup_{n \geq N} A_{n,\epsilon}^c) \geq \\ &\lim_{N \rightarrow \infty} P(A_{N,\epsilon}^c) = \lim_{N \rightarrow \infty} P(|X_N(w) - X(w)| > \epsilon). \end{aligned}$$

Portanto,  $X_n \rightarrow^P X$ . ■

O próximo teorema prova que convergência na  $r$ -ésima média implica convergência em probabilidade.

**Teorema 6.2.19:**  $X_n \rightarrow^r X \Rightarrow X_n \rightarrow^P X$ .

**Prova:** Primeiro note que  $\frac{|X_n - X|^r}{\epsilon^r} \geq I_{\{w: |X_n - X| > \epsilon\}}$ . Logo, tem-se que

$$E\left(\frac{|X_n - X|^r}{\epsilon^r}\right) \geq E(I_{\{w: |X_n - X| > \epsilon\}}),$$

ou seja,

$$\frac{E(|X_n - X|^r)}{\epsilon^r} \geq P(\{w : |X_n - X| > \epsilon\}).$$

Se  $X_n \rightarrow^r X$ , tem-se que  $\lim_{n \rightarrow \infty} E(|X_n - x|^r) = 0$ . Então, para todo  $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(\{w : |X_n - X| > \epsilon\}) = 0,$$

ou seja,  $X_n \rightarrow^P X$ . ■

O próximo exemplo prova que nem convergência em probabilidade, nem convergência na  $r$ -ésima média implicam convergência quase certa.

**Exemplo 6.2.20:** Seja  $X$  uma variável aleatória com distribuição uniforme no intervalo  $[0, 1]$ , e considere a seqüência de intervalos definida por

$$I_{2^m+i} = \left[\frac{i}{2^m}, \frac{i+1}{2^m}\right],$$

para  $m = 0, 1, 2, \dots$  e  $i = 0, 1, \dots, 2^m - 1$ .

Note que tem-se  $2^m$  intervalos de comprimento  $2^{-m}$  que cobrem todo o intervalo  $[0, 1]$ , e o comprimento dos intervalos fica cada vez menor tendendo a 0. Definamos

$$Y_n(w) = \begin{cases} 1 & \text{se } X(w) \in I_n, \\ 0 & \text{se } X(w) \notin I_n. \end{cases}$$

A seqüência  $Y_1, Y_2, \dots$  converge em probabilidade para 0, pois para  $0 < \epsilon \leq 1$ ,

$$P(|Y_n| \geq \epsilon) = P(Y_n = 1) = P(X \in I_n),$$

e esta probabilidade, que é igual ao comprimento de  $I_n$ , converge para zero quando  $n \rightarrow \infty$ .

Esta seqüência também converge na  $r$ -ésima média para todo  $r > 0$ , visto que  $E(|Y_n|^r) = P(Y_n = 1) \rightarrow 0$  quando  $n \rightarrow \infty$ . Logo,  $Y_n$  converge na  $r$ -ésima média para 0.

Porém para todo  $w \in \Omega$ ,  $Y_n(w) = 1$  para um número infinito de  $n$ 's e  $Y_n(w) = 0$  para um número infinito de  $n$ 's. Portanto,  $Y_n(w)$  não converge para todo  $w$ , o que implica que  $Y_n$  não converge quase certamente. □

O próximo teorema estabelece mais uma relação entre convergência quase certa e convergência em probabilidade.

**Teorema 6.2.21:**  $X_n \rightarrow^P X$  se, e somente se, toda subseqüência  $\{X_{n_k}\}$  possui uma outra subseqüência  $\{X_{n_{k(i)}}\}$  tal que  $X_{n_{k(i)}} \rightarrow X$  cp1 para  $i \rightarrow \infty$ .

**Prova:** Suponha que  $X_n \rightarrow^P X$ , então dada qualquer subseqüência  $\{X_{n_k}\}$ , escolha uma outra subseqüência  $\{X_{n_{k(i)}}\}$  tal que  $j \geq k(i)$  implica que  $P(|X_{n_j} - X| \geq i^{-1}) < 2^{-i}$ . Em particular, temos que  $P(|X_{n_{k(i)}} - X| \geq i^{-1}) < 2^{-i}$ . Seja  $A_i = \{|X_{n_{k(i)}} - X| \geq i^{-1}\}$ , então  $\sum_{i=1}^{\infty} P(A_i) < \sum_{i=1}^{\infty} 2^{-i} = 1 < \infty$ . Logo, pelo Lema de Borel-Cantelli, temos que  $P(A_i \text{ infinitas vezes}) = 0$ , ou seja,  $P(A_i \text{ finitas vezes}) = 1$ . Portanto,  $|X_{n_{k(i)}} - X| < i^{-1}$  exceto para um número finito de  $i$ 's com probabilidade 1. Portanto,  $X_{n_{k(i)}} \rightarrow X$  cp1.

Se  $X_n$  não converge para  $X$  em probabilidade, existe um  $\epsilon > 0$  e uma subseqüência  $\{X_{n_k}\}$  tal que  $P(|X_{n_k} - X| > \epsilon) > \epsilon$ . Logo nenhuma subseqüência de  $\{X_{n_k}\}$  pode convergir para  $X$



em probabilidade, logo pelo Teorema 6.2.18, nenhuma subsequência converge para  $X$  quase certamente. ■

O próximo exemplo mostra que convergência em probabilidade não implica convergência na  $r$ -ésima média

**Exemplo 6.2.22:** Seja  $X$  uma variável aleatória com distribuição uniforme no intervalo  $[0, 1]$ . Considere a seguinte seqüência de variáveis aleatórias

$$Y_n(w) = \begin{cases} 2^n & \text{se } X(w) \in (0, \frac{1}{n}), \\ 0 & \text{se } X(w) \notin (0, \frac{1}{n}). \end{cases}$$

Então,  $P(|Y_n| > \epsilon) = P(X(w) \in (0, \frac{1}{n})) = \frac{1}{n} \rightarrow 0$ , mas  $E(|Y_n|^r) = 2^{nr} \frac{1}{n} \rightarrow \infty$ .

O próximo exemplo mostra que convergência quase-certa não implica convergência na  $r$ -ésima média.

**Exemplo 6.2.23:** Seja  $\{Y_n, n \geq 1\}$  uma seqüência de variáveis aleatórias onde

$$P(Y_n = 0) = 1 - n^{-2} \text{ e } P(Y_n = e^n) = n^{-2}.$$

Portanto, para todo  $\epsilon > 0$ ,

$$P(|Y_n| > \epsilon) = P(Y_n > \epsilon) \leq P(Y_n = e^n) = n^{-2}.$$

Logo,

$$\sum_{n=1}^{\infty} P(|Y_n| > \epsilon) \leq \sum_{n=1}^{\infty} n^{-2} < \infty.$$

Então, Borel-Cantelli implica que  $|Y_n| > \epsilon$  infinitas vezes com probabilidade 0, o que por sua vez implica que  $Y_n \rightarrow 0$  com probabilidade 1, ou seja,  $Y_n \rightarrow 0$  cp1. Porém,

$$E|Y_n|^r = \frac{e^{nr}}{n^2} \rightarrow \infty,$$

para todo  $r > 0$ . Portanto,  $Y_n \rightarrow 0$  cp1, mas  $Y_n \not\rightarrow^r 0$  para todo  $r > 0$ .

O próximo teorema trata da relação entre convergência em distribuição e convergência em probabilidade.

**Teorema 6.2.24:** *As seguintes relações entre os tipos de convergência são válidas:*

(a)  $X_n \rightarrow^P X \Rightarrow X_n \rightarrow^D X$

(b) *Se  $X_n \rightarrow^D c$ , onde  $c$  é uma constante, então  $X_n \rightarrow^P c$ .*

**Prova:** Para parte (a), suponha que  $X_n \xrightarrow{P} X$  e seja  $x$  um ponto de continuidade de  $F_X$ . Queremos provar que  $F_{X_n}(x) \rightarrow F_X(x)$  quando  $n \rightarrow \infty$ .

Como para  $\epsilon > 0$ ,  $X_n \leq x \Rightarrow X \leq x + \epsilon$  ou  $|X - X_n| > \epsilon$ , temos  $\{w : X_n(w) \leq x\} \subseteq \{w : X(w) \leq x + \epsilon\} \cup \{w : |X_n(w) - X(w)| > \epsilon\}$ . Logo,

$$F_{X_n}(x) = P(X_n \leq x) \leq F_X(x + \epsilon) + P(|X_n - X| > \epsilon).$$

Por outro lado,  $X \leq x - \epsilon \Rightarrow X_n \leq x$  ou  $|X_n - X| > \epsilon$  de modo que

$$F_X(x - \epsilon) \leq F_{X_n}(x) + P(|X_n - X| > \epsilon).$$

Juntando as duas desigualdades, temos que  $\forall \epsilon > 0$ , and  $\forall n$ ,

$$F_X(x - \epsilon) - P(|X_n - X| > \epsilon) \leq F_{X_n}(x) \leq F_X(x + \epsilon) + P(|X_n - X| > \epsilon).$$

Como  $X_n \xrightarrow{P} X$ , para qualquer  $\delta > 0$ , existe  $N$  tal que para  $n \geq N$ , temos que

$$F_X(x - \epsilon) - \delta \leq F_{X_n}(x) \leq F_X(x + \epsilon) + \delta.$$

Finalmente, como  $x$  é ponto de continuidade de  $F_X$ , para  $\epsilon$  suficientemente pequeno, temos que

$$F_X(x) - 2\delta \leq F_X(x - \epsilon) - \delta \leq F_{X_n}(x) \leq F_X(x + \epsilon) + \delta \leq F_X(x) + 2\delta.$$

Ou seja,  $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$ .

Para parte (b), suponha que  $X_n \xrightarrow{D} c$ . Note que a função de distribuição de uma variável aleatória constante  $c$  é:

$$F_c(x) = \begin{cases} 1 & \text{se } x \geq c, \\ 0 & \text{se } x < c. \end{cases}$$

Pela convergência em distribuição, tem-se que  $\lim_{n \rightarrow \infty} F_{X_n}(x) = 0$ , se  $x < c$  e  $\lim_{n \rightarrow \infty} F_{X_n}(x) = 1$ , se  $x > c$ . Logo, para  $\epsilon > 0$ ,

$$P(|X_n - c| \leq \epsilon) = P(c - \epsilon \leq X_n \leq c + \epsilon) \geq P(c - \epsilon < X_n \leq c + \epsilon) = F_{X_n}(c + \epsilon) - F_{X_n}(c - \epsilon) \rightarrow 1 \text{ quando } n \rightarrow \infty.$$

Ou seja,  $\forall \epsilon > 0$ ,  $\lim_{n \rightarrow \infty} P(|X_n - c| > \epsilon) = 0$ . ■

A Figura 6.1 resume a relação entre os tipos de convergência.

**Exemplo 6.2.25:** Para  $n \geq 1$ ,  $X_n \sim U(0, 1)$  são variáveis aleatórias i.i.d. Defina  $Y_n = \min(X_1, X_2, \dots, X_n)$  e  $U_n = nY_n$ . Mostre que

(a)  $Y_n \xrightarrow{P} 0$ ,

(b)  $U_n \xrightarrow{D} U$ , sendo  $U \sim Exp(1)$ .

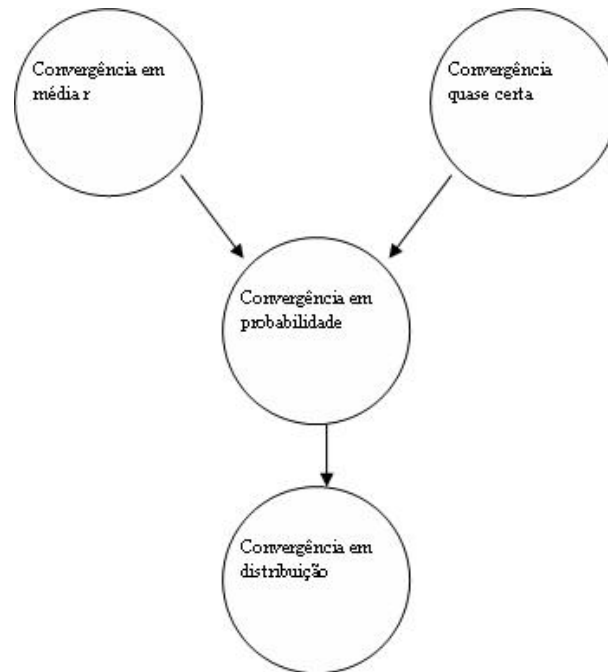


Figura 6.1: Relação entre os tipos de convergência.

**Solução:** Para parte (a), note que

$$P(|Y_n| > \epsilon) = P(Y_n > \epsilon) = P(X_1 > \epsilon, X_2 > \epsilon, \dots, X_n > \epsilon).$$

Como os  $X_n$  são independentes temos que a última expressão é igual a

$$(P(X_1 > \epsilon))^n = (1 - \epsilon)^n.$$

Como  $(1 - \epsilon)^n \rightarrow 0$  quando  $n \rightarrow \infty$ , temos que  $Y_n \xrightarrow{P} 0$ .

Para parte (b), note que

$$F_{U_n}(x) = P(U_n \leq x) = 1 - P(U_n > x) = 1 - P(nY_n > x) = 1 - P(Y_n > x/n)$$

De acordo com a parte (a), esta expressão é igual a  $1 - (1 - x/n)^n$ , que por sua vez converge para  $1 - e^{-x}$  quando  $n \rightarrow \infty$ , que é igual a  $F_U(x)$ .

## 6.3 Convergência de Vetores Aleatórios

Para o caso vetorial as definições de convergência sofrem algumas adaptações. Para as convergências quase certa e em probabilidade, precisamos avaliar a proximidade entre os vetores aleatórios  $X_n$  e  $X$  pelo comportamento da norma da diferença entre eles. Em geral, essa norma é calculada por  $\|X_n - X\| = (\sum_{j=1}^k (X_{nj} - X_j)^2)^{1/2}$ , onde  $k$  é a dimensão dos vetores e  $X_{nj}$  a coordenada  $j$  do vetor  $X_n$ . Pode-se verificar que a convergência do vetor aleatório, quase certamente ou em probabilidade, ocorre se, e somente se, existir a mesma

convergência em cada uma das variáveis que compõe o vetor aleatório. Dessa forma, o caso multidimensional pode ser estudado a partir de repetidas aplicações do caso univariado.

Para convergência em distribuição de vetores aleatórios, requeremos que a função de distribuição conjunta  $F_n(x)$  convirja para  $F(x)$ , em todos os pontos de continuidade da função  $F$ . Entretanto, lembremos que da função de distribuição conjunta podemos obter as marginais, mas o caminho inverso nem sempre é possível. Por essa razão, diferentemente das convergências quase certa e em probabilidade, não podemos reduzir o estudo da convergência em distribuição de vetores aleatórios, ao comportamento das suas respectivas coordenadas. Não temos equivalência, mas apenas implicação, em uma das direções. Ou seja, se o vetor converge em distribuição então cada componente também converge em distribuição, para a correspondente marginal da função de distribuição limite. Entretanto a recíproca não é em geral, verdadeira.

# Capítulo 7

## Funções Características

### 7.1 Motivação

Em matemática e suas aplicações, é sempre valioso ter maneiras alternativas de representar o mesmo objeto matemático. Uma analogia pode ser que um conjunto de vetores pode ser representado em vários sistemas de coordenadas. No nosso caso de probabilidade, o conceito básico é o de uma medida de probabilidade  $P$  que dá um valor real numérico a um conjunto de eventos em uma  $\sigma$ -álgebra. Para  $X$  uma variável aleatória, sabe-se que existem outras maneiras de representar a probabilidade  $P$ , como por exemplo através de sua função de distribuição acumulada  $F_X$ . Se  $X$  for uma variável aleatória discreta, pode-se equivalentemente representar  $P$  pela função de probabilidade de  $X$ ,  $p_X$ . Se  $X$  for absolutamente contínua, então  $P$  pode ser representada pela função densidade de probabilidade de  $X$ ,  $f_X$ . Uma função característica  $\varphi_X$  de uma variável aleatória  $X$  é uma outra maneira de representar  $P$ . Algumas vantagens do uso da função característica são: pode-se calcular os momentos de uma variável aleatória  $X$  diferenciando-se a função característica (o que geralmente é mais simples que usar diretamente as definições de momento que envolvem integrais), pode-se calcular mais facilmente a distribuição de soma de variáveis aleatórias independentes, e finalmente o uso de funções características ajuda na prova de uma família de Teoremas Centrais do Limite que ajudam a explicar a prevalência de distribuições normal ou Gaussianas na Natureza.

Uma função geratriz de momento é uma outra representação alternativa da distribuição de uma variável aleatória. As vantagens desta representação são as mesmas da função característica, mas como a função característica é mais robusta (no sentido que ela sempre existe), nós focaremos no uso da mesma, e apenas no final deste capítulo mencionaremos a definição de uma função geratriz de momento.

Até aqui, só tratamos com variáveis reais, mas o caso complexo é similar. Sem aprofundar o assunto, diremos que uma variável aleatória  $X$  é *complexa*, se pode ser escrita como  $X = X_1 + iX_2$ , onde  $i = \sqrt{-1}$ , e  $X_1$  e  $X_2$  são variáveis aleatórias reais. Logo, para verificar que uma função complexa é variável aleatória, precisamos verificar propriedades da imagem inversa nas suas duas partes. Para o valor esperado de  $X$ , exige-se que as duas partes sejam finitas. Assim, temos:  $EX = EX_1 + iEX_2$ , onde  $EX_1$  e  $EX_2$  são ambas finitas. Para efeitos práticos, quando realizando integração de funções complexas, podemos operar como

se estivéssemos com funções reais (trata-se  $i$  como se fosse uma constante real).

## 7.2 Definição

**Definição 7.2.1:** A função característica  $\varphi_X$  de uma variável aleatória  $X$  é dada por:

$$\varphi_X(t) = Ee^{itX} = E \cos(tX) + iE \operatorname{sen}(tX), \text{ onde } i \doteq \sqrt{-1}.$$

Note que como  $\cos(tX)$  e  $\operatorname{sen}(tX)$  são variáveis aleatórias limitadas, a esperança na definição acima é finita e, conseqüentemente, a função característica de qualquer variável aleatória é bem definida. Note também que de acordo com esta definição, a função de distribuição acumulada determina a função característica de uma variável aleatória.

No caso particular de uma variável aleatória discreta, temos:

$$\varphi_X(t) = \sum_k e^{itx_k} p(x_k),$$

onde  $p(x_k)$  é a função probabilidade de  $X$ .

Analogamente, se  $X$  for uma variável aleatória contínua, temos:

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x) dx,$$

onde  $f_X(x)$  é a função densidade de probabilidade de  $X$ .

**Observação 7.2.2:** A função característica de uma variável aleatória contínua é a transformada de Fourier da densidade de probabilidade de  $X$ . ■

### 7.2.1 Propriedades

A seguir listamos algumas propriedades da função característica.

P1. A função característica é limitada por 1:  $|\varphi_X(t)| \leq 1, \forall t \in R$ .

**Prova:** Como pela desigualdade de Jensen,  $E^2 \cos(tx) \leq E \cos^2(tx)$  e  $E^2 \operatorname{sen}(tx) \leq E \operatorname{sen}^2(tx)$ , temos

$$|\varphi_X(t)| = \sqrt{E^2 \cos(tX) + E^2 \operatorname{sen}(tX)} \leq \sqrt{E(\cos^2(tX) + \operatorname{sen}^2(tX))} = E1 = 1.$$

■

P2. A função característica assume o valor 1 no ponto 0:  $\varphi_X(0) = 1$ .

**Prova:**  $\varphi_X(0) = Ee^{i0X} = E1 = 1$ . ■

P3.  $\overline{\varphi_X(t)} = \varphi_X(-t)$ , onde  $\bar{c}$  é o complexo conjugado de  $c$ . (Se  $c = x + iy$ , o seu complexo conjugado é  $\bar{c} = x - iy$ .)

**Prova:**  $\varphi_X(-t) = E \cos(-tX) + iE \operatorname{sen}(-tX) = E \cos(tX) - iE \operatorname{sen}(tX) = \overline{\varphi_X(t)}$ . ■

P4.  $\varphi_X$  é uniformemente contínua na reta.

**Prova:** Uma função  $\varphi$  é uniformemente contínua, se para todo  $\epsilon > 0$  existe  $\delta > 0$  tal que para todo  $t, s \in R$   $|\varphi(t) - \varphi(s)| < \epsilon$  quando  $|t - s| < \delta$ . Logo,

$$|\varphi(t) - \varphi(s)| = |E(e^{itx} - e^{isx})| \leq E|e^{isx}(e^{i(t-s)x} - 1)| = E|e^{i(t-s)x} - 1|.$$

Seja  $h(u) = |e^{iu} - 1|$ . Como  $0 \leq |e^{iu} - 1| \leq 2$ , 2 é integrável, e  $\lim_{u \rightarrow 0} h(u) = 0$ , pelo teorema da convergência dominada, temos que  $\lim_{u \rightarrow 0} Eh(u) = 0$ . Então, para todo  $\epsilon > 0$  existe  $\delta > 0$  tal que  $|u| < \delta$  implica que  $Eh(u) < \epsilon$ , ou seja, para todo  $\epsilon > 0$  existe  $\delta > 0$  tal que  $|t - s| < \delta$  implica que  $|\varphi(t) - \varphi(s)| \leq E|e^{i(t-s)x} - 1| < \epsilon$ . ■

P5. Se  $X$  e  $Y$  são independentes, então  $\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t), \forall t \in R$ .

**Prova:**  $\varphi_{X+Y}(t) = Ee^{it(X+Y)} = E(e^{itX}e^{itY}) = E(e^{itX})E(e^{itY}) = \varphi_X(t) \cdot \varphi_Y(t)$ . ■

É fácil provar por indução que se  $X_1, \dots, X_n$  são variáveis aleatórias independentes, então  $\varphi_{X_1+\dots+X_n}(t) = \prod_{k=1}^n \varphi_{X_k}(t), \forall t \in R$ .

P6. A variável aleatória  $X$  tem distribuição simétrica em torno de 0 se, e somente se,  $\varphi_X(t)$  é real para todo  $t \in R$ .

**Prova:**  $X$  é simétrica em torno de 0 se e somente se  $P(X \leq x) = P(X \geq -x), \forall x \in R$ . Como  $X \geq -x \Leftrightarrow -X \leq x$ , nós temos que  $F_X = F_{-X}$ , ou seja,  $\varphi_X = \varphi_{-X}$ . Como

$$\varphi_{-X}(t) = Ee^{it(-X)} = Ee^{i(-t)X} = \varphi_X(-t) = \overline{\varphi_X(t)}.$$

Então,  $X$  é simétrica em torno de 0 se e somente se  $\varphi_X(t) = \overline{\varphi_X(t)}$ , ou seja, se  $\varphi_X(t)$  é real para todo  $t \in R$ . ■

P7. Se  $E|X|^n < \infty$ , então  $\varphi_X^{(k)}(0) = i^k EX^k$  para  $k \in \{1, \dots, n\}$ , de modo que a função característica é uma espécie de função geradora de momentos.

**Prova:** Suponhamos que  $X$  seja integrável; queremos provar que  $\varphi_X'(t) = E(iXe^{itX})$ .

Note que para  $h \neq 0$ , temos  $\frac{\varphi_X(t+h) - \varphi_X(t)}{h} = E(e^{itX} \frac{e^{ihX} - 1}{h})$ . Como  $\frac{e^{ihx} - 1}{h} \rightarrow ix$  quando  $h \rightarrow 0$  (regra de L'Hopital),  $\forall x \in R$ , temos que o resultado decorre se pudermos trocar a ordem do limite e da esperança. Mas como para todo  $x$ ,

$$\left| \frac{e^{ihx} - 1}{h} \right| = \left| \frac{\int_0^h ix e^{isx} ds}{h} \right| = |x| \cdot \left| \frac{\int_0^h e^{isx} ds}{h} \right| \leq |x|.$$

Portanto, como  $|e^{itX}| = 1$ , temos

$$\left| e^{itX} \frac{e^{ihX} - 1}{h} \right| \leq |X|.$$

Como  $X$  é integrável, o Teorema da Convergência Dominada implica que

$$\begin{aligned}\varphi'_X(t) &= \lim_{h \rightarrow 0} \frac{\varphi_X(t+h) - \varphi_X(t)}{h} \\ &= \lim_{h \rightarrow 0} E\left(e^{itX} \frac{(e^{ihX} - 1)}{h}\right) = E\left(\lim_{h \rightarrow 0} e^{itX} \frac{(e^{ihX} - 1)}{h}\right) = E(iXe^{itX}).\end{aligned}$$

Logo,  $\varphi'_X(0) = iEX$ . O restante da prova segue por indução em  $n$ . ■

P8. Se  $Y = aX + b$ , onde  $a$  e  $b$  são números reais constantes,  $\varphi_Y(t) = e^{itb}\varphi_X(at)$ .

**Prova:**  $\varphi_Y(t) = Ee^{itY} = Ee^{it(aX+b)} = Ee^{itb}e^{itaX} = e^{itb}Ee^{i(at)X} = e^{itb}\varphi_X(at)$ . ■

P9.  $\varphi_X(t)$  é positiva definida. Isto é, para todo  $n = 1, 2, \dots$ , tem-se

$$\sum_{j=1}^n \sum_{k=1}^n \varphi_X(t_j - t_k) z_j \bar{z}_k \geq 0,$$

para quaisquer números reais  $t_1, t_2, \dots, t_n$  e complexos  $z_1, z_2, \dots, z_n$ .

**Prova:**

$$\begin{aligned}& \sum_{j=1}^n \sum_{k=1}^n \varphi_X(t_j - t_k) z_j \bar{z}_k \\ &= \sum_{j=1}^n \sum_{k=1}^n E(e^{iX(t_j - t_k)}) z_j \bar{z}_k \\ &= \sum_{j=1}^n \sum_{k=1}^n E(z_j e^{iX(t_j)} \bar{z}_k e^{-iX t_k}) \\ &= E\left(\sum_{j=1}^n \sum_{k=1}^n z_j e^{iX(t_j)} \overline{z_k e^{iX t_k}}\right) \\ &= E\left[\left(\sum_{j=1}^n z_j e^{iX(t_j)}\right) \overline{\left(\sum_{k=1}^n z_k e^{iX t_k}\right)}\right] \\ &= E\left[\left(\sum_{j=1}^n z_j e^{iX(t_j)}\right) \overline{\left(\sum_{k=1}^n z_k e^{iX t_k}\right)}\right] \\ &= E\left(\left|\sum_{j=1}^n z_j e^{iX(t_j)}\right|^2\right) \geq 0\end{aligned}$$

Portanto,  $\varphi_X$  é positiva definida. ■

Os resultados a seguir conhecidos como *Fórmula de Inversão* e *Teorema da Unicidade* garantem que a função característica determina a função de distribuição de uma variável aleatória.



**Teorema 7.2.3:** *Seja  $X$  uma variável aleatória qualquer, então sua função característica  $\varphi_X(t)$  determina a função de distribuição de  $X$ , através da seguinte Fórmula de Inversão:*

$$\tilde{F}(b) - \tilde{F}(a) = \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-iat} - e^{-ibt}}{it} \varphi_X(t) dt;$$

onde  $\tilde{F}(w) = \frac{1}{2}(F(w^+) + F(w^-))$ ,  $\forall w \in \mathbb{R}$  e  $a, b, c$  são números reais tais que  $c > 0$  e  $a < b$ .

**Prova:** Note que se  $F$  for contínua em  $w$ , então  $\tilde{F}(w) = F(w)$ . A função  $\frac{e^{-iat} - e^{-ibt}}{it}$  é definida para ser igual a  $b - a$ , quando  $t = 0$ , coincidindo com seu limite quando  $t \rightarrow 0$ . Logo, ela será contínua para todo  $t$  real e limitada, pois:

$$\begin{aligned} \left| \frac{e^{-iat} - e^{-ibt}}{it} \right| &= \left| e^{\frac{i(a+b)t}{2}} \right| \times \left| \frac{e^{-iat} - e^{-ibt}}{it} \right| \\ &= \left| \frac{e^{\frac{1}{2}i(b-a)t} - e^{\frac{1}{2}i(a-b)t}}{it} \right| = \left| \frac{2 \operatorname{sen}\left[\frac{(b-a)t}{2}\right]}{t} \right| \leq b - a, \end{aligned}$$

onde a última desigualdade decorre do fato que  $|\operatorname{sen} w| \leq w$ ,  $\forall w \in \mathbb{R}$ .

Denotando por  $Int(c)$  a integral da fórmula da inversão, temos

$$\begin{aligned} Int(c) &= \frac{1}{2\pi} \int_{-c}^c \frac{e^{-iat} - e^{-ibt}}{it} \varphi_X(t) dt \\ &= \frac{1}{2\pi} \int_{-c}^c \frac{e^{-iat} - e^{-ibt}}{it} E(e^{iXt}) dt \\ &= \frac{1}{2\pi} \int_{-c}^c E\left(\frac{e^{-i(a-X)t} - e^{-i(b-X)t}}{it}\right) dt \\ &= E\left[\frac{1}{2\pi} \int_{-c}^c \frac{e^{-i(a-X)t} - e^{-i(b-X)t}}{it} dt\right], \end{aligned}$$

onde a última igualdade decorre da troca da ordem de integração que é justificada tendo em vista que o integrando é limitado conforme provamos acima. Portanto, trabalhando o termo entre colchetes, temos

$$\begin{aligned} &\frac{1}{2\pi} \int_{-c}^c \frac{e^{-i(a-X)t} - e^{-i(b-X)t}}{it} dt \\ &= \frac{1}{2\pi} \int_{-c}^c \frac{1}{it} [\cos((X-a)t) + i \operatorname{sen}((X-a)t) - \cos((X-b)t) - i \operatorname{sen}((X-b)t)] dt \\ &= \frac{1}{\pi} \int_0^c \left( \frac{\operatorname{sen}((X-a)t) - \operatorname{sen}((X-b)t)}{t} \right) dt \\ &= \frac{1}{\pi} \int_0^c \frac{\operatorname{sen}((X-a)t)}{t} dt - \frac{1}{\pi} \int_0^c \frac{\operatorname{sen}((X-b)t)}{t} dt \\ &= \frac{1}{\pi} \int_0^{c(X-a)} \frac{\operatorname{sen}(u)}{u} du - \frac{1}{\pi} \int_0^{c(X-b)} \frac{\operatorname{sen}(u)}{u} du \\ &= g(c(X-a)) - g(c(X-b)), \end{aligned}$$

onde  $g$  é a função dada por  $g(w) = \frac{1}{\pi} \int_0^w \frac{\operatorname{sen}(u)}{u} du, w \in \mathbb{R}$ . Logo, temos

$$\operatorname{Int}(c) = E[g(c(X - a)) - g(c(X - b))].$$

Como vamos passar ao limite para  $c \rightarrow \infty$ , precisamos verificar se será possível trocar a ordem entre limite e esperança. Como  $g$  é contínua e  $\lim_{w \rightarrow \pm\infty} g(w) = \pm\frac{1}{2}$ , temos que  $g$  é limitada. Então a troca de ordem do limite e da esperança é justificada pelo Teorema da Convergência Dominada. Seja  $Y = \frac{1}{2}I_{a \leq X < b} + \frac{1}{2}I_{a < X \leq b}$ . Temos que

$$\lim_{c \rightarrow \infty} g(c(X - a)) - g(c(X - b)) = Y.$$

Então,

$$\lim_{c \rightarrow \infty} \operatorname{Int}(c) = E[\lim_{c \rightarrow \infty} g(c(X - a)) - g(c(X - b))] = EY.$$

Mas o valor esperado de  $Y$  é dado por:

$$\begin{aligned} EY &= \frac{1}{2}P(X = a) + \frac{1}{2}P(X = b) + P(a < X < b) \\ &= \frac{1}{2}(F(a) - F(a^-)) + \frac{1}{2}(F(b) - F(b^-)) + (F(b^-) - F(a)) \\ &= \frac{1}{2}(F(b) + F(b^-)) - \frac{1}{2}(F(a) + F(a^-)) = \tilde{F}(b) - \tilde{F}(a). \end{aligned}$$

Portanto,  $\lim_{c \rightarrow \infty} \operatorname{Int}(c) = \tilde{F}(b) - \tilde{F}(a)$ , como queríamos demonstrar. ■

Agora podemos utilizar a fórmula da inversão para provar o Teorema da Unicidade.

**Teorema 7.2.4: Teorema da Unicidade.** *Se as variáveis aleatórias  $X$  e  $Y$  têm a mesma função característica, então elas têm a mesma distribuição.*

**Prova:** Por hipótese,  $X$  e  $Y$  têm a mesma função característica e, como consequência da Fórmula da Inversão, temos que para quaisquer  $a, b$  reais e  $a < b$ ,

$$\tilde{F}_X(b) - \tilde{F}_X(a) = \tilde{F}_Y(b) - \tilde{F}_Y(a).$$

Tomando o limite quando  $a \rightarrow -\infty$ , temos que  $\tilde{F}_X(a) \rightarrow 0$  e  $\tilde{F}_Y(a) \rightarrow 0$ . Portanto,  $\tilde{F}_X(b) = \tilde{F}_Y(b), \forall b \in \mathbb{R}$ . Seja  $c < b$ , pela monotonicidade de  $F_X$  e  $F_Y$  e pela definição de  $\tilde{F}$ , temos

$$F_X(c) \leq \tilde{F}_X(b) \leq F_X(b) \text{ e } F_Y(c) \leq \tilde{F}_Y(b) \leq F_Y(b).$$

Então pela continuidade à direita da função de distribuição, temos que

$$\lim_{b \downarrow c} \tilde{F}_X(b) = F_X(c) \text{ e } \lim_{b \downarrow c} \tilde{F}_Y(b) = F_Y(c).$$

Logo,  $F_X(c) = F_Y(c), \forall c \in \mathbb{R}$  como queríamos demonstrar. ■

Note que o Teorema da Unicidade junto com a definição de função característica implicam que existe uma correspondência 1-1 entre funções características e funções de distribuições.

**Exemplo 7.2.5:** Se  $\varphi_X(t) = \frac{1}{1+t^2}$ , calcule  $\text{Var}X$ .

**Solução:** Diferenciando  $\varphi_X$ , temos  $\varphi'_X(t) = \frac{-2t}{(1+t^2)^2}$ . Diferenciando mais uma vez,  $\varphi''_X(t) = \frac{-2(1+t^2)^2 + 2t(2(1+t^2)2t)}{(1+t^2)^4}$ . Portanto,  $EX = \frac{\varphi'_X(0)}{i} = 0$  e  $EX^2 = \frac{\varphi''_X(0)}{i^2} = -(-2) = 2$ . Logo,  $\text{Var}X = EX^2 - (EX)^2 = 2$ .

**Exemplo 7.2.6:** Seja  $\varphi(t) = \cos(at)$ , onde  $a > 0$ . Mostraremos que  $\varphi$  é função característica, achando a distribuição correspondente. Já que assume valores reais, se  $\varphi$  fosse função característica de alguma variável aleatória  $X$ , então por P6,  $X$  possuiria distribuição simétrica em torno de zero. Com efeito teríamos  $\cos(at) = \varphi(t) = E \cos(tX)$ , pois a parte imaginária seria nula. Como  $\cos(at) = \cos(-at)$ , é evidente que uma distribuição simétrica concentrada nos dois pontos  $a$  e  $-a$  corresponderia a função característica  $\varphi$ . Portanto,  $\varphi$  é função característica de  $X$ , se, e somente se,  $P(X = a) = 1/2 = P(X = -a)$ .

**Exemplo 7.2.7:** Sejam  $X_1$  e  $X_2$  duas variáveis aleatórias i.i.d. e seja  $Y = X_1 - X_2$ . Qual a função característica de  $Y$ ?

**Solução:** Seja  $\varphi$  a função característica de  $X_1$  e  $X_2$ . Por P8 e P3, temos que  $\varphi_{-X_2}(t) = \varphi(-t) = \overline{\varphi(t)}$ . Então, como  $X_1$  e  $X_2$  são independentes, por P5, temos que

$$\varphi_Y(t) = \varphi(t)\varphi_{-X_2}(t) = |\varphi(t)|^2.$$

**Teorema 7.2.8:** Uma função contínua  $\psi : \mathbb{R} \rightarrow \mathbb{C}$  com  $\psi(0) = 1$  é função característica de alguma variável aleatória se, e somente se, ela for positiva definida.

**Prova:** Conforme propriedades já demonstradas, se for função característica, é contínua, positiva definida e aplicada em 0, resulta o valor 1. A prova da recíproca será omitida. ■

## 7.2.2 Exemplos de Funções Características

**Bernoulli.** Suponhamos que  $X \sim \text{Bernoulli}(p)$ , onde  $P(X = 1) = p = 1 - P(X = 0)$ . Então,

$$\varphi_X(t) = Ee^{itX} = pe^{it} + (1 - p).$$

**Poisson.** Suponhamos que  $X \sim \text{Poisson}(\lambda)$ . Então,

$$\varphi_X(t) = Ee^{itX} = \sum_{n=0}^{\infty} e^{itn} e^{-\lambda} \frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^{it})^n}{n!} = e^{\lambda(e^{it}-1)}.$$

**Uniforme.** Suponhamos que  $X \sim \text{Uniforme}(-a, a)$ . Então,  $f_X(x) = \frac{1}{2a}$  para  $-a < x < a$ , e  $f_X(x) = 0$  caso contrário. Logo, se  $t = 0$ , então  $\varphi_X(0) = 1$ , e para  $t \neq 0$ ,

$$\varphi_X(t) = Ee^{itX} = \int_{-a}^a \frac{e^{itx}}{2a} dx = \frac{1}{2a} \left( \frac{e^{ita} - e^{-ita}}{it} \right) = \frac{\text{sen}(ta)}{ta}.$$

**Normal.** Suponhamos que  $X \sim N(0, 1)$ . Então,

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi}} \int e^{itx} e^{-\frac{x^2}{2}} dx = e^{-\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int e^{-\frac{(x-it)^2}{2}} dx = e^{-\frac{t^2}{2}},$$

onde esta última integral pode ser calculada utilizando o Teorema de Cauchy tendo em vista que  $e^{-\frac{z^2}{2}}$  é uma função analítica no plano complexo.

**Exponencial.** Suponhamos que  $X \sim Exp(\alpha)$ . Então,

$$\varphi_X(t) = \int_0^\infty e^{itx} \alpha e^{-\alpha x} dx = \alpha \int_0^\infty e^{x(-\alpha+it)} dx = \left[ \frac{\alpha}{-\alpha+it} e^{x(-\alpha+it)} \right]_0^\infty = \frac{\alpha}{\alpha-it}.$$

**Exemplo 7.2.9:** Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes e identicamente distribuídas, seguindo o modelo de Poisson com parâmetro  $\lambda$ . Queremos obter a distribuição de  $X_1 + X_2 + \dots + X_n$ .

**Solução:** Temos

$$\varphi_{X_1+\dots+X_n}(t) = E(e^{it(X_1+\dots+X_n)}) = \prod_{j=1}^n E(e^{itX_j}) = e^{n\lambda(e^{it}-1)}.$$

Portanto,  $X_1 + X_2 + \dots + X_n$  tem uma distribuição Poisson com parâmetro  $n\lambda$ .

## 7.3 Teorema da Continuidade de Levy

Nosso objetivo nesta seção é provar que  $X_n \rightarrow^D X$  se, e somente se,  $\varphi_{X_n}(t) \rightarrow \varphi_X(t), \forall t \in R$ . Antes de provarmos a necessidade desta afirmação, considere a seguinte definição de convergência de funções de distribuição.

**Definição 7.3.1:** Seja  $X, X_1, X_2, \dots$  uma seqüência de variáveis aleatórias com funções de distribuição acumuladas dadas respectivamente por  $F, F_1, F_2, \dots$ . Diz-se que  $F_n$  converge fracamente para  $F$ , se  $X_n \rightarrow^D X$ . ■

**Teorema 7.3.2: Teorema de Helly-Bray.** Sejam  $F, F_1, F_2, \dots$  funções de distribuição. Se  $F_n$  converge fracamente para  $F$ , então

$$\int g(x) dF_n(x) \rightarrow \int g(x) dF(x)$$

para toda função  $g : R \rightarrow R$  contínua e limitada.

**Prova:** Para  $-\infty < a < b < \infty$ , onde  $a$  e  $b$  são pontos de continuidade de  $F$ ,

$$\left| \int g dF_n - \int g dF \right| \leq \left| \int g dF_n - \int_a^b g dF_n \right| + \left| \int_a^b g dF_n - \int_a^b g dF \right| + \left| \int_a^b g dF - \int g dF \right| = I + II + III.$$

Seja  $c = \sup_{x \in R} |g(x)| < \infty$  e seja  $\epsilon > 0$ . Então,

$$\begin{aligned} III &= \left| \int_a^b g dF - \int g dF \right| = \left| \int_{-\infty}^a g dF + \int_b^{\infty} g dF \right| \leq \left| \int_{-\infty}^a g dF \right| + \left| \int_b^{\infty} g dF \right| \\ &\leq \int_{-\infty}^a |g| dF + \int_b^{\infty} |g| dF \leq \int_{-\infty}^a c dF + \int_b^{\infty} c dF = c(F(a) + 1 - F(b)) \end{aligned}$$

Logo, para qualquer  $\epsilon > 0$ , podemos escolher  $a$  suficientemente pequeno e  $b$  suficientemente grande tal que  $III < \epsilon$ , pois  $\lim_{x \rightarrow -\infty} F(x) = 0$  e  $\lim_{x \rightarrow \infty} F(x) = 1$ . Para esses valores de  $a$  e  $b$ , e para  $n$  suficientemente grande, como  $a$  e  $b$  são pontos de continuidade de  $F$ , e como  $F_n$  converge fracamente para  $F$ , temos que  $I \leq c(F_n(a) + 1 - F_n(b)) < 2\epsilon$ .

Consideremos agora  $II$ . Sejam  $a$  e  $b$  os pontos já escolhidos. Já que  $g$  é uniformemente contínua em  $[a, b]$ ,<sup>1</sup> podemos escolher  $x_0, x_1, \dots, x_N$  tais que  $a = x_0 < x_1 < \dots < x_N = b$ , onde  $x_i$  são pontos de continuidade de  $F$  e  $|g(x) - g(x_i)| < \epsilon$  para todo  $x \in [x_i, x_{i+1}]$ ,  $i \in \{0, \dots, N-1\}$ . Então,

$$m_{ni} = (g(x_i) - \epsilon)(F_n(x_{i+1}) - F_n(x_i)) \leq \int_{x_i}^{x_{i+1}} g(x) dF_n(x) \leq (g(x_i) + \epsilon)(F_n(x_{i+1}) - F_n(x_i)) = M_{ni}$$

e

$$m_i = (g(x_i) - \epsilon)(F(x_{i+1}) - F(x_i)) \leq \int_{x_i}^{x_{i+1}} g(x) dF(x) \leq (g(x_i) + \epsilon)(F(x_{i+1}) - F(x_i)) = M_i.$$

Portanto,

$$m_{ni} - M_i \leq \int_{x_i}^{x_{i+1}} g(x) dF_n(x) - \int_{x_i}^{x_{i+1}} g(x) dF(x) \leq M_{ni} - m_i,$$

para  $i \in \{0, \dots, N-1\}$ . Somando, temos

$$\sum_{i=0}^{N-1} (m_{ni} - M_i) \leq \int_a^b g(x) dF_n(x) - \int_a^b g(x) dF(x) \leq \sum_{i=0}^{N-1} (M_{ni} - m_i).$$

Quando  $n \rightarrow \infty$ , temos que  $m_{ni} \rightarrow m_i$  e  $M_{ni} \rightarrow M_i$ , logo,

$$\sum_{i=0}^{N-1} (m_{ni} - M_i) \rightarrow \sum_{i=0}^{N-1} (m_i - M_i) = -2\epsilon(F(b) - F(a)) \geq -2\epsilon$$

e

$$\sum_{i=0}^{N-1} (M_{ni} - m_i) \rightarrow \sum_{i=0}^{N-1} (M_i - m_i) = 2\epsilon(F(b) - F(a)) \leq 2\epsilon$$

Como para  $n$  suficientemente grande temos que  $|\sum_{i=0}^{N-1} (m_{ni} - M_i) - \sum_{i=0}^{N-1} (m_i - M_i)| < \epsilon$  e  $|\sum_{i=0}^{N-1} (M_{ni} - m_i) - \sum_{i=0}^{N-1} (M_i - m_i)| < \epsilon$ , segue que  $\sum_{i=0}^{N-1} (m_{ni} - M_i) \geq -3\epsilon$  e  $\sum_{i=0}^{N-1} (M_{ni} - m_i) \leq 3\epsilon$ .

<sup>1</sup>Uma função  $g$  é uniformemente contínua em  $[a, b]$  se para todo  $\epsilon > 0$ , existe  $\delta > 0$  tal que para todo  $x, y \in [a, b]$  se  $|x - y| < \delta$ , então  $|g(x) - g(y)| < \epsilon$ . É fácil provar que toda função contínua em um intervalo fechado é uniformemente contínua neste intervalo.

$m_i) \leq 3\epsilon$ . Então, para  $n$  suficientemente grande, temos que  $II \leq 3\epsilon$ . Portanto,  $|\int gdF_n - \int gdF| \leq 6\epsilon$  para  $n$  grande o suficiente. ■

Como  $\cos(tx)$  e  $\sin(tx)$  são funções contínuas e limitadas, tem-se que para  $t$  fixo

$$E(\cos(tX_n)) \rightarrow E(\cos(tX))$$

e

$$E(\sin(tX_n)) \rightarrow E(\sin(tX))$$

Logo,  $\varphi_{X_n}(t) \rightarrow \varphi_X(t)$ .

É fácil definir a função característica  $\varphi$  dada uma função de distribuição  $F$ :  $\varphi(t) = \int e^{itx} dF(x), \forall t \in R$ . O próximo teorema implica a suficiência do nosso objetivo nesta seção, ou seja, se  $\varphi_{X_n} \rightarrow \varphi_X$ , então  $X_n \rightarrow^D X$ .

**Teorema 7.3.3:** *Sejam  $F_1, F_2, \dots$  funções de distribuições e  $\varphi_1, \varphi_2, \dots$  suas funções características. Se  $\varphi_n$  converge pontualmente para um limite  $\varphi$  e se  $\varphi$  é contínua no ponto zero, então*

- (a) *existe uma função de distribuição  $F$  tal que  $F_n \rightarrow F$  fracamente; e*
- (b)  *$\varphi$  é a função característica de  $F$ .*

**Prova:** Note que o teorema anterior implica que, sob as hipóteses, (a) implica (b). Para provar que  $F_n$  converge fracamente para alguma função de distribuição, vamos primeiro provar que para toda seqüência de funções de distribuição satisfazendo as condições do teorema, existem uma subseqüência  $F_{n_1}, F_{n_2}, \dots$  e uma função de distribuição  $F$  tais que  $F_{n_j} \rightarrow F$  fracamente, quando  $j \rightarrow \infty$ . Provaremos isso em duas etapas:

- (i) existem uma subseqüência  $F_{n_1}, F_{n_2}, \dots$  e uma função  $F : R \rightarrow [0, 1]$  tais que  $F$  é não-decrescente e contínua à direita e  $F_{n_j}(x) \rightarrow F(x)$ , quando  $j \rightarrow \infty$ , para todo  $x$  ponto de continuidade de  $F$ ; e
- (ii)  $F(\infty) = 1$  e  $F(-\infty) = 0$ .

Para provar (i), usaremos o método da diagonalização. Sejam  $r_1, r_2, \dots$ , uma enumeração dos racionais da reta. Considere a seguinte matriz:

$$\begin{array}{cccccc} F_1 & F_2 & F_3 & F_4 & \dots \\ F_1^1 & F_2^1 & F_3^1 & F_4^1 & \dots \\ F_1^2 & F_2^2 & F_3^2 & F_4^2 & \dots \\ F_1^3 & F_2^3 & F_3^3 & F_4^3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

Nesta matriz temos que a seqüência  $(F_1^j, F_2^j, F_3^j, \dots)$  contida na  $(j + 1)$ -ésima linha da matriz é uma subseqüência da seqüência contida na  $j$ -ésima linha que converge no racional  $r_j$ , para  $j \geq 1$ . Note que como a seqüência  $(F_1^{j-1}(r_j), F_2^{j-1}(r_j), F_3^{j-1}(r_j), \dots)$  é uma seqüência limitada de números reais, ela possui uma subseqüência convergente; logo pode-se escolher a

seqüência  $(F_1^j, F_2^j, F_3^j, \dots)$  indutivamente conforme descrito acima. Seja  $F_{n_j} = F_j^j$ , para  $j \geq 1$ , então temos que a subseqüência  $(F_{n_j})_j$  converge em todos os racionais da reta. Chamemos o limite de  $F(r_k)$ , de modo que  $F_{n_j}(r_k) \rightarrow F(r_k), \forall k$ . É óbvio que  $0 \leq F(r_k) \leq 1$  e que  $F$  é não decrescente nos racionais. Definamos  $F$  em  $x$  irracional por  $F(x) = \lim_{r \downarrow x, r \text{ racional}} F(r)$ .  $F$  assim definida é não-decrescente, mas não é necessariamente contínua à direita. Vamos provar que  $F_{n_j}(x) \rightarrow F(x)$  para todo ponto  $x$  de continuidade de  $F$ . Suponha que  $x$  é um ponto de continuidade de  $F$  e sejam  $r'$  e  $r''$  racionais tais que  $r' < x < r''$  e  $F(r'') - \epsilon < F(x) < F(r') + \epsilon$ . Então,

$$\begin{aligned} F(x) - \epsilon < F(r') &= \lim_{j \rightarrow \infty} F_{n_j}(r') \leq \liminf_{j \rightarrow \infty} F_{n_j}(x) \leq \\ \limsup_{j \rightarrow \infty} F_{n_j}(x) &\leq \lim_{j \rightarrow \infty} F_{n_j}(r'') = F(r'') < F(x) + \epsilon \end{aligned}$$

Como  $\epsilon$  é arbitrário, temos  $F_{n_j}(x) \rightarrow F(x)$  quando  $j \rightarrow \infty$ . Finalmente, podemos redefinir  $F$  nos seus pontos de descontinuidade de modo que  $F$  seja contínua à direita.

Para provar (ii), note que

$$\int_0^t \varphi_{n_j}(s) ds = \int_0^t \int_{-\infty}^{\infty} e^{isx} dF_{n_j}(x) ds.$$

Mas como o integrando é limitado podemos trocar a ordem de integração, logo

$$\int_0^t \varphi_{n_j}(s) ds = \int_{-\infty}^{\infty} \int_0^t e^{isx} ds dF_{n_j}(x) = \int_{-\infty}^{\infty} \frac{e^{itx} - 1}{ix} dF_{n_j}(x)$$

Considere a função,  $h(x) = \frac{e^{itx} - 1}{ix}$  para  $x \neq 0$  e  $h(0) = t$ .  $h$  é limitada e contínua e um argumento similar ao utilizado na prova do teorema anterior, pode ser utilizado para provar que quando  $j \rightarrow \infty$

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{e^{itx} - 1}{ix} dF_{n_j}(x) &= \int_{-\infty}^{\infty} h(x) dF_{n_j}(x) \rightarrow \int_{-\infty}^{\infty} h(x) dF(x) = \\ \int_{-\infty}^{\infty} \frac{e^{itx} - 1}{ix} dF(x) &= \int_0^t \int_{-\infty}^{\infty} e^{isx} dF(x) ds \end{aligned}$$

Como  $\varphi_{n_j}(t) \rightarrow \varphi(t)$ ,  $\varphi$  é contínua em zero, implica que  $\varphi$  é limitada e mensurável, então pelo teorema da convergência dominada, tem-se que

$$\int_0^t \varphi_{n_j}(s) ds \rightarrow \int_0^t \varphi(s) ds.$$

Igualando-se os limites iguais e dividindo-se por  $t$ , temos

$$\frac{1}{t} \int_0^t \varphi(s) ds = \frac{1}{t} \int_0^t \int_{-\infty}^{\infty} e^{isx} dF(x) ds, t \neq 0.$$

Fazendo  $t \rightarrow 0$  e usando a continuidade em  $s = 0$  das duas funções  $\varphi(s)$  e  $\int_{-\infty}^{\infty} e^{isx} dF(x)$ , tem-se

$$\varphi(0) = \int_{-\infty}^{\infty} 1 dF(x) = F(\infty) - F(-\infty).$$

Como  $\varphi(0) = \lim_{n \rightarrow \infty} \varphi_n(0) = 1$ , temos que  $F(\infty) - F(-\infty) = 1$ , ou seja, o que implica que  $F(\infty) = 1$  e  $F(-\infty) = 0$ .

Para terminar a prova suponha por contradição que  $F_n$  não convirja fracamente para  $F$ , onde  $F_{n_j} \rightarrow F$  fracamente. Então, existirão  $x$ , ponto de continuidade de  $F$  e uma subsequência  $F_{1'}, F_{2'}, \dots$  tais que  $F_{n'}(x) \rightarrow a \neq F(x)$ . Como essa subsequência também satisfaz as condições do teorema, (i) e (ii) implicam que existe uma subsequência  $F_{1''}, F_{2''}, \dots$  e uma função de distribuição  $G$  tais que  $F_{n''} \rightarrow G$  fracamente. Como  $F$  e  $G$  possuem a mesma função característica ( $\varphi$ ), temos que  $F = G$ , ou seja  $F_{n''}(x) \rightarrow a = G(x) = F(x)$ , uma contradição. ■

**Exemplo 7.3.4:** Suponha que  $X_n$  e  $Y_n$  são independentes para cada  $n \geq 0$  e que  $X_n \rightarrow^D X_0$  e  $Y_n \rightarrow^D Y_0$ . Prove que  $X_n + Y_n \rightarrow^D X_0 + Y_0$ .

**Solução:** Pelo Teorema da Continuidade sabemos que  $\varphi_{X_n}(t) \rightarrow \varphi_{X_0}(t)$  e que  $\varphi_{Y_n}(t) \rightarrow \varphi_{Y_0}(t)$ . Como  $X_n$  e  $Y_n$  são independentes temos que  $\varphi_{X_n+Y_n}(t) = \varphi_{X_n}(t)\varphi_{Y_n}(t)$ . Portanto,

$$\lim_n \varphi_{X_n+Y_n}(t) = \lim_n (\varphi_{X_n}(t)\varphi_{Y_n}(t)) = \varphi_{X_0}(t)\varphi_{Y_0}(t) = \varphi_{X_0+Y_0}(t).$$

Logo, pelo Teorema da Continuidade, temos que  $X_n + Y_n \rightarrow^D X_0 + Y_0$ .

**Exemplo 7.3.5:** Suponha que a variável aleatória  $X_n$  tenha distribuição Binomial, ou seja,

$$P(X_n = k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k}, k = 0, 1, 2, \dots, n.$$

Se  $p_n \rightarrow 0$  quando  $n \rightarrow \infty$  de tal modo que  $np_n \rightarrow \lambda > 0$ , então

$$X_n \rightarrow^D Y,$$

onde  $Y \sim Poisson(\lambda)$ . Para verificar isto lembre que podemos representar uma variável aleatória Binomial como a soma de variáveis aleatórias Bernoulli i.i.d., então

$$\varphi_{X_n}(t) = Ee^{itX_n} = (1 - p_n + e^{it}p_n)^n = (1 + p_n(e^{it} - 1))^n = \left(1 + \frac{np_n(e^{it} - 1)}{n}\right)^n \rightarrow e^{\lambda(e^{it} - 1)},$$

onde a expressão final é a função característica de uma variável aleatória  $Poisson(\lambda)$ . Portanto, pelo Teorema da Continuidade,  $X_n \rightarrow^D Y$ .

## 7.4 Soma de um Número Aleatório de Variáveis Aleatórias

Nesta seção, nós estudaremos somas de um número aleatório de variáveis aleatórias, ou seja,

$$S = \sum_{i=0}^N X_i,$$



onde  $N$  é uma variável aleatória inteira e não negativa, e assume-se que ela é independente das parcelas  $X_i$ . Por exemplo,  $N$  pode ser o número de clientes, pacotes ou trabalhos chegando em uma fila em um dado intervalo de tempo e  $X_i$  pode ser o tempo necessário para finalizar o  $i$ -ésimo trabalho.  $S$  então seria o tempo total do serviço. Em nossas aplicações assumiremos que  $N = 0$  significa que  $S = 0$ , ou seja,  $X_0 = 0$  com função característica  $\varphi_{X_0}(u) = 1$ . Sabemos que  $ES = E[E(S|N)]$  e que

$$E(S|N = n) = \sum_{i=0}^n E(X_i|N = n).$$

Como assumimos que  $N$  é independente de  $X_i$ , temos

$$E(S|N = n) = \sum_{i=0}^n EX_i.$$

Se as variáveis aleatórias  $\{X_i, i > 0\}$  têm esperança igual a  $m$ , então  $E(S|N = n) = nm$  e  $ES = mEN$ .

Para informações mais detalhadas sobre  $S$ , vamos calcular sua função característica  $\varphi_S$  assumindo que as variáveis aleatórias  $\{N, X_1, X_2, \dots\}$  são independentes:

$$\varphi_S(t) = Ee^{itS} = E(E(e^{itS}|N)).$$

Por outro lado, utilizando a hipótese de independência, podemos calcular,

$$E(e^{itS}|N = n) = E\left(\prod_{i=0}^n e^{itX_i}|N = n\right) = \prod_{i=0}^n \varphi_{X_i}(t).$$

Logo,

$$\varphi_S(t) = \sum_{n=0}^{\infty} P(N = n) \prod_{i=0}^n \varphi_{X_i}(t).$$

Se as parcelas  $\{X_1, X_2, \dots\}$  forem também identicamente distribuídas com função característica  $\varphi_X$ , então

$$\varphi_S(t) = \sum_{n=0}^{\infty} P(N = n) \varphi_X^n(t),$$

onde utilizamos o fato que  $\varphi_X^0 = 1 = \varphi_{X_0}(t)$ . Note que a função característica de  $N$  é:

$$\varphi_N(t) = \sum_{n=0}^{\infty} P(N = n) e^{itn} = \sum_{n=0}^{\infty} P(N = n) [e^{it}]^n.$$

Comparando as expressões de  $\varphi_S$  e  $\varphi_N$ , nós vemos que escolhendo  $t$  em  $\varphi_N(t)$  de forma que  $e^{it} = \varphi_X$ , nós podemos reescrever:

$$\varphi_S(t) = \varphi_N(-i \log \varphi_X(t)).$$

Portanto, nós provamos o seguinte teorema:

**Teorema 7.4.1:** Se  $N$  é uma variável aleatória inteira não-negativa,  $S = \sum_{i=0}^N X_i$ ,  $X_0 = 0$ , onde  $\{X_i, i \geq 1\}$  são i.i.d. com função característica comum  $\varphi_X$ , e elas são independentes de  $N$  que é descrita pela função característica  $\varphi_N$ , então

$$\varphi_S(t) = \varphi_N(-i \log \varphi_X(t)).$$

**Exemplo 7.4.2:** Suponha que  $N \sim \text{Poisson}(\lambda)$  representa o número de clientes que são atendidos em um dado tempo  $T$ . Suponha ainda que com probabilidade  $p$  o  $i$ -ésimo cliente fica satisfeito com o atendimento. Assuma que os clientes ficam satisfeitos com o serviço de maneira independente e que  $N$ , é independente da probabilidade que clientes ficam satisfeitos. Determine a distribuição de probabilidade de  $S$  o número total de clientes satisfeitos no tempo  $T$ .

**Solução:** Seja  $X_i \sim \text{Bernoulli}(p)$ ,  $i \geq 1$ , a variável aleatória que descreve se o  $i$ -ésimo cliente ficou ou não satisfeito com o atendimento. Então temos,

$$S = \sum_{i=0}^N X_i,$$

onde  $X_0 = 0$ . Desta forma, sabemos que

$$\varphi_S(t) = \varphi_N(-i \log \varphi_X(t)),$$

onde  $\varphi_X(t) = pe^{it} + (1-p)$  e  $\varphi_N(t) = e^{\lambda(e^{it}-1)}$ . Substituindo temos:

$$\varphi_S(t) = e^{\lambda(e^{i(-i \log(pe^{it} + (1-p)))-1})} = e^{\lambda(pe^{it} + (1-p) - 1)} = e^{p\lambda(e^{it}-1)}.$$

Pela unicidade da função característica, temos que  $S \sim \text{Poisson}(p\lambda)$ .

## 7.5 Função Característica de um Vetor Aleatório

**Definição 7.5.1:** Seja  $\vec{X} = (X_1, \dots, X_k)$  um vetor aleatório  $k$ -dimensional. A função característica de  $\vec{X}$  é a função  $\varphi_{\vec{X}} : \mathbb{R}^k \rightarrow \mathbb{C}$  definida por

$$\varphi_{\vec{X}}(\vec{t}) = Ee^{i\vec{t} \cdot \vec{X}} = E \exp\left(i \sum_{j=1}^k t_j X_j\right).$$

$\varphi_{\vec{X}}$  é também chamada de função característica conjunta de  $X_1, \dots, X_k$ .

A função característica multivariada tem propriedades análogas a todas as propriedades enunciadas para a função característica de uma variável aleatória. As propriedades P1–P4 e P6 são válidas com as óbvias modificações (a reta é substituída por  $\mathbb{R}^k$ ). Para P5, supõe-se que  $\vec{X}$  e  $\vec{Y}$  sejam vetores de mesma dimensão. Sob esta condição, a independência de  $\vec{X}$  e  $\vec{Y}$  implica que

$$\varphi_{\vec{X}+\vec{Y}}(\vec{t}) = \varphi_{\vec{X}}(\vec{t})\varphi_{\vec{Y}}(\vec{t}).$$

Quanto ao Teorema da Unicidade, também existe uma fórmula da inversão para a função característica multidimensional que pode ser usada para provar a unicidade da função característica:

**Teorema 7.5.2: Teorema da Unicidade.** *Se  $\vec{X}$  e  $\vec{Y}$  forem vetores aleatórios  $k$ -dimensionais tais que  $\varphi_{\vec{X}}(\vec{t}) = \varphi_{\vec{Y}}(\vec{t}), \forall \vec{t} \in \mathbb{R}^k$ , então  $\vec{X}$  e  $\vec{Y}$  têm a mesma distribuição. Em outras palavras, a função característica determina a distribuição, e podemos escrever:  $\varphi_{\vec{X}} = \varphi_{\vec{Y}} \Leftrightarrow F_{\vec{X}} = F_{\vec{Y}}$ .*

Analogamente a P7, correlações de ordem maiores podem ser facilmente calculadas diferenciando-se a função característica conjunta repetidamente. Formalmente, seja  $p = \sum_{k=1}^n p_k$  para números naturais quaisquer  $p_k$ , temos

$$E\left(\prod_1^n X_k^{p_k}\right) = \frac{1}{i^p} \frac{\partial^p \varphi_{\vec{X}}(\vec{t})}{\partial t_1^{p_1} \dots \partial t_n^{p_n}} \Big|_{\vec{t}=\vec{0}}.$$

No caso particular de  $\vec{X} = (X_1, X_2)$ , temos que

$$EX_1X_2 = -\frac{\partial^2 \varphi_{X_1, X_2}(t_1, t_2)}{\partial t_1 \partial t_2} \Big|_{t_1=t_2=0}.$$

Também é fácil analisar o comportamento da função característica multivariada de transformações lineares de vetores aleatórios em analogia a propriedade P8. (Assumiremos que um vetor  $\vec{X}$   $k$ -dimensional é uma matriz coluna com dimensão  $k \times 1$ . Deste modo  $\vec{t} \cdot \vec{X} = (\vec{t})^T \vec{X}$ .) Por exemplo, seja  $\vec{Y} = \mathbf{A}\vec{X} + \vec{b}$ , então

$$\begin{aligned} \varphi_{\vec{Y}}(\vec{t}) &= Ee^{i(\vec{t})^T \vec{Y}} = Ee^{i(\vec{t})^T (\mathbf{A}\vec{X} + \vec{b})} \\ &= E(e^{i(\vec{t})^T \vec{b}} e^{i(\mathbf{A}^T \vec{t})^T \vec{X}}) = e^{i(\vec{t})^T \vec{b}} \varphi_{\vec{X}}(\mathbf{A}^T \vec{t}), \end{aligned}$$

onde utilizamos o fato que  $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$  e que  $e^{i(\vec{t})^T \vec{b}}$  não é aleatório e pode sair fora da operação de esperança.

Assim como é fácil obter a distribuição marginal dada uma distribuição conjunta de variáveis aleatórias, também é fácil obter a função característica de qualquer distribuição marginal. Para isso basta fazer todos os termos “extras” iguais a zero na função característica multivariada. Por exemplo, para as variáveis aleatórias  $X, Y$ , e  $Z$ , temos  $Ee^{i(xX+yY)} = Ee^{i(xX+yY+0Z)}$ , ou seja,  $\varphi_{X,Y}(x,y) = \varphi_{X,Y,Z}(x,y,0), \forall (x,y) \in \mathbb{R}^2$ .

Como no caso unidimensional, temos convergência em distribuição se, e somente se, as funções características convergem.

**Teorema 7.5.3:**  $\vec{X}_n \rightarrow^D \vec{X}$  se, e somente se,  $\varphi_{\vec{X}_n}(\vec{t}) \rightarrow \varphi_{\vec{X}}(\vec{t}), \forall \vec{t} \in \mathbb{R}^k$ .

**Prova:** Omitida. ■

O próximo teorema mostra que convergência em distribuição de vetores aleatórios é equivalente à convergência em distribuição de todas as combinações lineares das coordenadas.

**Teorema 7.5.4: Cramér-Wold.** *Sejam  $\vec{X}_n = (X_{n1}, X_{n2}, \dots, X_{nk})$  e  $\vec{X} = (X_1, \dots, X_n)$  vetores aleatórios  $k$ -dimensionais. Então,  $\vec{X}_n \rightarrow \vec{X}$  se, e somente se,  $\sum_{j=1}^k t_j X_{nj} \rightarrow^D \sum_{j=1}^k t_j X_j$ , para todo  $(t_1, \dots, t_k) \in \mathbb{R}^k$ .*

**Prova:** Suponhamos primeiro que  $\sum_{j=1}^k t_j X_{nj} \rightarrow^D \sum_{j=1}^k t_j X_j, \forall (t_1, \dots, t_k)$ . Então,

$$\begin{aligned} \varphi_{\vec{X}_n}(t_1, \dots, t_k) &= E e^{i \sum_{j=1}^k t_j X_{nj}} \\ &= \varphi_{\sum_{j=1}^k t_j X_{nj}}(1) \rightarrow \varphi_{\sum_{j=1}^k t_j X_j}(1) = \varphi_{\vec{X}}(t_1, \dots, t_k), \end{aligned}$$

onde utilizamos o Teorema da Continuidade de Levy. Também pelo Teorema da Continuidade de Levy no caso multidimensional, temos que como  $\varphi_{\vec{X}_n} \rightarrow \varphi_{\vec{X}}, \vec{X}_n \rightarrow^D \vec{X}$ .

Agora suponha que  $\vec{X}_n \rightarrow^D \vec{X}$ . Para  $(t_1, \dots, t_k) \in \mathbb{R}^k$ , queremos provar que  $\sum_{j=1}^k t_j X_{nj} \rightarrow^D \sum_{j=1}^k t_j X_j$ . Para tanto, basta provarmos que  $\varphi_{\sum_{j=1}^k t_j X_{nj}}(t) \rightarrow \varphi_{\sum_{j=1}^k t_j X_j}(t), \forall t \in \mathbb{R}$ . Mas, utilizando novamente o Teorema da Continuidade de Levy, temos que

$$\begin{aligned} \varphi_{\sum_{j=1}^k t_j X_{nj}}(t) &= E e^{it \sum_{j=1}^k t_j X_{nj}} = E e^{i \sum_{j=1}^k (tt_j) X_{nj}} \\ \varphi_{\vec{X}_n}(tt_1, \dots, tt_k) &\rightarrow \varphi_{\vec{X}}(tt_1, \dots, tt_k) = \varphi_{\sum_{j=1}^k t_j X_j}(t) \end{aligned}$$

■

Terminaremos nossa discussão de funções características multidimensionais considerando um critério para independência de vetores aleatórios.

**Teorema 7.5.5:** *Sejam  $\vec{X} = (X_1, \dots, X_m)$  e  $\vec{Y} = (Y_1, \dots, Y_n)$  vetores aleatórios, onde  $m \geq 1, n \geq 1$ .  $\vec{X}$  e  $\vec{Y}$  são independentes se, e somente se,*

$$\varphi_{X_1, \dots, X_m, Y_1, \dots, Y_n}(x_1, \dots, x_m, y_1, \dots, y_n) = \varphi_{\vec{X}}(x_1, \dots, x_m) \varphi_{\vec{Y}}(y_1, \dots, y_n),$$

para todo  $(x_1, \dots, x_m) \in \mathbb{R}^m$  e  $(y_1, \dots, y_n) \in \mathbb{R}^n$ .

**Prova:** Suponhamos primeiro que  $\vec{X}$  e  $\vec{Y}$  sejam variáveis aleatórias  $X$  e  $Y$  ( $m = 1, n = 1$ ), com  $X$  e  $Y$  independentes. Então temos,

$$\varphi_{X,Y}(x, y) = E e^{i(xX+yY)} = E e^{ixX} e^{iyY} = E e^{ixX} E e^{iyY} = \varphi_X(x) \varphi_Y(y), \forall (x, y) \in \mathbb{R}^2.$$

Reciprocamente, suponha que  $\varphi_{X,Y}(x, y) = \varphi_X(x) \varphi_Y(y)$  para todo  $(x, y) \in \mathbb{R}^2$ . Então a independência de  $X$  e  $Y$  é consequência do Teorema da Unicidade: se  $X$  e  $Y$  fossem independentes, elas teriam função característica conjunta  $\varphi_{X,Y}(x, y) = \varphi_X(x) \varphi_Y(y)$  pela parte inicial desta demonstração. Se não fossem independentes, elas teriam uma função característica diferente, o que contraria a hipótese. Logo, são independentes.

A prova no caso geral é análoga e omitida. ■

Um resultado semelhante vale para um número finito qualquer de vetores aleatórios. Consideremos o caso mais simples em que  $X_1, \dots, X_n$  são variáveis aleatórias. Então, temos  $X_1, \dots, X_n$  independentes se, e somente se,

$$\varphi_{X_1, \dots, X_n}(t_1, \dots, t_n) = \prod_{j=1}^n \varphi_{X_j}(t_j), \forall (t_1, \dots, t_n) \in \mathbb{R}^n.$$

## 7.6 Funções Geratrizes de Momento

**Definição 7.6.1:** Uma função geratriz de momento  $\hat{F}_X(t)$  de uma variável aleatória  $X$  com função de distribuição  $F_X$  existe se,

$$\hat{F}_X(t) := Ee^{tX} < \infty, \forall t \in I,$$

onde  $I$  é um intervalo contendo 0 no seu interior.

O problema de utilizar funções geratrizes de momento é que elas nem sempre existem. Por exemplo, a função geratriz de momento de uma variável aleatória com distribuição de Cauchy não existe. Pode-se provar que a existência da função geratriz de momento é equivalente a cauda da distribuição de  $X$  ser limitada exponencialmente, ou seja,  $P(|X| > x) \leq Ke^{-cx}$ , para algum  $K > 0$  e  $c > 0$ . Se a função geratriz de momento existe, pode-se provar que ela também determina a função de distribuição.

## 7.7 Teorema de Slutsky

Nesta seção, estudaremos o Teorema de Slutsky que trata do comportamento da soma e do produto de variáveis aleatórias, uma convergindo em distribuição e outra em probabilidade. Antes disso, iremos provar que funções contínuas preservam convergência.

**Teorema 7.7.1:** *Sejam  $\{X_n : n \geq 1\}$  e  $X$  variáveis aleatórias com funções de distribuição  $\{F_n : n \geq 1\}$  e  $F$ , respectivamente. Seja  $g : \mathbb{R} \rightarrow \mathbb{R}$  uma função contínua. Então, se  $X_n$  converge para  $X$  quase certamente, em probabilidade ou em distribuição, o mesmo ocorre com  $g(X_n)$  para  $g(X)$ , no mesmo modo de convergência.*

**Prova:** Suponha que  $X_n \rightarrow X$  cp1. Então, existe um conjunto  $A \in \mathcal{F}$  tal que  $P(A) = 0$  e  $X_n(w) \rightarrow X(w)$  para  $w \in A^c$ . Como  $g$  é contínua,  $g(X_n(w)) \rightarrow g(X(w))$  para  $w \in A^c$  e, portanto,  $g(X_n) \rightarrow g(X)$  cp1.

Considere que  $X_n \rightarrow^P X$  e vamos verificar que  $g(X_n) \rightarrow^P g(X)$ . Dado  $\epsilon > 0$  arbitrário, fixemos  $m$  grande o suficiente tal que  $P(|X| > m/2) < \epsilon$ . A função  $g$  sendo contínua em  $\mathbb{R}$ , será uniformemente contínua no intervalo fechado  $[-m, m]$ , logo para  $\epsilon' > 0$  arbitrário existe  $\delta$  tal que  $0 < \delta \leq m/2$  e se  $x, y \in [-m, m]$  e  $|x - y| < \delta$ , então  $|g(x) - g(y)| < \epsilon'$ .

Observe que se  $P(A_n) \rightarrow 1$ , então  $P(A_n \cap A) \rightarrow P(A)$ , pois  $P(A_n) + P(A) - 1 \leq P(A_n \cap A) \leq P(A)$  e  $P(A_n) + P(A) - 1 \rightarrow P(A)$ . Portanto, como  $P(|X_n - X| < \delta) \rightarrow 1$ , temos que  $P(|X| \leq m/2, |X_n - X| < \delta) \rightarrow P(|X| \leq m/2) > 1 - \epsilon$ . Mas

$$[|X| \leq m/2, |X_n - X| < \delta] \subseteq [|X| \leq m, |X_n| \leq m, |X_n - X| < \delta] \subseteq [|g(X_n) - g(X)| < \epsilon'],$$

logo  $P(|g(X_n) - g(X)| < \epsilon') > 1 - 2\epsilon$  para  $n$  suficientemente grande. Como  $\epsilon$  é arbitrário, temos que  $P(|g(X_n) - g(X)| < \epsilon') \rightarrow 1$  quando  $n \rightarrow \infty$ , ou seja  $g(X_n) \rightarrow^P g(X)$ .

Finalmente, considere que  $X_n \rightarrow^D X$ . Pelo Teorema da Continuidade de Levy, para que  $g(X_n) \rightarrow^D g(X)$ , basta a convergência das respectivas funções características. Por definição,

$$\varphi_{g(X_n)}(t) = Ee^{itg(X_n)} = E \cos(tg(X_n)) + iE \operatorname{sen}(tg(X_n)).$$

Como as funções  $\cos(tg(x))$  e  $\sin(tg(x))$  são contínuas e limitadas na reta, para  $t$  fixo, decorre do Teorema de Helly-Bray que

$$\varphi_{g(X_n)}(t) \rightarrow E \cos(tg(X)) + iE \sin(tg(X)) = \varphi_{g(X)}(t), \forall t \in \mathbb{R}.$$

■

**Teorema 7.7.2:** Considere  $\{X_n : n \geq 1\}$ ,  $\{Y_n : n \geq 1\}$  e  $X$  variáveis aleatórias tais que valem as convergências  $X_n \rightarrow^D X$  e  $Y_n \rightarrow^P c$ , com  $c$  constante. Então,

(i)  $X_n + Y_n \rightarrow^D X + c$ ;

(ii)  $X_n Y_n \rightarrow^D cX$ ;

(iii) Se  $c \neq 0$ ,  $\frac{X_n}{Y_n} \rightarrow^D \frac{X}{c}$ , desde que  $P(Y_n \neq 0) = 1$ .

**Prova:** Prova de (i): Temos

$$\varphi_{X_n+Y_n}(t) = E(e^{it(X_n+Y_n)}) = E(e^{it(X_n+c)}) + E[(e^{itX_n})(e^{itY_n} - e^{itc})].$$

Por hipótese temos,

$$\lim_n E(e^{it(X_n+c)}) = \lim_n e^{itc} E(e^{itX_n}) = e^{itc} E(e^{itX}) = E(e^{it(X+c)}).$$

Observe que  $|e^{itX_n}| = 1$  e, assim, vem

$$|E[(e^{itX_n})(e^{itY_n} - e^{itc})]| \leq E[|(e^{itX_n})(e^{itY_n} - e^{itc})|] = E[|(e^{itY_n} - e^{itc})|].$$

Seja  $Z_n = |(e^{itY_n} - e^{itc})|$ , temos  $0 \leq Z_n \leq 2$ . Logo, para  $\epsilon > 0$ , temos

$$\begin{aligned} E[|(e^{itY_n} - e^{itc})|] &= EZ_n = E(Z_n I_{Z_n \leq \epsilon}) + E(Z_n I_{Z_n > \epsilon}) \\ &\leq \epsilon + 2E(I_{Z_n > \epsilon}) \leq \epsilon + 2P(Z_n > \epsilon). \end{aligned}$$

Como  $Z_n$  é uma função contínua de  $Y_n$  e lembrando que funções contínuas preservam convergência em probabilidade, temos que  $Z_n \rightarrow^P 0$ , pois  $Y_n \rightarrow^P c$ . Nessas condições, para  $n$  grande o suficiente,

$$|E[(e^{itX_n})(e^{itY_n} - e^{itc})]| \leq E[|(e^{itY_n} - e^{itc})|] < 2\epsilon.$$

Logo, tomando o limite de  $\varphi_{X_n+Y_n}(t)$  quando  $n \rightarrow \infty$ , concluímos a demonstração da parte (i).

Prova de (ii): Inicialmente consideramos  $c = 0$  e vamos verificar que  $X_n Y_n \rightarrow^P 0$ , e conseqüentemente,  $X_n Y_n \rightarrow^D 0$ . Sejam  $\epsilon, \delta > 0$  e  $x < 0 < y$  pontos de continuidade de  $F_X$  tais que  $F_X(y) - F_X(x) = P(x < X \leq y) > 1 - \delta$ . Como  $X_n \rightarrow^D X$ , temos  $P(x < X_n \leq y) = F_{X_n}(y) - F_{X_n}(x) > 1 - 2\delta$  para  $n$  suficientemente grande. Definamos  $M = \max(y, -x)$ , então a convergência em probabilidade de  $Y_n$  para zero implica que  $P(|Y_n| < \frac{\epsilon}{M}) > 1 - \delta$  para  $n$  suficientemente grande. Logo para  $n$  suficientemente grande, temos

$$P(x < X_n \leq y, |Y_n| < \frac{\epsilon}{M}) > 1 - 3\delta.$$

Como  $x < X_n \leq y$  e  $|Y_n| < \frac{\epsilon}{M}$  implicam  $|X_n Y_n| < \epsilon$ , temos  $P(|X_n Y_n| < \epsilon) > 1 - 3\delta$  para  $n$  grande o suficiente. Portanto, para todo  $\epsilon > 0$ ,  $P(|X_n Y_n| < \epsilon) \rightarrow 1$ , ou seja,  $X_n Y_n \xrightarrow{P} 0$ .

Agora consideremos o caso  $c$  geral. Como  $X_n Y_n = cX_n + (Y_n - c)X_n$  e  $Y_n - c \xrightarrow{P} 0$ . Pelo caso  $c = 0$ , temos que  $(Y_n - c)X_n \xrightarrow{P} 0$ . Além disso como  $cx$  é uma função contínua, temos  $cX_n \xrightarrow{D} cX$ . Como  $X_n Y_n$  é a soma de dois termos, o primeiro dos quais converge para  $cX$  em distribuição, e o segundo para zero em probabilidade, o resultado é consequência da parte (i).

Prova de (iii): Como  $1/x$  é contínua para  $x \neq 0$ , temos que  $1/Y_n \xrightarrow{P} 1/c$ . Agora, basta aplicar o ítem (ii). ■

# Capítulo 8

## Lei dos Grandes Números

### 8.1 Motivação

Entre outras coisas, a Lei dos grandes Números nos permite formalizar a idéia que à medida que o número de repetições de um experimento cresce, a frequência relativa  $f_A$  de algum evento  $A$  converge (quase certamente) para a probabilidade teórica  $P(A)$ . É este fato que nos permite estimar o valor da probabilidade de um evento  $A$ , baseado na frequência relativa de  $A$  em um grande número de repetições de um experimento. É também este fato que justifica a intuição que temos que eventos com probabilidade próximas de 1, quase sempre ocorrem; e que eventos com probabilidade próximas de 0 quase sempre não ocorrem.

Por exemplo, se uma nova peça for produzida e não tivermos conhecimento anterior sobre quão provável será que a peça seja defeituosa, poderemos proceder à inspeção de um grande número dessas peças, digamos  $N$ , contarmos o número de peças defeituosas dentre elas, por exemplo  $n$ , e depois empregarmos  $n/N$  com uma aproximação da probabilidade de que uma peça seja defeituosa. O número  $n/N$  é uma variável aleatória, e seu valor depende essencialmente de duas coisas. Primeira, o valor de  $n/N$  depende da probabilidade básica, mas desconhecida,  $p$  de que uma peça seja defeituosa. Segunda, depende daquelas  $N$  peças que tenham sido inspecionadas. O que a Lei dos Grandes Números mostra é que se a técnica de selecionar as  $N$  peças for aleatória, então o quociente  $n/N$  convergirá quase certamente para  $p$ . (Evidentemente, a seleção das  $N$  peças é importante. Se fôssemos escolher somente aquelas peças que exibissem algum defeito físico externo, por exemplo, poderíamos prejudicar seriamente nossos cálculos.)

Mais formalmente, considere um experimento básico, com a variável aleatória  $X$  representando o valor de um característico numérico do resultado (no caso anterior, temos que  $X$  seria a função indicadora do evento  $A$ ). Pensemos na realização deste experimento  $N$  vezes ( $N$  grande), de tal maneira que as realizações sejam independentes. Suponhamos que depois de cada realização do experimento registre-se o valor do característico numérico do resultado; chamemos este um valor observado. A Lei dos Grandes Números afirma que a média aritmética dos  $n$  valores observados converge, em certo sentido, para a média  $EX$ , quando  $N \rightarrow \infty$ .

Vamos agora construir um modelo para o experimento repetido que apresentamos acima. Para experimentos dessa natureza, um resultado possível é uma seqüência de  $N$  resultados



possíveis do experimento básico. Como estamos interessados em analisar a convergência para  $N$  grande, se  $\Omega_0$  é o espaço amostral do experimento básico, o espaço amostral do experimento global consiste nas seqüências infinitas de elementos de  $\Omega_0$ , ou seja,

$$\Omega = \{(w_1, w_2, \dots) : w_i \in \Omega_0, i = 1, 2, \dots\} = \Omega_0 \times \Omega_0 \times \dots = \Omega_0^\infty,$$

onde  $w_i$  é o resultado do  $i$ -ésimo ensaio do experimento básico. Podemos completar o modelo utilizando a  $\sigma$ -álgebra produto para  $\mathcal{A}$  e a probabilidade produto para  $P$ ,<sup>1</sup> pois os ensaios são independentes.

Já que vamos registrar um certo característico do  $i$ -ésimo resultado para todo  $i$ , estaremos registrando os valores de uma seqüência de variáveis aleatórias. Intuitivamente,  $X(w_0)$  representa o valor do característico numérico do experimento básico ( $w_0 \in \Omega_0$ ), então, quando o resultado da seqüência de realizações for  $w = (w_1, w_2, \dots)$ , os valores observados serão  $X(w_1), X(w_2), \dots$ . É conveniente representar por  $X_n$  o resultado observado na  $n$ -ésima realização. Assim,  $X_n$  é função do resultado  $w$  do experimento global, com  $X_n(w) = X(w_n)$ , e no decorrer serão registrados os valores das variáveis aleatórias  $X_1, X_2, \dots$ . Notemos que  $X_n$  tem a mesma distribuição de  $X$ , pois trata-se de uma seqüência de repetições do mesmo experimento. Como as  $X_n$  dependem de realizações independentes, elas são independentes, onde  $X_1, X_2, \dots$  são independentes se para todo  $n \geq 2$ ,  $X_1, \dots, X_n$  são independentes.

Uma versão da Lei dos Grandes Números diz que se  $X_1, X_2, \dots$  são i.i.d. e integráveis, então

$$\frac{X_1 + \dots + X_n}{n} \rightarrow EX_1.$$

Quando o tipo de convergência é convergência em probabilidade, chamamos de Lei Fraca dos Grandes Números, e quando temos convergência quase certa, chamamos de Lei Forte dos Grandes Números. Como vimos em capítulo anterior, convergência quase-certa implica convergência em probabilidade, portanto se uma seqüência de variáveis aleatórias satisfaz a Lei Forte dos Grandes Números, então ela também satisfaz a Lei Fraca.

Para esclarecer as diferenças entre as Leis Fraca e Leis Fortes, considere o caso em que  $X_i \sim \text{Bernoulli}(p)$  é a função indicadora de certo evento  $A$  e  $n_A$  é o número de vezes que o evento  $A$  ocorre em  $n$  realizações do experimento. Então, a Lei Fraca afirma que  $\frac{n_A}{n} \xrightarrow{P} p$ , o que é equivalente a dizer que para todo  $\epsilon > 0$  podemos encontrar um  $n$  suficientemente grande tal que, a probabilidade de  $\frac{n_A}{n}$  estar entre  $p - \epsilon$  e  $p + \epsilon$ , é maior que  $1 - \delta$  para qualquer  $\delta > 0$  especificado. Em outras palavras, se realizarmos muitas seqüências  $\text{Bernoulli}(p)$  de tamanho  $n$ , espera-se que apenas em uma fração delas menor que  $\delta$ , temos que  $\frac{n_A}{n}$  está fora do intervalo  $(p - \epsilon, p + \epsilon)$ . Note que a Lei Fraca não dá nenhuma informação sobre a existência ou o valor do limite de  $\frac{n_A}{n}$ . Em contraste, a Lei Forte garante que o conjunto de todas as realizações do experimento, para as quais  $\lim_n \frac{n_A}{n} = p$ , é um evento com probabilidade 1. Se fixarmos  $\epsilon > 0$ , o conjunto das realizações dos experimentos para os quais  $p - \epsilon < \frac{n_A}{n} < p + \epsilon$ , para  $n$  suficientemente grande é um evento com probabilidade 1. A Lei Forte assegura que dado  $\epsilon > 0$ , com probabilidade 1, os termos da seqüência de freqüência relativas de uma particular realização do experimento realmente estarão no intervalo  $(p - \epsilon, p + \epsilon)$ .

<sup>1</sup>Formalmente, dados uma seqüência de espaços de probabilidade  $(\Omega_i, \mathcal{A}_i, P_i)$ , a  $\sigma$ -álgebra produto  $\mathcal{A}$  em  $\times_i \Omega_i$  é definida como sendo a menor  $\sigma$ -álgebra contendo eventos da forma  $A_1 \times A_2 \times \dots$ , onde  $A_i \in \mathcal{A}_i$  para todo  $i$ ; e a probabilidade produto é tal que  $P(A_1 \times A_2 \times \dots) = \prod_{i=1}^\infty P_i(A_i)$ ; pode-se provar que existe uma única medida de probabilidade em  $\mathcal{A}$  que satisfaz esta condição.

## 8.2 Lei Fraca dos Grandes Números

Na seção anterior, motivamos o resultado da Leis dos Grandes Números para variáveis aleatórias independentes e identicamente distribuídas. Nesta seção, analisaremos duas versões da Lei Fraca dos Grandes Números, na primeira delas não é necessário assumir que as variáveis aleatórias são identicamente distribuídas. Vamos usar a desigualdade de Chebyshev para provar a Lei Fraca dos Grandes Números de Chebyshev.

**Teorema 8.2.1: Lei Fraca de Chebyshev** *Sejam  $X_1, X_2, \dots$  variáveis aleatórias independentes 2 a 2 com variâncias finitas e uniformemente limitadas (ou seja, existe  $c$  finito tal que para todo  $n$ ,  $\text{Var}X_n \leq c$ ). Então,  $X_1, X_2, \dots$  satisfazem a Lei Fraca dos Grandes Números:*

$$\frac{S_n - ES_n}{n} \xrightarrow{P} 0.$$

**Prova:** Precisamos provar que para todo  $\epsilon > 0$ ,

$$P\left(\frac{|S_n - ES_n|}{n} \geq \epsilon\right) \rightarrow 0 \text{ quando } n \rightarrow \infty.$$

Como as variáveis aleatórias são independentes 2 a 2, temos que

$$\text{Var}(S_n) = \sum_{i=1}^n \text{Var}(X_i) \leq nc.$$

Pela desigualdade de Chebyshev, temos que

$$P(|S_n - ES_n| \geq n\epsilon) \leq \frac{\text{Var}(S_n)}{\epsilon^2 n^2} \leq \frac{c}{\epsilon^2 n} \rightarrow 0.$$

■

**Corolário 8.2.2: Lei Fraca dos Grandes Números de Bernoulli.** *Consideremos uma seqüência de ensaios binomiais independentes, tendo a mesma probabilidade  $p$  de “sucesso” em cada ensaio. Se  $S_n$  é o número de sucessos nos primeiros  $n$  ensaios, então*

$$\frac{S_n}{n} \xrightarrow{P} p$$

**Prova:** Seja  $X_n = 1$  se o  $n$ -ésimo ensaio é sucesso,  $X_n = 0$  caso contrário. Então,  $X_1, X_2, \dots$  são i.i.d. e integráveis com média  $\mu = p$ . Como  $\text{Var}X_n = p(1-p)$ , a Lei Fraca de Chebyshev implica que  $\frac{S_n - np}{n} \xrightarrow{P} 0$ , ou, equivalentemente,  $\frac{S_n}{n} \xrightarrow{P} p$ . ■

Podemos utilizar a Lei Fraca dos Grandes Números para responder a seguinte questão: quantas repetições de um experimento devemos realizar a fim de termos uma probabilidade ao menos 0,95 para que a frequência relativa difira de  $p = P(A)$  por menos do que, digamos,

0,01? Utilizando a equação (8.1), onde  $S_n$  é o número de ocorrências do evento  $A$  em  $n$  realizações do experimento temos que  $S_n/n = f_A$ ,  $ES_n = np$ ,  $VarS_n = np(1-p)$ , e:

$$P(|f_A - p| \geq 0,01) \leq \frac{p(1-p)}{n(0,01)^2},$$

ou seja, queremos que  $\frac{p(1-p)}{n(0,01)^2} \leq 0,05$ , o que é equivalente a  $n \geq \frac{p(1-p)}{0,05(0,01)^2}$ . Substituindo os valores específicos de 0,05 e 0,01 por  $\delta$  e  $\epsilon$ , respectivamente, teremos

$$P(|f_A - p| < \epsilon) \geq 1 - \delta \text{ sempre que } n \geq \frac{p(1-p)}{\delta(\epsilon)^2}.$$

Em muitos problemas, não conhecemos o valor de  $p = P(A)$  e, por isso, não poderemos empregar o limite acima. Nesse caso, poderemos empregar o fato de que  $p(1-p)$  toma seu valor máximo quando  $p = 1/2$ , e esse valor máximo é igual a  $1/4$ . Conseqüentemente, estamos certamente seguros se afirmamos que para  $n \geq \frac{1}{4\epsilon^2\delta}$  teremos

$$P(|f_A - p| < \epsilon) \geq 1 - \delta.$$

**Exemplo 8.2.3:** Peças são produzidas de tal maneira que a probabilidade de uma peça ser defeituosa é  $p$  (admitida desconhecida). Um grande número de peças, digamos  $n$ , são classificadas como defeituosas ou perfeitas. Que valor deverá ter  $n$  de maneira que possamos estar 99% certos de que a freqüência relativa de defeituosas difere de  $p$  por menos de 0,05?

**Solução:** Porque não conhecemos o valor de  $p$ , deveremos aplicar a última fórmula com  $\epsilon = 0,05$ ,  $\delta = 0,01$ . Deste modo encontraremos que se  $n \geq \frac{1}{4(0,05)^2 0,01} = 10.000$ , a condição exigida será satisfeita.

A hipótese de variâncias finitas pode ser eliminada e o próximo teorema prova uma versão da Lei Fraca dos Grandes Números para variáveis aleatórias i.i.d. e integráveis.

**Teorema 8.2.4: Lei Fraca de Khintchin.** Se  $X_1, X_2, \dots$  são i.i.d. e integráveis com média comum  $\mu$ , então

$$\frac{S_n}{n} \xrightarrow{P} \mu.$$

**Prova:** É consequência da Lei Forte de Kolmogorov e do fato que convergência quase certa implica convergência em probabilidade. ■

**Exemplo 8.2.5:** Sejam  $\{X_n : n \geq 1\}$  variáveis i.i.d. com média  $\mu$  e variância  $\sigma^2$ , ambas finitas. Prove que  $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \xrightarrow{P} \sigma^2$ .

**Solução:**

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 &= \frac{1}{n} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - 2\bar{X} \frac{1}{n} \sum_{i=1}^n X_i + \frac{1}{n} \sum_{i=1}^n \bar{X}^2 \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \end{aligned}$$

Pela Lei Fraca de Kintchin, temos que

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \rightarrow^P E(X_i^2) = \sigma^2 + \mu^2$$

e

$$\bar{X} \rightarrow^P E(X_i) = \mu.$$

Como funções contínuas preservam convergência, temos que

$$\bar{X}^2 \rightarrow^P \mu^2.$$

Logo, temos que

$$\left(\frac{1}{n} \sum_{i=1}^n X_i^2, \bar{X}^2\right) \rightarrow^P (\sigma^2 + \mu^2, \mu^2).$$

Finalmente, como funções contínuas preservam convergência

$$\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \rightarrow^P \sigma^2.$$

### 8.3 Lei Forte dos Grandes Números

Antes de iniciarmos a prova da Lei Forte dos Grandes Números, vamos provar uma extensão da desigualdade de Chebyshev.

**Lema 8.3.1:** *Sejam  $X_1, \dots, X_n$  variáveis aleatórias independentes tais que  $EX_k = 0$  e  $Var X_k < \infty, k = 1, \dots, n$ . Então, para todo  $\lambda > 0$ ,*

$$P(\max_{1 \leq k \leq n} |S_k| \geq \lambda) \leq \frac{1}{\lambda^2} Var S_n = \frac{1}{\lambda^2} \sum_{k=1}^n Var X_k,$$

onde  $S_k = X_1 + \dots + X_k$ .

**Prova:** Queremos uma cota superior para  $P(\max_{1 \leq k \leq n} S_k^2 \geq \lambda^2)$ . Para tanto, seja  $A = [\max_{1 \leq k \leq n} S_k^2 \geq \lambda^2]$ . Vamos decompor  $A$  conforme a primeira vez que  $S_k^2 \geq \lambda^2$ , definamos:

$$\begin{aligned} A_1 &= [S_1^2 \geq \lambda^2], \\ A_2 &= [S_1^2 < \lambda^2, S_2^2 \geq \lambda^2], \\ A_k &= [S_1^2 < \lambda^2, \dots, S_{k-1}^2 < \lambda^2, S_k^2 \geq \lambda^2], \text{ para } 2 \leq k \leq n. \end{aligned}$$

Então os  $A_k$  são disjuntos e  $A = \cup_{k=1}^n A_k$ . Logo,  $I_A = \sum_{k=1}^n I_{A_k}$  e

$$S_n^2 \geq S_n^2 I_A = \sum_{k=1}^n S_n^2 I_{A_k} \Rightarrow ES_n^2 \geq \sum_{k=1}^n ES_n^2 I_{A_k}.$$

Queremos substituir  $S_n^2$  por  $S_k^2$  no somatório (pois  $S_k^2 \geq \lambda^2$  em  $A_k$ , e não vale necessariamente  $S_n^2 \geq \lambda^2$ ); o truque é escrever

$$S_n^2 = (S_n - S_k)^2 + S_k^2 + 2(S_n - S_k)S_k \geq S_k^2 + 2(S_n - S_k)S_k.$$

Portanto,

$$ES_n^2 I_{A_k} \geq ES_k^2 I_{A_k} + 2E((S_n - S_k)S_k I_{A_k}).$$

Como  $S_n - S_k = X_{k+1} + \dots + X_n$  e  $S_k I_{A_k}$  depende só de  $X_1, \dots, X_k$ , as duas são funções de famílias disjuntas de variáveis independentes, logo são independentes e a esperança fatora:

$$E((S_n - S_k)S_k I_{A_k}) = E(S_n - S_k)E(S_k I_{A_k}).$$

Como  $E(S_n - S_k) = 0$ , temos

$$ES_n^2 I_{A_k} \geq ES_k^2 I_{A_k} \geq E\lambda^2 I_{A_k} = \lambda^2 P(A_k).$$

Portanto,

$$ES_n^2 \geq \sum_{k=1}^n \lambda^2 P(A_k) = \lambda^2 P(A),$$

logo

$$P(A) \leq \frac{1}{\lambda^2} ES_n^2 = \frac{1}{\lambda^2} \text{Var} S_n.$$

■

O próximo teorema é conhecido como **Primeira Lei Forte de Kolmogorov**.

**Teorema 8.3.2:** *Sejam  $X_1, X_2, \dots$  variáveis aleatórias independentes e integráveis, e suponha que*

$$\sum_{n=1}^{\infty} \frac{\text{Var} X_n}{n^2} < \infty.$$

Então, as  $X_n$  satisfazem a Lei Forte dos Grandes Números, ou seja,

$$\frac{X_1 + \dots + X_n}{n} - \frac{(EX_1 + \dots + EX_n)}{n} \rightarrow 0 \text{ quase certamente.}$$

**Prova:** Suponhamos sem perda de generalidade que  $EX_n = 0, \forall n$ . Queremos mostrar que  $\frac{S_n}{n} \rightarrow 0$  cp1, onde  $S_n = X_1 + \dots + X_n$ . Para tanto, basta mostrar que

$$M_n = \max_{2^n < k \leq 2^{n+1}} \frac{|S_k|}{k} \rightarrow 0 \text{ cp1 quando } n \rightarrow \infty.$$

Provaremos isto em duas etapas:

- (i)  $\sum_{n=1}^{\infty} P(M_n \geq \frac{1}{m}) < \infty, \forall m = 1, 2, \dots$ ; e
- (ii)  $M_n \rightarrow 0$  cp1.

Para (i), considere  $m$  fixo. Então, para todo  $n$ ,

$$\begin{aligned} P(M_n \geq \frac{1}{m}) &\leq P(\max_{2^n < k \leq 2^{n+1}} |S_k| \geq \frac{2^n}{m}) \leq \\ &\leq P(\max_{1 < k \leq 2^{n+1}} |S_k| \geq \frac{2^n}{m}) \leq \frac{m^2}{4^n} \sum_{k=1}^{2^{n+1}} \text{Var}(X_k), \end{aligned}$$

onde vale a última passagem pelo lema anterior. Seja  $A_n = [M_n \geq \frac{1}{m}]$ , então

$$\begin{aligned} \sum_{n=1}^{\infty} P(A_n) &\leq m^2 \sum_{n=1}^{\infty} (\frac{1}{4^n} \sum_{k=1}^{2^{n+1}} \text{Var}(X_k)) = m^2 \sum_{k=1}^{\infty} \sum_{n: 2^{n+1} \geq k} (\frac{1}{4^n} \text{Var}(X_k)) = \\ &= m^2 \sum_{k=1}^{\infty} \text{Var}(X_k) \sum_{n: 2^{n+1} \geq k} (\frac{1}{4^n}). \end{aligned}$$

Como  $\sum_{n: 2^{n+1} \geq k} (\frac{1}{4^n}) \leq \frac{16}{3k^2}$ , temos

$$\sum_{n=1}^{\infty} P(A_n) \leq \frac{16m^2}{3} \sum_{k=1}^{\infty} \frac{\text{Var}(X_k)}{k^2} < \infty.$$

Para (ii), note que por Borel-Cantelli, tem-se  $P(A_n \text{ infinitas vezes}) = 0$ . Logo, para todo  $m$ , a probabilidade é 1 de que  $M_n$  assumira um valor  $\geq \frac{1}{m}$  para somente um número finito de  $n$ 's. Seja  $B_m$  o evento “ $M_n$  assumira um valor  $\geq \frac{1}{m}$  para somente um número finito de  $n$ 's”, então  $P(B_m) = 1, \forall m$ , o que implica que  $P(\cap_{m=1}^{\infty} B_m) = 1$ , e (ii) resulta da equivalência entre os eventos  $\cap_{m=1}^{\infty} B_m$  e  $[M_n \rightarrow 0]$ . ■

O próximo exemplo ilustra uma aplicação da Primeira Lei Forte de Kolmogorov.

**Exemplo 8.3.3:** Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes com  $X_n \sim \text{Poisson}(\sqrt{n})$ , para cada  $n \geq 1$ . Calcule o limite quase-certo de  $\bar{X}$ .

**Solução:** Como  $\text{Var} X_n = \sqrt{n}$ , temos que

$$\sum_{n=1}^{\infty} \frac{\text{Var} X_n}{n^2} = \sum_{n=1}^{\infty} \frac{\sqrt{n}}{n^2} < \infty.$$

Logo, a primeira Lei Forte de Kolmogorov implica que

$$\bar{X} - \frac{EX_1 + \dots + EX_n}{n} \rightarrow 0 \text{ cp1, ou seja}$$

$$\bar{X} - \frac{\sqrt{1} + \sqrt{2} + \dots + \sqrt{n}}{n} \rightarrow 0 \text{ cp1.}$$

Pelo teste da integral, pode-se verificar que

$$\sqrt{1} + \sqrt{2} + \dots + \sqrt{n} \geq \frac{2n^{3/2}}{3}.$$

Portanto,

$$\frac{\sqrt{1} + \sqrt{2} + \cdots + \sqrt{n}}{n} \geq \frac{2n^{1/2}}{3} \rightarrow \infty.$$

Logo,  $\bar{X} \rightarrow \infty$  cp1.

Antes de enunciarmos e provarmos a Segunda Lei Forte de Kolmogorov, considere o seguinte lema:

**Lema 8.3.4:** *Seja  $X$  uma variável aleatória integrável com função de distribuição  $F$ . Então,*

$$\sum_{n=1}^{\infty} \left( \frac{1}{n^2} \int_{-n}^n x^2 dF(x) \right) < \infty.$$

**Prova:** Vamos utilizar o seguinte fato:  $\sum_{n=j}^{\infty} \frac{1}{n^2} \leq \frac{2}{j}$  para  $j = 1, 2, \dots$ . Como

$$\int_{-n}^n x^2 dF(x) = \sum_{j=-n+1}^n \int_{j-1}^j x^2 dF(x),$$

temos

$$\begin{aligned} \sum_{n=1}^{\infty} \left( \frac{1}{n^2} \int_{-n}^n x^2 dF(x) \right) &= \sum_{n=1}^{\infty} \sum_{j=-n+1}^n \left( \frac{1}{n^2} \int_{j-1}^j x^2 dF(x) \right) = \\ &= \sum_{n=1}^{\infty} \sum_{j=1}^n \left( \frac{1}{n^2} \int_{j-1}^j x^2 dF(x) \right) + \sum_{n=1}^{\infty} \sum_{j=-n+1}^0 \left( \frac{1}{n^2} \int_{j-1}^j x^2 dF(x) \right) \\ &= \sum_{j=1}^{\infty} \sum_{n=j}^{\infty} \left( \frac{1}{n^2} \int_{j-1}^j x^2 dF(x) \right) + \sum_{j=-\infty}^0 \sum_{n=|j|+1}^{\infty} \left( \frac{1}{n^2} \int_{j-1}^j x^2 dF(x) \right) \leq \\ &\leq 2 \sum_{j=1}^{\infty} \int_{j-1}^j \frac{x^2}{j} dF(x) + 2 \sum_{j=-\infty}^0 \int_{j-1}^j \frac{x^2}{|j|+1} dF(x). \end{aligned}$$

Como  $\frac{x^2}{j} \leq x$  em  $(j-1, j]$ , para  $j \geq 1$ , e  $\frac{x^2}{|j|+1} \leq |x|$  em  $(j-1, j]$ , para  $j \leq 0$ , temos

$$\begin{aligned} \sum_{n=1}^{\infty} \left( \frac{1}{n^2} \int_{-n}^n x^2 dF(x) \right) &\leq 2 \sum_{j=1}^{\infty} \int_{j-1}^j x dF(x) + 2 \sum_{j=-\infty}^0 \int_{j-1}^j |x| dF(x) = \\ &= 2 \sum_{j=-\infty}^{\infty} \int_{j-1}^j |x| dF(x) = 2 \int_{-\infty}^{\infty} |x| dF(x) = 2E|X| < \infty. \end{aligned}$$

■

A seguir enunciarmos e provamos a **Segunda Lei Forte de Kolmogorov**.

**Teorema 8.3.5:** *Sejam  $X_1, X_2, \dots$  variáveis aleatórias independentes, identicamente distribuídas e integráveis, com  $EX_n = \mu$ . Então,*

$$\frac{X_1 + \dots + X_n}{n} \rightarrow \mu \text{ quase certamente.}$$

**Prova:** Suponhamos sem perda de generalidade que  $\mu = 0$ . Vamos truncar as variáveis  $X_n$ , definamos  $Y_n = X_n I_{[-n < X_n \leq n]}$ . Seja  $Z_n = X_n - Y_n$ , de modo que

$$\frac{X_1 + \dots + X_n}{n} = \frac{Y_1 + \dots + Y_n}{n} + \frac{Z_1 + \dots + Z_n}{n}.$$

A prova terá três partes:

- (a)  $\frac{Z_1 + \dots + Z_n}{n} \rightarrow 0$  quase certamente (usaremos Borel-Cantelli);
- (b)  $\frac{Y_1 + \dots + Y_n}{n} - \frac{EY_1 + \dots + EY_n}{n} \rightarrow 0$  quase certamente (usaremos a Primeira Lei Forte e o Lema 8.3.4); e
- (c)  $\frac{EY_1 + \dots + EY_n}{n} \rightarrow 0$  (usaremos o Teorema da Convergência Dominada).

É fácil ver que (a), (b), e (c) implicam o teorema. Para provar (a), note que  $Z_n \neq 0 \Leftrightarrow Y_n \neq X_n \Leftrightarrow X_n \notin (-n, n]$ . Logo,

$$P(Z_n \neq 0) = P(X_n \notin (-n, n]) \leq P(|X_n| \geq n).$$

Mas os eventos  $A_n = [Z_n \neq 0]$  satisfazem

$$\sum_{n=1}^{\infty} P(A_n) \leq \sum_{n=1}^{\infty} P(|X_n| \geq n) = \sum_{n=1}^{\infty} P(|X_1| \geq n) \leq E|X_1| < \infty.$$

Portanto, Borel-Cantelli implica que  $P(A_n \text{ infinitas vezes}) = 0$ , ou seja

$$P(Z_n \neq 0 \text{ infinitas vezes}) = 0.$$

Isso significa que

$$P(Z_n = 0 \text{ para todo } n \text{ suficientemente grande}) = 1.$$

Mas se  $Z_n = 0$  para  $n$  suficientemente grande, então  $Z_n \rightarrow 0$  e

$$\frac{Z_1 + \dots + Z_n}{n} \rightarrow 0, \text{ logo } P\left(\frac{Z_1 + \dots + Z_n}{n} \rightarrow 0\right) = 1.$$

Para provar (b), seja  $F$  a função de distribuição comum,  $F = F_{X_n}$ . Verifiquemos a condição da primeira Lei Forte de Kolmogorov para as variáveis aleatórias  $Y_n$ . Como  $Y_n = X_n I_{[-n < X_n \leq n]}$ , temos

$$\text{Var}(Y_n) \leq E(Y_n^2) = E(X_n^2 I_{[-n < X_n \leq n]}) = \int_{-n}^n x^2 dF(x).$$



Portanto,

$$\sum_{n=1}^{\infty} \frac{\text{Var}(Y_n)}{n^2} \leq \sum_{n=1}^{\infty} \frac{1}{n^2} \int_{-n}^n x^2 dF(x) < \infty,$$

onde a última desigualdade decorre do Lema 8.3.4. Portanto, (b) decorre da primeira Lei Forte de Kolmogorov.

Para provar (c), é suficiente mostrar que  $EY_n \rightarrow 0$ . Mas,

$$EY_n = E(X_n I_{[-n < X_n \leq n]}) = E(X_1 I_{[-n < X_1 \leq n]}) \rightarrow EX_1 = 0,$$

pelo teorema da convergência dominada que se aplica pois  $|X_1|$  domina  $X_1 I_{[-n \leq X_1 \leq n]}$ 's e é integrável. ■

**Exemplo 8.3.6:** As variáveis  $X_n$ ,  $n \geq 1$ , são independentes e todas têm distribuição Exponencial de parâmetro  $\lambda$ . Mostre que a seqüência  $\{X_n^2 : n \geq 1\}$  satisfaz a Lei Forte dos Grandes Números.

**Solução:** De acordo com a Segunda Lei Forte de Kolmogorov, precisamos mostrar que  $EX_n^2$  é finita para todo  $n$ . Como  $EX_n^2 = \text{Var}X_n + (EX_n)^2 = \frac{2}{\lambda^2} < \infty$ , temos que a seqüência  $\{X_n^2 : n \geq 1\}$  satisfaz a Lei Forte dos Grandes Números.

**Exemplo 8.3.7:** Seja  $\{X_n : n \geq 1\}$  uma seqüência de variáveis aleatórias i.i.d., seguindo o modelo Uniforme contínuo em  $(0, 1)$ . Calcule o limite, quase certo, para  $\frac{1}{n} \sum_{k=1}^n (-\log(X_k))$  quando  $n \rightarrow \infty$ .

**Solução:** Vamos tentar usar a Lei Forte dos Grandes Números. Para isso, precisamos calcular  $E(-\log X_k)$ .

$$E(-\log X_k) = \int_0^1 -\log x dx = -x \log x \Big|_0^1 + \int_0^1 dx = 1.$$

Portanto, temos que  $\frac{1}{n} \sum_{k=1}^n (-\log(X_k)) \rightarrow 1$  cp1.

A seguir veremos uma importante conseqüência da Lei Forte dos Grandes Números para a área de Estatística Aplicada. Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias em  $(\Omega, \mathcal{A}, P)$  independentes e identicamente distribuídas com função de distribuição  $F$ . Essas variáveis podem representar a amostra observada de uma certa quantidade de interesse. A *função de distribuição empírica* ou *amostral*, denotada por  $F_n^e$ , é definida para todo  $x \in \mathbb{R}$  e  $w \in \Omega$  por:

$$F_n^e(x, w) = \frac{1}{n} [\text{número de } i\text{'s tais que } X_i(w) \leq x, i = 1, 2, \dots, n].$$

Para uma particular trajetória  $w_0 \in \Omega$ , obtemos o conjunto de valores fixados  $X_1(w_0) = x_1, \dots, X_n(w_0) = x_n$ . Se os  $x_i$ 's são todos diferentes, então  $F_n^e(x, w_0)$  é uma função de distribuição com saltos  $1/n$  em cada um desses valores.

Considere um  $x_0 \in \mathbb{R}$  fixo. Então  $F_n^e(x_0, w)$  é uma variável aleatória, pois é uma função das variáveis  $X_1, X_2, \dots, X_n$ . Se  $Y_i = I_{X_i \leq x_0}$ ,  $i = 1, 2, \dots, n$ , então  $F_n^e(x_0, w) = \frac{1}{n} \sum_{i=1}^n Y_i(w)$ . Como as variáveis aleatórias  $Y_i$  são funções de famílias disjuntas de variáveis aleatórias

independentes, elas também são independentes. Além disso, temos que  $Y_i \sim \text{Bernoulli}(p)$  com

$$p = P(Y_i = 1) = P(X_i \leq x_0) = F(x_0).$$

Portanto, concluímos que pela Lei Forte de Kolmogorov, para cada valor  $x_0 \in \mathbb{R}$  fixo, temos  $F_n^e(x_0, w) \rightarrow F(x_0)$  cp1. O *Teorema de Glivenko-Cantelli* também conhecido como *Teorema Fundamental da Estatística* afirma que a função de distribuição empírica converge para a função de distribuição populacional, quase certamente em  $\Omega$  e uniformemente em  $\mathbb{R}$ .

**Teorema 8.3.8:** *Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias em  $(\Omega, \mathcal{A}, P)$ , independentes e identicamente distribuídas com função de distribuição  $F$ . Seja  $F_n^e$  a correspondente função de distribuição empírica, então:*

$$P(\limsup_n \sup_{x \in \mathbb{R}} |F_n^e(x, w) - F(x)| = 0) = 1.$$

**Prova:** Para cada  $x$  fixo, os argumentos anteriores garantem convergência quase certa. A prova de que este resultado pode ser estendido, usa técnicas de Análise Matemática e será omitida. ■

Por fim nós enunciaremos e provaremos a **Recíproca da Lei Forte de Kolmogorov**. A Lei Forte afirma que se as variáveis aleatórias  $X_n$  são integráveis, então  $\frac{S_n}{n}$  converge para um limite finito ( $= EX_1$ ) com probabilidade 1. A recíproca diz que se as  $X_n$  não forem integráveis, então com probabilidade 1,  $\frac{S_n}{n}$  não convergirá para um limite finito.

**Teorema 8.3.9:** *Sejam  $X_1, X_2, \dots$  variáveis aleatórias independentes e identicamente distribuídas. Se  $E|X_1| = \infty$ , então, com probabilidade 1, a seqüência  $\frac{|S_n|}{n}$  não é limitada.*

**Prova:** Se  $E|X_1| = \infty$ , então  $E(\frac{|X_1|}{k}) = \infty$ , para  $k = 1, 2, \dots$ . De acordo com Lema 4.6.2, temos que

$$\sum_{n=1}^{\infty} P(\frac{|X_1|}{k} \geq n) = \infty, \forall k.$$

Como as variáveis  $X_n$  são identicamente distribuídas, temos

$$\sum_{n=1}^{\infty} P(\frac{|X_1|}{k} \geq n) = \sum_{n=1}^{\infty} P(\frac{|X_n|}{k} \geq n) = \sum_{n=1}^{\infty} P(\frac{|X_n|}{n} \geq k).$$

Por independência dos  $X_n$ , os eventos  $A_n = [\frac{|X_n|}{n} \geq k]$  são independentes, e Borel-Cantelli implica

$$P(\frac{|X_n|}{n} \geq k \text{ infinitas vezes}) = 1, \forall k.$$

Fazendo  $B_k = [\frac{|X_n|}{n} \geq k \text{ infinitas vezes}]$ , temos  $P(\cap_{k=1}^{\infty} B_k) = 1$ , pois a intersecção de um número enumerável de eventos de probabilidade 1 também tem probabilidade 1. Mas o evento  $\cap_{k=1}^{\infty} B_k$  é o evento " $\frac{|X_n|}{n} > k$  para um número infinito de  $n$ , para todo  $k$ ", ou seja, é

o evento “a seqüência  $\frac{|X_n|}{n}$  é ilimitada.” Para terminar a prova, basta mostrar que se  $\frac{|X_n|}{n}$  é ilimitada, então  $\frac{|S_n|}{n}$  também é ilimitada. Agora, com  $S_0 = 0$ , temos

$$\frac{|X_n|}{n} = \frac{|S_n - S_{n-1}|}{n} \leq \frac{|S_n|}{n} + \frac{|S_{n-1}|}{n},$$

para  $n = 1, 2, \dots$ . Portanto, se  $\frac{|X_n|}{n}$  é ilimitada, então  $\frac{|S_n|}{n}$  é ilimitada ou  $\frac{|S_{n-1}|}{n}$  é ilimitada. Mas,

$$\frac{|S_{n-1}|}{n} = \frac{|S_{n-1}|}{(n-1)} \cdot \frac{(n-1)}{n},$$

então  $\frac{|S_{n-1}|}{n}$  é ilimitada se, e somente se,  $\frac{|S_n|}{n}$  também for. ■

## 8.4 Um Exemplo de Divergência das Médias

Uma variável aleatória tem distribuição de Cauchy de parâmetro  $a$  se, para  $a > 0$

$$f_X(x) = \frac{1}{\pi} \cdot \frac{a}{a^2 + x^2}.$$

Assuma que  $X_n$  são i.i.d. segundo uma distribuição de Cauchy de parâmetro  $a$ . Seja  $S_n = \frac{1}{n} \sum_{i=1}^n X_i$ . Utilizando a definição e as propriedades da função característica pode-se provar que

$$\varphi_{X_n}(u) = e^{-a|u|}, \text{ e } \varphi_{S_n}(u) = e^{-a|u|}.$$

Então, as médias  $S_n$  são distribuídas exatamente como uma das parcelas da soma. Para  $n \geq m$ , após alguma manipulação algébrica, temos que

$$S_n - S_m = \left(1 - \frac{m}{n}\right)([Z_{n,m}] - [Y_{n,m}]),$$

onde  $Z_{n,m} = \frac{1}{n-m} \sum_{i=m+1}^n X_i$  e  $Y_{n,m} = \frac{1}{m} \sum_{i=1}^m X_i$ . Observe que como  $Z_{n,m}$  e  $Y_{n,m}$  são médias de conjuntos disjuntos de variáveis aleatórias independentes, elas são independentes uma da outra. Ainda mais, pelo resultado para  $\varphi_{S_n}$ , é o caso que elas são identicamente distribuídas com função característica igual a  $e^{-a|u|}$ . Seja  $W_{n,m} = Z_{n,m} - Y_{n,m}$ , nós vemos que  $S_n - S_m = \left(1 - \frac{m}{n}\right)W_{n,m}$ . Contudo,

$$\varphi_{W_{n,m}}(u) = \varphi_{Z_{n,m}}(u)\varphi_{Y_{n,m}}(-u) = e^{-2a|u|}.$$

Então,  $W_{n,m}$  tem uma distribuição fixa, não degenerada que é independente de  $n$  e  $m$ . Fixando,  $n = 2m$ , temos que

$$\varphi_{S_{2m}-S_m}(u) = e^{-a|u|}.$$

Portanto, quando  $m \rightarrow \infty$ ,  $S_{2m} - S_m$  não converge para zero, mas para todo  $m$ , tem uma distribuição Cauchy de parâmetro  $a$ . Portanto,  $S_n$  não satisfaz o critério de convergência de Cauchy e não é convergente.

Observe que isto não é um contra-exemplo a Lei Forte de Kolmogorov, tendo em vista que uma variável aleatória que tem distribuição de acordo com uma Cauchy não tem valor esperado definido, ou seja

$$EX = - \int_{-\infty}^0 \frac{1}{\pi} \frac{a|x|}{a^2 + x^2} dx + \int_0^{\infty} \frac{1}{\pi} \frac{ax}{a^2 + x^2} dx,$$

é indefinido, visto que ambas as integrais são infinitas. Este exemplo serve para ilustrar que a suposição da existência de  $EX$  é necessária para a Lei Forte dos Grandes Números.

# Capítulo 9

## Teorema Central do Limite

### 9.1 Motivação

Consideremos uma seqüência de variáveis aleatórias independentes,  $X_1, X_2, \dots$ , definidas no mesmo espaço de probabilidade  $(\Omega, \mathcal{A}, P)$ , e seja  $S_1, S_2, \dots$  a seqüência de somas parciais, definidas por  $S_n = X_1 + X_2 + \dots + X_n$ . A Lei dos Grandes Números trata da convergência de  $\frac{1}{n}(S_n - ES_n)$  para 0, quando  $n \rightarrow \infty$ , supondo que as variáveis aleatórias  $X_i$ 's sejam integráveis. Quando a seqüência obedece à lei dos grandes números, existe uma tendência da variável aleatória  $\frac{S_n}{n}$ , a média amostral no caso de variáveis aleatórias independentes e identicamente distribuídas, para concentrar-se em torno de sua média. O Teorema Central do Limite prova que sob certas hipóteses gerais, a distribuição da média amostral padronizada tende à normal. O problema consiste em achar condições sob as quais

$$\frac{S_n - ES_n}{\sqrt{VarS_n}} \rightarrow^D N(0, 1).$$

Resumidamente, estas condições exigem que cada parcela da soma contribua com um valor sem importância para a variação da soma, ou seja é muito improvável que qualquer parcela isolada dê uma contribuição muito grande para a soma.

O Teorema Central do Limite dá apoio ao uso da normal como distribuição de erros, pois em muitas situações reais é possível interpretar o erro de uma observação como resultante de muitos erros pequenos e independentes. Há também outras situações que o Teorema Central do Limite pode justificar o uso da normal. Por exemplo, a distribuição de alturas de homens adultos de certa idade pode ser considerada aproximadamente normal, pois a altura pode ser pensada como soma de muitos efeitos pequenos e independentes.

### 9.2 Teoremas e provas

Existem vários Teoremas Centrais do Limite que variam de acordo com as hipóteses sobre as distribuições das variáveis aleatórias  $X_i$ 's na seqüência. Como teoremas centrais do limite tratam de convergência em distribuição e como, pelo Teorema da Continuidade de Levy, sabe-se que uma seqüência de variáveis aleatórias  $Y_n \rightarrow^D Y$  se, e somente se,  $\varphi_{Y_n} \rightarrow \varphi_Y$ ,

a idéia será provar que a função característica de  $\frac{S_n - ES_n}{\sqrt{VarS_n}}$  converge para  $e^{-\frac{t^2}{2}}$  que é a função característica da  $N(0, 1)$ . Nós iremos agora enunciar e provar alguns desses teoremas, começando pelo caso de variáveis aleatórias independentes e identicamente distribuídas.

**Teorema 9.2.1:** *Sejam  $X_1, X_2, \dots$  variáveis aleatórias iid com  $E(X_n) = \mu$  e  $Var(X_n) = \sigma^2$ . Suponha que  $N$  é uma variável aleatória com distribuição  $N(0, 1)$ . Se  $S_n = X_1 + X_2 + \dots + X_n$ , então*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \rightarrow^D N.$$

**Prova:** Sem perda de generalidade, seja  $E(X_n) = 0$  e  $E(X_n^2) = 1$  (caso este não seja o caso, pode-se provar o resultado para

$$X_i^* = \frac{X_i - \mu}{\sigma},$$

já que  $E(X_i^*) = 0$  e  $E(X_i^*)^2 = 1$ ).

Seja  $\varphi_n(t) = E(e^{it\frac{S_n}{\sqrt{n}}})$  e  $\varphi(t) = E(e^{itX_1})$ . Como a função característica de uma soma de variáveis aleatórias independentes é igual ao produto das funções características das variáveis aleatórias, tem-se que

$$\varphi_n(t) = (E(e^{it\frac{X_1}{\sqrt{n}}}))^n = \varphi^n(t/\sqrt{n}).$$

Como os dois primeiros momentos existem,  $\varphi$  possui duas derivadas contínuas. Então, utilizando a expansão de Taylor de  $\varphi$  e o fato que  $\varphi^{(k)}(0) = i^k E(X_1^k)$ , temos que

$$\varphi(t) = 1 + t\varphi'(0) + \frac{t^2}{2}\varphi''(\theta(t)),$$

onde  $|\theta(t)| \leq |t|$ . Logo, como  $\varphi''$  é contínua em 0, temos que  $\varphi''(\theta(t)) - \varphi''(0) \rightarrow 0$  quando  $t \rightarrow 0$ . Então, tem-se

$$\varphi(t) = 1 - \frac{t^2}{2} + \frac{t^2}{2}e(t),$$

onde  $e(t) = \varphi''(\theta(t)) + 1$  e  $\lim_{t \rightarrow 0} e(t) = 0$ . Então, para  $t$  fixo

$$\varphi^n\left(\frac{t}{\sqrt{n}}\right) = \left[1 - \frac{t^2}{2n} + \frac{t^2}{2n}e\left(\frac{t}{\sqrt{n}}\right)\right]^n = \left[1 + \frac{-t^2}{2n}\left[1 - e\left(\frac{t}{\sqrt{n}}\right)\right]\right]^n \rightarrow e^{-\frac{t^2}{2}},$$

quando  $n \rightarrow \infty$ , pois  $[1 - e(\frac{t}{\sqrt{n}})] \rightarrow 1$  e para números complexos  $c_n \rightarrow c \Rightarrow (1 + \frac{c_n}{n})^n \rightarrow e^c$  (Esse limite é conhecido como limite de Euler e sua prova será omitida). ■

Um caso especial do Teorema Central do Limite para variáveis aleatórias independentes e identicamente distribuídas é quando estas variáveis são distribuídas de acordo com a distribuição de Bernoulli, este caso é conhecido como Teorema Central do Limite de De Moivre e Laplace.

**Corolário 9.2.2:** *Seja  $X_1, X_2, \dots$  uma seqüência de variáveis aleatórias independentes e distribuídas de acordo com a distribuição de Bernoulli com parâmetro  $p$ , ou seja,  $P(X_i = 1) = p = 1 - P(X_i = 0)$  para  $0 < p < 1$ . Então, se  $S_n = X_1 + \dots + X_n$ ,*

$$\frac{S_n - np}{\sqrt{np(1-p)}} \rightarrow^D N(0, 1).$$

**Prova:** É imediata dado o teorema anterior, já que  $E(X_i) = p$  e  $E(X_i^2) = p$ . ■

**Exemplo 9.2.3:** Suponha que temos algumas voltagens de ruídos independentes, por exemplo  $V_i, i = 1, 2, \dots, n$ , as quais são recebidas naquilo que se denomina um “somador”. Seja  $V$  a soma das voltagens recebidas. Suponha também que cada variável aleatória  $V_i$  seja uniformemente distribuída sobre o intervalo  $[0, 10]$ . Daí,  $EV_i = 5$  volts e  $VarV_i = \frac{100}{12}$ . De acordo com o Teorema Central do Limite, se  $n$  for suficientemente grande, a variável aleatória

$$S = \frac{(V - 5n)\sqrt{12}}{10\sqrt{n}}$$

terá aproximadamente a distribuição  $N(0, 1)$ . Portanto, se  $n = 20$ , podemos calcular que a probabilidade de que a voltagem total na entrada exceda 105 volts da seguinte maneira:

$$P(V > 105) = P\left(\frac{(V - 100)\sqrt{12}}{10\sqrt{20}} > \frac{(105 - 100)\sqrt{12}}{10\sqrt{20}}\right) \simeq 1 - \Phi(0, 388) = 0, 352.$$

Agora analisaremos um resultado mais forte que dá condições gerais que garantem convergência da média amostral padronizada para normal: o Teorema Central do Limite de Lindeberg.

**Teorema 9.2.4:** *Sejam  $X_1, X_2, \dots$  variáveis aleatórias independentes tais que  $E(X_n) = \mu_n$  e  $Var(X_n) = \sigma_n^2 < \infty$ , onde pelo menos um  $\sigma_i^2 > 0$ . Sejam  $S_n = X_1 + \dots + X_n$  e  $s_n = \sqrt{Var(S_n)} = \sqrt{\sigma_1^2 + \dots + \sigma_n^2}$ . Considere a seguinte condição, conhecida como condição de Lindeberg,*

$$\forall \epsilon > 0, \lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{k=1}^n \int_{|x - \mu_k| > \epsilon s_n} (x - \mu_k)^2 dF_k(x) = 0.$$

Então, se a condição de Lindeberg é satisfeita

$$\frac{S_n - ES_n}{s_n} \rightarrow^D N(0, 1).$$

Antes de provarmos este teorema, vamos primeiro dar alguma intuição sobre a condição de Lindeberg. Esta condição diz que, para  $n$  grande, a parcela da variância devida às caudas das  $X_k$  é desprezível.

A condição de Lindeberg implica que as parcelas  $X_k$  da soma têm variâncias uniformemente pequenas para  $n$  grande, em outras palavras nenhuma parcela tem muito peso na soma. Formalmente, a condição de Lindeberg implica que  $\max_{1 \leq k \leq n} \frac{\sigma_k^2}{s_n^2} \rightarrow 0$  quando  $n \rightarrow \infty$ . Para ver isto, observe que para todo  $k$ ,

$$\begin{aligned} \frac{\sigma_k^2}{s_n^2} &= \frac{1}{s_n^2} \int_{|x - \mu_k| \leq \epsilon' s_n} (x - \mu_k)^2 dF_k(x) + \frac{1}{s_n^2} \int_{|x - \mu_k| > \epsilon' s_n} (x - \mu_k)^2 dF_k(x) \leq \\ &\frac{1}{s_n^2} \int_{|x - \mu_k| \leq \epsilon' s_n} (\epsilon' s_n)^2 dF_k(x) + \frac{1}{s_n^2} \sum_{j=1}^n \int_{|x - \mu_j| > \epsilon' s_n} (x - \mu_j)^2 dF_j(x) \leq \\ &\frac{1}{s_n^2} \int_{-\infty}^{\infty} (\epsilon' s_n)^2 dF_k(x) + \frac{1}{s_n^2} \sum_{j=1}^n \int_{|x - \mu_j| > \epsilon' s_n} (x - \mu_j)^2 dF_j(x). \end{aligned}$$

Este último termo não depende de  $k$ , pois a primeira parcela é igual a  $(\epsilon')^2$ . Portanto, temos

$$\max_{1 \leq k \leq n} \frac{\sigma_k^2}{s_n^2} \leq (\epsilon')^2 + \frac{1}{s_n^2} \sum_{k=1}^n \int_{|x-\mu_k| > \epsilon' s_n} (x - \mu_k)^2 dF_k(x),$$

que converge para  $(\epsilon')^2$ , pela condição de Lindeberg. Como isto vale para todo  $\epsilon'$ , temos  $\max_{1 \leq k \leq n} \frac{\sigma_k^2}{s_n^2} \rightarrow 0$ .

Portanto, o Teorema Central do Limite de Lindeberg pode ser aplicado para justificar o seguinte raciocínio: a soma de um grande número de pequenas quantidades independentes tem aproximadamente uma distribuição normal.

**Exemplo 9.2.5:** Vamos verificar neste exemplo que uma seqüência  $X_1, X_2, \dots$  de variáveis aleatórias i.i.d. com  $EX_i = \mu$  e  $Var X_i = \sigma^2$  satisfaz a condição de Lindeberg. Note que  $s_n = \sqrt{Var S_n} = \sigma\sqrt{n}$ . Então para  $\epsilon > 0$ , e  $F$  a distribuição comum das variáveis aleatórias:

$$\begin{aligned} \frac{1}{s_n^2} \sum_{k=1}^n \int_{|x-\mu_k| > \epsilon s_n} (x - \mu_k)^2 dF_k(x) &= \frac{1}{n\sigma^2} \sum_{k=1}^n \int_{|x-\mu| > \epsilon\sigma\sqrt{n}} (x - \mu)^2 dF(x) \\ &= \frac{1}{n\sigma^2} n \int_{|x-\mu| > \epsilon\sigma\sqrt{n}} (x - \mu)^2 dF(x). \end{aligned}$$

Então, finalmente,

$$\lim_n \frac{1}{\sigma^2} \int_{|x-\mu| > \epsilon\sigma\sqrt{n}} (x - \mu)^2 dF(x) = 0.$$

Agora iremos provar o Teorema Central do Limite de Lindeberg.

**Prova:** Assim como no caso de variáveis aleatórias i.i.d., mostraremos que a função característica de  $\frac{S_n - ES_n}{s_n}$  converge para  $e^{-\frac{t^2}{2}}$ .

Para tanto, fixemos  $t \in R$ . Usaremos duas versões da fórmula de Taylor aplicada à função  $g(x) = e^{itx}$ :

$$e^{itx} = 1 + itx + \theta_1(x) \frac{t^2 x^2}{2}, \text{ onde } |\theta_1(x)| \leq 1$$

e

$$e^{itx} = 1 + itx - \frac{t^2 x^2}{2} + \theta_2(x) \frac{t^3 x^3}{6}, \text{ onde } |\theta_2(x)| \leq 1.$$

Seja  $\epsilon > 0$ . Usando a primeira fórmula para  $|x| > \epsilon$  e a segunda para  $|x| \leq \epsilon$ , podemos escrever  $e^{itx}$  da seguinte forma geral:

$$e^{itx} = 1 + itx - \frac{t^2 x^2}{2} + r_\epsilon(x),$$

onde

$$r_\epsilon(x) = \begin{cases} (1 + \theta_1(x)) \frac{t^2 x^2}{2} & \text{se } |x| > \epsilon, \\ \theta_2(x) \frac{t^3 x^3}{6} & \text{se } |x| \leq \epsilon. \end{cases}$$

Portanto,



$$\begin{aligned}
E(e^{it\frac{X_k-\mu_k}{s_n}}) &= \int e^{it\frac{x-\mu_k}{s_n}} dF_k(x) = \int (1 + it\frac{x-\mu_k}{s_n} - \frac{t^2(\frac{x-\mu_k}{s_n})^2}{2} + \\
&+ r_\epsilon(\frac{x-\mu_k}{s_n}))dF_k(x) = 1 + itE(\frac{X_k-\mu_k}{s_n}) - \frac{t^2}{2}E((\frac{X_k-\mu_k}{s_n})^2) + \\
&+ \frac{t^2}{2} \int_{|x-\mu_k|>\epsilon s_n} (1 + \theta_1(\frac{x-\mu_k}{s_n}))(\frac{x-\mu_k}{s_n})^2 dF_k(x) + \\
&\frac{t^3}{6} \int_{|x-\mu_k|\leq\epsilon s_n} \theta_2(\frac{x-\mu_k}{s_n})(\frac{x-\mu_k}{s_n})^3 dF_k(x).
\end{aligned}$$

Como  $EX_k = \mu_k$  e  $Var(X_k) = \sigma_k^2$ , temos

$$E(e^{it\frac{X_k-\mu_k}{s_n}}) = 1 - \frac{t^2\sigma_k^2}{2s_n^2} + e_{n,k},$$

onde o resto  $e_{n,k}$  satisfaz

$$\begin{aligned}
|e_{n,k}| &\leq t^2 \int_{|x-\mu_k|>\epsilon s_n} (\frac{x-\mu_k}{s_n})^2 dF_k(x) + \frac{|t^3|}{6} \int_{|x-\mu_k|\leq\epsilon s_n} \epsilon(\frac{x-\mu_k}{s_n})^2 dF_k(x) \\
&\leq \frac{t^2}{s_n^2} \int_{|x-\mu_k|>\epsilon s_n} (x-\mu_k)^2 dF_k(x) + \frac{\epsilon|t^3|}{6s_n^2} \int_{-\infty}^{\infty} (x-\mu_k)^2 dF_k(x).
\end{aligned}$$

Temos então,

$$\sum_{k=1}^n |e_{n,k}| \leq \frac{t^2}{s_n^2} \sum_{k=1}^n \int_{|x-\mu_k|>\epsilon s_n} (x-\mu_k)^2 dF_k(x) + \frac{\epsilon|t^3|}{6}.$$

Pela condição de Lindeberg, a primeira parcela do termo à direita tende a zero quando  $n \rightarrow \infty$ . Logo, para  $n$  suficientemente grande,

$$\sum_{k=1}^n |e_{n,k}| \leq \frac{\epsilon|t^3|}{3}.$$

Vamos então escolher uma seqüência de  $\epsilon$ 's que converge para zero. Para  $\epsilon = \frac{1}{m}$ , existe  $n_m$  tal que para  $n \geq n_m$ ,

$$\sum_{k=1}^n |e_{n,k}| \leq \frac{|t^3|}{3m}, \quad (9.1)$$

onde os restos  $e_{n,k}$  são os determinados pela fórmula baseada em  $\epsilon = \frac{1}{m}$ . Portanto, existe uma seqüência de inteiros positivos  $n_1 < n_2 < \dots$  tal que (9.1) é satisfeita para  $n_m \leq n < n_{m+1}$ , onde para estes valores de  $n$  os restos são baseados em  $\epsilon = \frac{1}{m}$ . É importante lembrar durante o restante da prova que o valor de  $\epsilon$  que determina o resto  $e_{n,k}$  depende da posição de  $n$  em relação aos  $n_m$ . Temos, então,

$$\sum_{k=1}^n |e_{n,k}| \rightarrow 0 \text{ quando } n \rightarrow \infty.$$

Como  $X_i$ 's são independentes,

$$\varphi_{\frac{S_n - ES_n}{s_n}}(t) = \prod_{k=1}^n E\left(e^{it \frac{X_k - \mu_k}{s_n}}\right) = \prod_{k=1}^n \left(1 - \frac{t^2 \sigma_k^2}{2s_n^2} + e_{n,k}\right).$$

Para provar que o termo à direita converge para  $e^{-\frac{t^2}{2}}$ , usaremos o seguinte Lema sobre números complexos.

**Lema 9.2.6:** *Sejam  $c_{n,k}$  números complexos tais que  $\sum_{k=1}^n c_{n,k} \rightarrow c$  quando  $n \rightarrow \infty$ . Se*

$$\max_{1 \leq k \leq n} |c_{n,k}| \rightarrow 0 \text{ quando } n \rightarrow \infty$$

e

$$\sum_{k=1}^n |c_{n,k}| \leq M < \infty,$$

onde  $M$  é uma constante que não depende de  $n$ , então

$$\prod_{k=1}^n (1 + c_{n,k}) \rightarrow e^c \text{ quando } n \rightarrow \infty.$$

**Prova:** Nós omitimos a prova deste lema que pode ser encontrada no livro do Chung seção 7.1. ■

Em nosso caso, sejam  $c_{n,k} = -\frac{t^2 \sigma_k^2}{2s_n^2} + e_{n,k}$  e  $c = -\frac{t^2}{2}$ . Temos que

$$\sum_{k=1}^n |c_{n,k}| \leq \frac{t^2}{2} + \sum_{k=1}^n |e_{n,k}| \rightarrow \frac{t^2}{2},$$

logo existe  $M < \infty$  tal que  $\forall n, \sum_{k=1}^n |c_{n,k}| < M$ . Para aplicar o lema resta verificar a condição sobre o máximo

$$\max_{1 \leq k \leq n} |c_{n,k}| \leq \max_{1 \leq k \leq n} \frac{t^2 \sigma_k^2}{2s_n^2} + \max_{1 \leq k \leq n} |e_{n,k}| \leq \max_{1 \leq k \leq n} \frac{t^2 \sigma_k^2}{2s_n^2} + \max_{1 \leq k \leq n} |e_{n,k}|$$

Como já provamos que os dois termos acima tendem a zero, a prova está terminada. ■

**Exemplo 9.2.7:** Seja  $\{X_n : n \geq 1\}$  uma sequência de variáveis i.i.d. com média 0 e variância 1. Também, seja  $\{Y_n : n \geq 1\}$  uma sequência de variáveis independentes com

$$P(Y_n = \pm n) = \frac{1}{2n^2} \text{ e } P(Y_n = 0) = 1 - \frac{1}{n^2}, n \geq 1.$$

Sendo  $X_n$  e  $Y_n$  independentes para  $n \geq 1$ , temos  $\frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k + Y_k) \xrightarrow{D} N(0, 1)$ , mas a condição de Lindeberg não está satisfeita.

**Solução:** Pelo TCL para variáveis i.i.d., temos que  $\frac{1}{\sqrt{n}} \sum_{k=1}^n X_k \xrightarrow{D} N(0, 1)$ , vamos provar que  $\frac{1}{\sqrt{n}} \sum_{k=1}^n Y_k \xrightarrow{P} 0$ . Deste modo o resultado segue por Slutsky. Pela desigualdade de Markov, temos

$$P\left(\left|\frac{1}{\sqrt{n}} \sum_{k=1}^n Y_k\right| > \epsilon\right) \leq \frac{E\left|\sum_{k=1}^n Y_k\right|}{\epsilon\sqrt{n}} \leq \frac{\sum_{k=1}^n E|Y_k|}{\epsilon\sqrt{n}} = \frac{\sum_{k=1}^n 1/k}{\epsilon\sqrt{n}} \xrightarrow{n \rightarrow \infty} 0,$$

(onde o último limite pode ser visto pelo fato de que usando o teste da integral para séries pode-se provar que  $\frac{1}{\log n} \sum_{k=1}^n 1/k \xrightarrow{n \rightarrow \infty} 1$ ). Logo,  $\frac{1}{\sqrt{n}} \sum_{k=1}^n Y_k \xrightarrow{P} 0$ .

Como  $Var(X_k + Y_k) = Var(X_k) + Var(Y_k) = 2$ , temos que se a condição de Lindeberg fosse satisfeita, teríamos  $\frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k + Y_k) \xrightarrow{D} N(0, 2)$ . Logo, a condição de Lindeberg não é satisfeita, caso contrário teríamos uma contradição. ■

**Corolário 9.2.8: Teorema Central do Limite de Liapunov.** *Sejam  $X_1, X_2, \dots$  variáveis aleatórias independentes tais que  $EX_n = \mu_n$  e  $Var X_n = \sigma_n^2 < \infty$  com pelo menos um  $\sigma_j^2 > 0$ . Seja  $S_n = X_1 + \dots + X_n$  e  $s_n^2 = Var S_n$ . Se existir  $m > 0$  tal que*

$$\frac{1}{s_n^{2+m}} \sum_{k=1}^n E(|X_k - \mu_k|^{2+m}) \rightarrow 0 \text{ quando } n \rightarrow \infty,$$

então,

$$\frac{S_n - ES_n}{s_n} \rightarrow^D N(0, 1).$$

**Prova:** Para provar este teorema, é suficiente verificar que as condições do Teorema de Liapunov implicam as condições do Teorema de Lindeberg. A condição de Lindeberg estabelece uma integral na região  $|x - \mu_k| > \epsilon s_n$ ,  $\epsilon > 0$ . Nessa região, temos que  $\frac{|x - \mu_k|}{\epsilon s_n} > 1$ , o que por sua vez implica  $\frac{|x - \mu_k|^m}{\epsilon^m s_n^m} > 1$ . Desse modo, temos que:

$$\begin{aligned} & \frac{1}{s_n^2} \sum_{k=1}^n \int_{|x - \mu_k| > \epsilon s_n} (x - \mu_k)^2 dF_k(x) \leq \frac{1}{s_n^2} \sum_{k=1}^n \int_{|x - \mu_k| > \epsilon s_n} (x - \mu_k)^2 \frac{|x - \mu_k|^m}{\epsilon^m s_n^m} dF_k(x) \\ & = \frac{1}{\epsilon^m s_n^{2+m}} \sum_{k=1}^n \int_{|x - \mu_k| > \epsilon s_n} |x - \mu_k|^{2+m} dF_k(x) \leq \frac{1}{\epsilon^m s_n^{2+m}} \sum_{k=1}^n \int_{-\infty}^{\infty} |x - \mu_k|^{2+m} dF_k(x) \\ & = \frac{1}{\epsilon^m s_n^{2+m}} \sum_{k=1}^n E|X_k - \mu_k|^{2+m}. \end{aligned}$$

Mas a condição de Liapunov implica que o último termo tende a zero quando  $n \rightarrow \infty$ . Portanto, a condição de Lindeberg está satisfeita. ■

Antes de verificarmos um exemplo do Teorema Central do Limite de Liapunov, vamos considerar o seguinte Lema.

**Lema 9.2.9:** Para  $\lambda > 0$ ,

$$\frac{1}{n^{\lambda+1}} \sum_{k=1}^n k^\lambda \rightarrow \frac{1}{\lambda+1},$$

quando  $n \rightarrow \infty$ , de maneira que  $\sum_{k=1}^n k^\lambda$  é da ordem de  $n^{\lambda+1}$ .

**Prova:** Como  $x^\lambda \leq k^\lambda$  se  $k-1 \leq x \leq k$ , e  $k^\lambda \leq x^\lambda$  se  $k \leq x \leq k+1$ , segue-se que

$$\int_{k-1}^k x^\lambda dx \leq \int_{k-1}^k k^\lambda dx = k^\lambda = \int_k^{k+1} k^\lambda dx \leq \int_k^{k+1} x^\lambda dx,$$

somando-se em  $k$  de 1 até  $n$ , temos

$$\int_0^n x^\lambda dx \leq \sum_{k=1}^n k^\lambda \leq \int_1^{n+1} x^\lambda dx.$$

Logo,

$$\frac{n^{\lambda+1}}{\lambda+1} \leq \sum_{k=1}^n k^\lambda \leq \frac{(n+1)^{\lambda+1} - 1}{\lambda+1} \leq \frac{(n+1)^{\lambda+1}}{\lambda+1},$$

o que é equivalente a

$$\frac{1}{\lambda+1} \leq \frac{1}{n^{\lambda+1}} \sum_{k=1}^n k^\lambda \leq \frac{1}{\lambda+1} \cdot \left(\frac{n+1}{n}\right)^{\lambda+1}.$$

Como  $\left(\frac{n+1}{n}\right)^{\lambda+1} \rightarrow 1$  quando  $n \rightarrow \infty$ , o lema está provado. ■

**Exemplo 9.2.10:** Sejam  $X_1, X_2, \dots$ , independentes,  $X_n \sim U[-n, n]$ . Prove que  $\frac{S_n - ES_n}{s_n} \rightarrow^D N(0, 1)$ .

**Solução:** Vamos verificar a condição de Liapunov para  $\delta = 1$ . Temos

$$E|X_k - \mu_k|^3 = E|X_k|^3 = \frac{1}{2k} \int_{-k}^k |x|^3 dx = \frac{1}{k} \int_0^k x^3 dx = \frac{k^3}{4}.$$

Logo, o Lema anterior implica que  $\sum_{k=1}^n E|X_k - \mu_k|^3$  é da ordem de  $n^4$ . Vamos determinar a ordem de  $s_n^3$ . Como  $\mu_k = EX_k = 0$  e

$$\sigma_k^2 = \text{Var}(X_k) = EX_k^2 = \frac{1}{2k} \int_{-k}^k x^2 dx = \frac{k^2}{3}, \text{ temos}$$

$$s_n^2 = \sum_{k=1}^n \frac{k^2}{3}.$$

Portanto, aplicando o resultado do Lema, temos:

$$\frac{s_n^2}{n^3} \rightarrow \frac{1}{9}.$$

Então,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{s_n^3} \sum_{k=1}^n E|X_k - \mu_k|^3 &= \lim_{n \rightarrow \infty} \left( \frac{n^{9/2}}{s_n^3} \cdot \frac{\sum_{k=1}^n E|X_k - \mu_k|^3}{n^4} \cdot \frac{1}{n^{1/2}} \right) \\ &= 9^{3/2} \cdot \frac{1}{16} \cdot \lim_{n \rightarrow \infty} \frac{1}{n^{1/2}} = 0. \end{aligned}$$

**Exemplo 9.2.11:**

Sejam  $X_n$ ,  $n \geq 1$ , variáveis independentes com

$$P(X_n = \pm 2^n) = 2^{-n-1} \text{ e } P(X_n = \pm 1) = \frac{1}{2}(1 - 2^{-n}), n \geq 1.$$

Verifique que  $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \xrightarrow{D} N(0, 1)$ .

**Solução:** Defina  $Y_n = X_n I_{\{|X_n| \leq n\}}$ . Deste modo,  $P(Y_n = \pm 1) = \frac{1}{2}(1 - 2^{-n})$  e  $P(Y_n = 0) = 2^{-n}$ . Vamos verificar que  $Y_n$  satisfaz a condição de Liapunov para  $m = 1$ . Temos que  $EY_n = 0$ ,  $Var(Y_n) = EY_n^2 = (1 - 2^{-n})$ , e  $E|Y_n|^3 = (1 - 2^{-n}) = Var(Y_n)$ . Logo,  $s_n^2 = \sum_{k=1}^n Var(Y_k) = \sum_{k=1}^n (1 - 2^{-k}) = n - \frac{\frac{1}{2} - (\frac{1}{2})^{n+1}}{\frac{1}{2}}$ . Portanto,

$$\frac{1}{s_n^3} \sum_{k=1}^n E|Y_k|^3 = \frac{1}{s_n^3} \sum_{k=1}^n Var(Y_k) = \frac{1}{s_n} = \frac{1}{n - \frac{\frac{1}{2} - (\frac{1}{2})^{n+1}}{\frac{1}{2}}} \xrightarrow{n \rightarrow \infty} 0.$$

O Teorema Central do Limite de Liapunov implica que,

$$\frac{1}{\sqrt{n - \frac{\frac{1}{2} - (\frac{1}{2})^{n+1}}{\frac{1}{2}}}} \sum_{k=1}^n Y_k \xrightarrow{D} N(0, 1).$$

Portanto,

$$\frac{\sqrt{n}}{\sqrt{n - \frac{\frac{1}{2} - (\frac{1}{2})^{n+1}}{\frac{1}{2}}}} \frac{1}{\sqrt{n}} \sum_{k=1}^n Y_k \xrightarrow{D} N(0, 1).$$

Como  $\frac{\sqrt{n}}{\sqrt{n - \frac{\frac{1}{2} - (\frac{1}{2})^{n+1}}{\frac{1}{2}}}} \xrightarrow{n \rightarrow \infty} 1$ , temos que

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n Y_k \xrightarrow{D} N(0, 1).$$

Seja  $Z_n = X_n - Y_n$ . Então,

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n X_k = \frac{1}{\sqrt{n}} \sum_{k=1}^n Y_k + \frac{1}{\sqrt{n}} \sum_{k=1}^n Z_k.$$

Se conseguirmos provar que  $\frac{1}{\sqrt{n}} \sum_{k=1}^n Z_k \xrightarrow{P} 0$ , então o resultado segue por Slutsky. Mas  $P(Z_n = \pm 2^n) = 2^{-n-1}$  e  $P(Z_n = 0) = 1 - 2^{-n}$ . Como  $P(|Z_n| > \frac{1}{k}) = P(|Z_n| = 2^n) = 2^{-n}$ , temos que

$$\sum_{n=1}^{\infty} P(|Z_n| > \frac{1}{k}) = \sum_{n=1}^{\infty} 2^{-n} < \infty, \forall k \geq 1.$$

Portanto,  $Z_n \rightarrow 0$  cp1, ou seja,  $P(\{w \in \Omega : \lim_{n \rightarrow \infty} Z_n(w) = 0\}) = 1$ . Como

$$\begin{aligned} \lim_{n \rightarrow \infty} Z_n(w) = 0 &\Leftrightarrow \forall \epsilon > 0, \exists N \text{ tal que } |Z_n(w)| < \epsilon, \forall n \geq N \\ &\Rightarrow \exists N \text{ tal que } |Z_n(w)| < 1, \forall n \geq N \\ &\Rightarrow \exists N \text{ tal que } |Z_n(w)| = 0, \forall n \geq N \\ &\Rightarrow \left| \sum_{i=1}^{\infty} Z_i(w) \right| < \infty \\ &\Rightarrow \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i(w) = 0, \end{aligned}$$

temos que  $\{w \in \Omega : \lim_{n \rightarrow \infty} Z_n(w) = 0\} \subseteq \{w \in \Omega : \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i(w) = 0\}$ . Logo,  $P(\{w \in \Omega : \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i(w) = 0\}) = 1$ , o que por sua vez implica que,  $\frac{1}{\sqrt{n}} \sum_{k=1}^n Z_k \xrightarrow{P} 0$ .

### 9.3 Teorema Central do Limite: Caso Multivariado

Concluimos dizendo que o Teorema Central do Limite também pode ser estendido ao caso de vetores aleatórios. Neste caso, tem-se que a distribuição da média amostral centrada converge para uma distribuição normal multivariada. A seguir, nós enunciamos formalmente o teorema sem prová-lo.

**Teorema 9.3.1:** *Seja  $\vec{X}_1, \vec{X}_2, \dots$  uma seqüência de vetores aleatórios  $k$ -dimensionais, independentes e identicamente distribuídos. Suponha que  $\vec{X}_1$  tenha variância finita, e sejam  $\vec{\mu}$  a média e  $\Sigma$  a matriz de covariância de  $\vec{X}_1$ . Seja  $\vec{X}_n$  a média amostral, definida como a média aritmética dos vetores  $\vec{X}_1, \dots, \vec{X}_n$ . Então,*

$$\sqrt{n}(\vec{X}_n - \vec{\mu}) \rightarrow^D N(\vec{0}, \Sigma), \text{ quando } n \rightarrow \infty.$$

### 9.4 Método Delta

O método Delta é um resultado que aumenta significativamente a relevância do Teorema Central do Limite. Antes de enunciarmos o teorema, vamos provar dois lemas. Dizemos que uma seqüência de variáveis aleatórias  $\{Y_n\}$  é *limitada em probabilidade* se para todo  $\epsilon > 0$ , existir  $K$  e  $n_0$  tal que  $P(|Y_n| \leq K) > 1 - \epsilon$  para todo  $n > n_0$ .

**Lema 9.4.1:** Se  $\{Y_n\}$  converge em distribuição para uma variável aleatória com função de distribuição  $H$ , então a seqüência é limitada em probabilidade.

**Prova:** Fixemos  $K_1$  e  $K_2$  pontos de continuidade de  $H$  tal que  $H(K_1) > 1 - \epsilon/4$  e  $H(-K_2) < \epsilon/4$ . Escolhamos  $n_0$  tal que,  $\forall n > n_0$ ,

$$H_n(K_1) > H(K_1) - \epsilon/4 > 1 - \epsilon/2$$

e

$$H_n(-K_2) < H(-K_2) + \epsilon/4 < \epsilon/2.$$

Então,

$$P(-K_2 \leq Y_n \leq K_1) \geq H_n(K_1) - H_n(K_2) > 1 - \epsilon.$$

O resultado está provado se escolhermos  $K = \max(|K_1|, |K_2|)$ . ■

**Lema 9.4.2:** Se  $\{Y_n\}$  é limitada em probabilidade e  $X_n = o(Y_n)$ , então  $X_n \xrightarrow{P} 0$ .

**Prova:** Dados quaisquer  $\epsilon > 0$  e  $\delta > 0$ , precisamos mostrar que existe  $N$  tal que  $P(|X_n| > \epsilon) < \delta$  para todo  $n \geq N$ . Como  $\{Y_n\}$  é limitada em probabilidade, existe  $K$  e  $n_1$  tal que  $P(|Y_n| \leq K) > 1 - \delta$  para todo  $n \geq n_1$ . Como  $X_n = o(Y_n)$ , sabemos que existe  $n_2$  tal que  $\frac{|X_n|}{|Y_n|} < \frac{\epsilon}{K}$  para todo  $n \geq n_2$ . Façamos  $N = \max(n_1, n_2)$ , então para  $n \geq N$ ,  $|X_n| > \epsilon \Rightarrow |Y_n| > K$ . Logo

$$P(|X_n| > \epsilon) \leq P(|Y_n| > K) < \delta.$$

■

**Teorema 9.4.3:** Se  $\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \tau^2)$ , então

$$\sqrt{n}[f(T_n) - f(\theta)] \xrightarrow{D} N(0, \tau^2[f'(\theta)]^2), \quad (9.2)$$

desde que  $f'(\theta)$  exista e não seja zero.

**Prova:** Utilizaremos a versão da série de Taylor em torno de  $T_n = \theta$  que diz que:

$$f(T_n) = f(\theta) + (T_n - \theta)f'(\theta) + o(T_n - \theta),$$

e então

$$\sqrt{n}[f(T_n) - f(\theta)] = \sqrt{n}(T_n - \theta)f'(\theta) + o(\sqrt{n}(T_n - \theta)).$$

O primeiro termo do lado direito converge em distribuição para  $N(0, \tau^2[f'(\theta)]^2)$ . Por outro lado, como  $\sqrt{n}(T_n - \theta)$  converge em distribuição, pelo Lema 9.4.1, temos que  $\sqrt{n}(T_n - \theta)$  é limitada em probabilidade. Então pelo Lema 9.4.2,  $o(\sqrt{n}(T_n - \theta))$  converge para zero em probabilidade. O resultado portanto é uma consequência do Teorema de Slutsky. ■

Este teorema pode parecer uma surpresa, já que se  $X$  é distribuído normalmente, a distribuição de  $f(X)$ , por exemplo,  $1/X$ ,  $\log X$ , ou  $e^X$  não será tipicamente normal. A explicação para este paradoxo aparente pode ser encontrada na prova. Como  $o(T_n - \theta) \xrightarrow{P} 0$ , nós estamos quase certos que quando  $n$  for grande,  $T_n$  é aproximadamente linear, e uma função linear de uma variável normal é também normal. O processo de aproximar a diferença  $f(T_n) - f(\theta)$  pela função linear  $(T_n - \theta)f'(\theta)$  e o limite em (9.2) é chamado de *método delta*.

**Exemplo 9.4.4:** Para estimar  $p^2$ , suponha que temos a escolha entre

- (a)  $n$  ensaios binomiais com probabilidade  $p^2$  de sucesso; ou
- (b)  $n$  ensaios binomiais com probabilidade  $p$  de sucesso.

Sejam  $X$  e  $Y$  o número de sucessos no primeiro e segundo tipo de ensaios, e suponha que como estimadores de  $p^2$  nos dois casos, nós usaríamos  $X/n$  e  $(Y/n)^2$ , respectivamente. Então nós temos:

$$\sqrt{n}\left(\frac{X}{n} - p^2\right) \rightarrow^D N(0, p^2(1 - p^2))$$

e

$$\sqrt{n}\left(\left(\frac{Y}{n}\right)^2 - p^2\right) \rightarrow^D N(0, p(1 - p)4p^2).$$

Então, pelo menos para  $n$  grande,  $X/n$  será mais acurado que  $(Y/n)^2$ , desde que

$$p^2(1 - p^2) < p(1 - p)4p^2.$$

Dividindo ambos os lados por  $p^2(1 - p)$ , podemos ver que

$$\frac{X}{n} \text{ ou } \frac{Y^2}{n^2} \text{ é preferível se } p > 1/3 \text{ ou } p < 1/3, \text{ respectivamente.}$$

O método delta proporciona a base para derivar transformações que estabilizam a variância, ou seja, transformações que levem a uma variância assintótica que é independente do parâmetro. Suponha, por exemplo, que  $X_1, \dots, X_n$  são variáveis Poisson com parâmetro  $\lambda$ . Segue do Teorema Central do Limite que

$$\sqrt{n}(\bar{X} - \lambda) \rightarrow N(0, \lambda).$$

Para problemas de inferência que se referem a  $\lambda$ , é quase sempre inconveniente que  $\lambda$  ocorre não somente na esperança mas também na variância da distribuição limite. É portanto de interesse achar uma função  $f$  para a qual  $\sqrt{n}[f(T_n) - f(\theta)]$  tende em distribuição para  $N(0, c^2)$ , onde  $c^2$  não depende de  $\lambda$ . Em geral, suponha que  $\sqrt{n}(T_n - \theta) \rightarrow^D N(0, \tau^2(\theta))$ . Então, pelo método delta:

$$\sqrt{n}[f(\bar{X}) - f(\lambda)] \rightarrow^D N(0, \tau^2(\theta)(f')^2(\theta)),$$

desde que a derivada de  $f$  exista em  $\theta$  e seja diferente de 0. A distribuição limite do lado direito terá portanto variância constante  $c^2$  se  $f'(\theta) = \frac{c}{\tau(\theta)}$ . A transformação resultante é dita ser estabilizadora de variância.

**Exemplo 9.4.5: Poisson.** No caso de Poisson, temos  $\theta = \lambda$  e  $\tau(\theta) = \sqrt{\lambda}$ . Logo,

$$f'(\lambda) = \frac{c}{\sqrt{\lambda}} \text{ ou } f(\lambda) = 2c\sqrt{\lambda}.$$

Fazendo  $c = 1$ , temos que

$$2\sqrt{n}(\sqrt{\bar{X}} - \sqrt{\lambda}) \rightarrow^D N(0, 1).$$



**Exemplo 9.4.6: Chi-Quadrado.** Seja  $Y_i = X_i^2$ , onde as  $X_i$ 's são i.i.d.  $N(0, \sigma^2)$ . Então,  $EY_i = \sigma^2$  e  $VarY_i = 2\sigma^4$  e pelo Teorema Central do Limite, temos

$$\sqrt{n}(\bar{Y} - \sigma^2) \rightarrow^D N(0, 2\sigma^4),$$

ou seja,  $T_n = \bar{Y}$ ,  $\theta = \sigma^2$ , e  $\tau^2(\theta) = 2\theta^2$ . Logo,

$$f'(\theta) = \frac{c}{\sqrt{2}\theta} \text{ ou } f(\theta) = \frac{c}{\sqrt{2}} \log \theta.$$

Fazendo  $c = 1$ , vemos que

$$\sqrt{\frac{n}{2}} \log\left(\frac{\bar{Y}}{\sigma^2}\right) \rightarrow^D N(0, 1).$$

# Referências Bibliográficas

1. James, B. (1981), "Probabilidade: um curso em nível intermediário" - Projeto Euclides
2. Magalhães, Marcos M. (2006), "Probabilidade e Variáveis Aleatórias", 2a. edição, edusp.
3. Lima, E. (1976), "Curso de Análise", vol.1 - Projeto Euclides
4. Resnick, S. I., "A Probability Path", Birkhauser, 2005.
5. Halpern, Joseph Y. (2003), "Reasoning About Uncertainty", The MIT press.
6. Davenport Jr., W. (1970), "Probability and Random Processes", McGraw-Hill Book Company Inc.
7. deFinetti, B. (1972), "Probability, Induction, and Statistics", New York: Willey.
8. Fine, T. (2006), "Probability and Probabilistic Reasoning for Electrical Engineering", Prentice Hall.