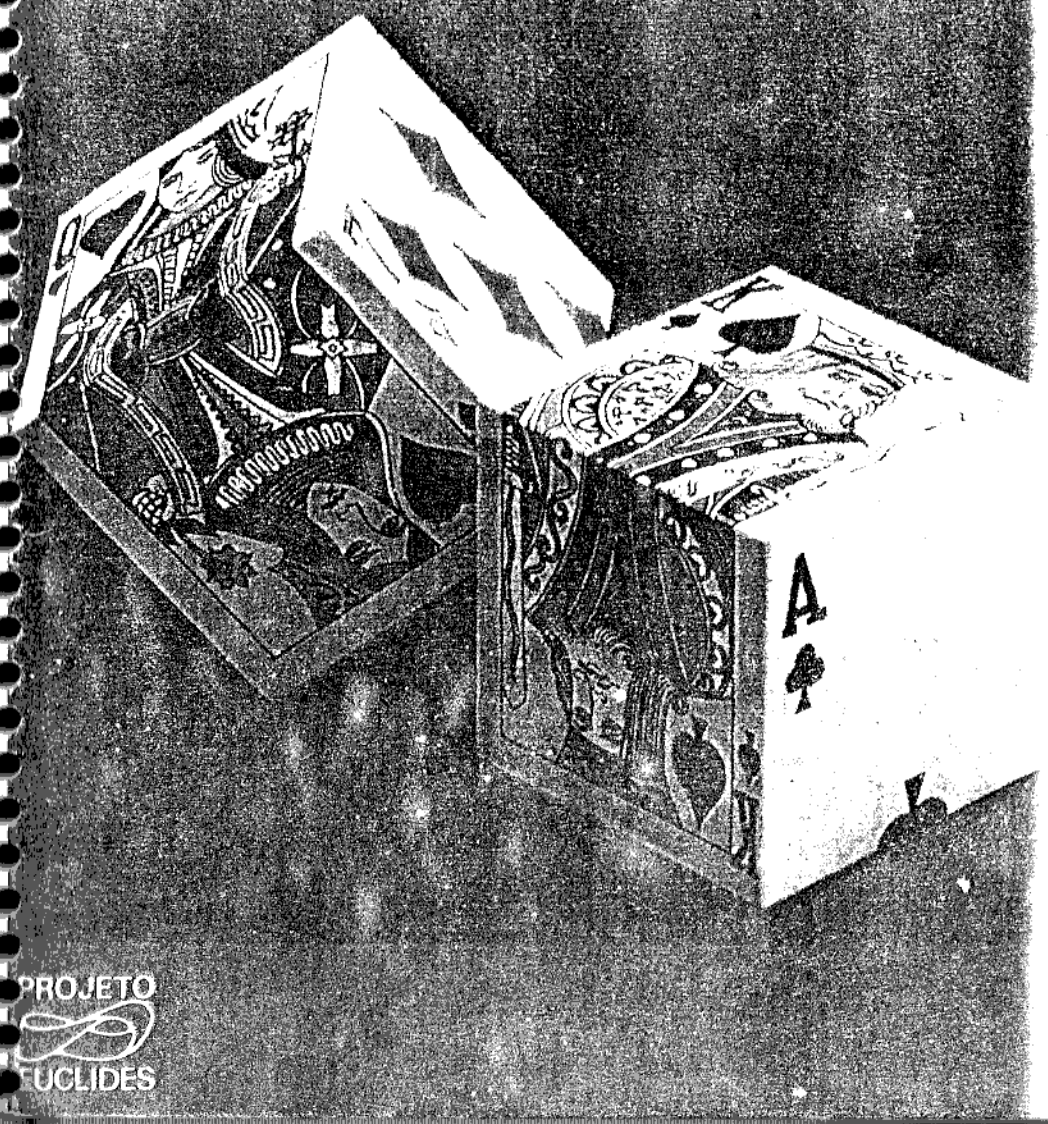
barry r. james

probabilidade: um curso em nível intermediário



PROJETO EUCLIDES

barry r. james

probabilidade: um curso em nível intermediário

DONACIÓN
DE LA
A F A M U S M
Rio de Janeiro



impa
Instituto de Matemática Pura e Aplicada — CNPq

Copyright ©, by Barry R. James
Direitos reservados, 1981, por Conselho Nacional
de Desenvolvimento Científico e Tecnológico, CNPq,
Av. W-3 Norte, Brasília, DF

Impresso no Brasil/*Printed in Brazil*

Capa: Gian Calvi Criação Visual Ltda.

Projeto Euclides

Comissão Editorial

Cesar Camacho Manco
Chaim Samuel Hönig
Djairo Guedes de Figueiredo
Elon Lages Lima
Guilherme de La Penha
Imre Simon
Jacob Palis Junior
Lindolpho de Carvalho Dias
Manfredo Perdigão do Carmo
Pedro Jesus Fernandez

Títulos já publicados:

1. Curso de Análise, vol. 1, Elon Lages Lima
2. Medida e Integração, Pedro Jesus Fernandez
3. Aplicações da Topologia à Análise, Chaim Samuel Hönig
4. Espaços Métricos, Elon Lages Lima
5. Análise de Fourier e Equações Diferenciais Parciais, Djairo Guedes de Figueiredo
6. Introdução aos Sistemas Dinâmicos, Jacob Palis Junior e Welington C. de Melo
7. Introdução à Álgebra, Adilson Gonçalves
8. Aspectos Teóricos da Computação, Cláudio L. Lucchesi, Imre Simon, Istvan Simon, Janos Simon e Tomasz Kowaltowski
9. Teoria Geométrica das Folheações, Alcides Lins Neto e Cesar Camacho Manco
10. Geometria Riemanniana, Manfredo P. do Carmo
11. Lições de Equações Diferenciais Ordinárias, Jorge Sotomayor
12. Probabilidade: Um Curso em Nível Intermediário, Barry R. James
13. Curso de Análise, vol. 2, Elon Lages Lima

Publicação feita com auxílio da FINEP — Financiadora de Estudos e Projetos.

Composição e arte:
AM Produções Gráficas Ltda.

Impresso por:
Gráfica Editora Hamburg Ltda.
Rua Apeninos, 294 — São Paulo — SP

Distribuído por:
Livros Técnicos e Científicos Editora S.A.
Avenida Venezuela, 163
20.200 — Rio de Janeiro, RJ — Brasil



CONTEÚDO

PREFÁCIO VII

CAPÍTULO 1 DEFINIÇÕES BÁSICAS

§ 1.1 Modêlo matemático para um experimento (modêlo probabilístico) .. 1
§ 1.2 Probabilidade condicional 14
§ 1.3 Independência 18
Exercícios 27

CAPÍTULO 2 VARIÁVEIS ALEATÓRIAS

§ 2.1 Variáveis aleatórias e funções de distribuição 35
§ 2.2 Tipos de variáveis aleatórias 41
§ 2.3 A distribuição de uma variável aleatória 47
§ 2.4 Vetores aleatórios 55
§ 2.5 Independência 59
§ 2.6 Distribuições de funções de variáveis e vetores aleatórios 67
§ 2.7 O método do jacobiano 74
§ 2.8 Observações adicionais — variáveis e vetores aleatórios 84
Exercícios 86

CAPÍTULO 3 ESPERANÇA MATEMÁTICA

§ 3.1 Preliminares: a integral de Stieltjes 97
§ 3.2 Esperança 104
§ 3.3 Propriedades da esperança 113
§ 3.4 Esperanças de funções de variáveis aleatórias 117
§ 3.5 Momentos 120
§ 3.6 Esperanças de funções de vetores aleatórios 126
§ 3.7 Teoremas de convergência 133
Exercícios 137

CAPÍTULO 4 DISTRIBUIÇÃO E ESPERANÇA CONDICIONAIS

§ 4.1 Distribuição condicional de X dada Y discreta 144
§ 4.2 Distribuição condicional de X dada Y: caso geral 154
§ 4.3 Definições formais e teoremas de existência 162
§ 4.4 Exemplos 166
§ 4.5 Esperança condicional 174

Exercícios 180

912323

CAPÍTULO 5 A LEI DOS GRANDES NÚMEROS

§ 5.1 Introdução às Leis Fraca e Forte dos Grandes Números 189
§ 5.2 Seqüências de eventos e o Lema de Bovel-Cantelli..... 197
§ 5.3 A Lei Forte 202
Exercícios 215

CAPÍTULO 6 FUNÇÕES CARACTERÍSTICAS E CONVERGÊNCIA

§ 6.1 Funções características 221
§ 6.2 Convergência em distribuição..... 231
§ 6.3 Função característica de um vetor aleatório..... 238
§ 6.4 Observações e complementos 244
Exercícios 253

CAPÍTULO 7 O TEOREMA CENTRAL DO LIMITE

§ 7.1 O Teorema Central do Limite para seqüências de variáveis aleatórios 261
§ 7.2 A distribuição normal multivariada 273
§ 7.3 O Teorema Central do Limite — caso multivariado..... 277
Exercícios 280
REFERÊNCIAS 286
ÍNDICE DE NOTAÇÕES 287
ÍNDICE ALFABÉTICO 288

DONACIÓN
DE LA
AFAMUSM
Rio de Janeiro

PREFÁCIO

Le calcul des probabilités n'est au fond que le bon sens réduit au calcul.
— Laplace

A. *Ao leitor.*

O presente volume é oferecido para uso em cursos de Probabilidade em nível "intermediário", que devia ser entendido como o nível entre um curso elementar de Introdução à Probabilidade e um curso mais avançado que trata de Probabilidade com base na Teoria da Medida e Integração. O material aqui apresentado tem sido usado várias vezes em disciplinas de Probabilidade em nível de início de Mestrado, mas também poderá ser usado em nível de graduação se alguns assuntos de natureza mais técnica forem pulados. (Tais assuntos são indicados no texto pelo aviso que podem ser omitidas "em uma primeira leitura".)

Este livro não pretende ser introdutório, embora eventualmente possa ser usado como tal. Por exemplo, um assunto muito comum nos livros de Introdução à Probabilidade, e que não será considerado aqui, é a teoria combinatória. É preferível, mas não necessário, que o leitor já tenha alguma noção das distribuições discretas clássicas baseadas na contagem de permutações, combinações, etc., tais como a binomial, hipergeométrica e multinomial. Uma boa discussão dessas distribuições pode ser encontrada em Breiman ([6], p. 19-31, em inglês) ou Feller ([8], traduzido para o Português).

Para se poder acompanhar o texto deste livro, os reais prerequisitos são um curso de Cálculo Diferencial e Integral e alguma familiaridade com os conceitos básicos de conjuntos e funções. É bom que o leitor saiba lidar com uniões, interseções e complementares de conjuntos, e conheça os conceitos de imagem inversa de um Conjunto por uma função, supremo e ínfimo de um conjunto e limite de uma seqüência de números reais. Em caso de uma eventual lacuna na formação do leitor nesta área de Análise, recomenda-se uma consulta ao livro *Curso de Análise* (Volume 1), de Elon Lima [14]. Lá se pode achar, também, definições e discussões de outros

conceitos analíticos que aparecem de vez em quando aqui, tais como o limite superior de uma seqüência de números reais, ou a integral de Riemann. Além disso, integrais múltiplas entram em jogo no Capítulo 2, e será bom se até o §2.7 o leitor já tiver ouvido falar na matriz jacobiana e derivadas parciais, cuja definição poderá ser encontrada em qualquer bom texto de Cálculo Avançado. Finalmente, nos parágrafos 7.2 e 7.3 usam-se alguns conceitos elementares de Álgebra Linear; em particular, é preciso conhecer as regras de multiplicação de matrizes e vetores.

Em caso de qualquer dúvida que surgir sobre o significado de um símbolo usado no texto, o leitor deverá consultar a lista de símbolos e notações que aparece no final do livro. Procurei fazer uma lista completa, mas se persistir a dúvida, é recomendável consultar os livros, citados acima, de Lima [14] ou Feller [8]. (A única diferença importante entre a notação de [14] e a presente, é que aqui o complementar de um conjunto A é representado por A^c.)

Como é sempre o caso com livros de Matemática deste nível, as listas de exercícios são parte importantíssima do livro. Há grande número de exercícios puramente computacionais, para ajudar o leitor a se treinar no cálculo básico de probabilidades. Há muitos exercícios que estendem e desenvolvem idéias abordadas no texto, e alguns outros que introduzem idéias não consideradas ali. Os exercícios estão arranjados por seção, e de um modo geral os problemas foram colocados na ordem que achei mais conveniente para sua resolução posterior. Mas esta regra não foi seguida com muita fidelidade, e recomendo que o leitor ao menos leia todos os exercícios quando chegar ao final de cada seção, resolvendo em seguida os que ele achar mais interessantes. Sugestões foram incluídas com os exercícios que considerei mais difíceis, ou com os que fogem ao nível da seção em que se encontram.

B. *Ao professor.*

O material no livro é mais que o suficiente para um curso de semestre. No IMPA, o curso de Probabilidade é de quatro meses e, cada vez que foi usada a apostila precursora deste livro, sentiu-se a necessidade de correr no final para abordar todos os assuntos aqui presentes. Capítulos 1, 2 e 3 formam a base do livro e podem ser feitos em mais ou menos a metade de um período. Muitas demonstrações nos Capítulos 5 e 6 são técnicas e poderão ser omitidas. Os parágrafos 6.3, 7.2 e 7.3 constituem um assunto especial — a parte da Estatística Multivariada — que poderá ser omitido ou até condensado para apresentação em uma aula no final do curso.

O Capítulo 4 é muito mais extenso do que costuma ser o tratamento de distribuição e esperança condicionais em textos da Teoria da Probabilidade. Em caso de tempo exíguo, esse capítulo poderá ser abreviado, pois consiste, na maior parte, em exemplos. Mas será bom se o aluno terminar o capítulo com, pelo menos, uma boa idéia intuitiva do princípio da preservação de chances relativas, pois assim terá condições de lidar no futuro com aplicações de condicionamento, que é da maior importância nas áreas de Confiabilidade, Processos Estocásticos e Estatística (tanto Bayesiana quanto Não-Bayesiana).

Quanto aos prerequisitos, quero enfatisar que não é necessário que o aluno tenha conhecimentos da Teoria da Medida e a integral de Lebesgue para poder acompanhar o livro. A esperança matemática de uma variável aleatória é tratada utilizando-se da integral de Stieltjes, cuja definição e propriedades são abordadas no §3.1. A integral de Lebesgue é mencionada apenas em observações. Distribuição e esperança condicionais são tratadas em plena generalidade, ou seja, sem a restrição aos casos simples que costumam ser tratados em livros de Introdução à Probabilidade. Mas o enfoque adotado é o que chega às definições partindo da consideração de limites de probabilidades condicionais. A relação entre este tratamento mais intuitivo e a abordagem de Probabilidade Avançada, que utiliza o Teorema de Radon-Nikodym, é formalizada no §4.3 (esta formalização poderá ser omitida, como é indicado no texto).

C. *Uma reta sobre terminologia.*

Durante o processo de escrever este livro, esbarrei frequentemente no problema de terminologia. Como o português não é minha língua natal, procurei ser fiel ao idioma ao máximo possível. Mas descobri vários conceitos que, embora representados por uma palavra ou frase em outras línguas, possuiam mais de uma versão aqui no Brasil. Era, por exemplo, o caso de "esperança" e "expectância" para indicar a média de uma variável aleatória. Precisei, então, fazer uma escolha. Gostaria agora de explicar minha escolha em três casos.

(1) Parece-me que a palavra "expectância" surgiu em português como tradução do inglês "expectation". Porém, a palavra usada em espanhol é "esperanza" e em francês é "espérance", esta última sendo usada desde, pelo menos, o século 18. Portanto, optei por "esperança".

(2) Resolvi usar o adjetivo "condicional", em vários lugares em que também se usa "condicionada", a saber, em "probabilidade condicional", "distribuição condicional" e "esperança condicional".

Escolhi assim porque quando ocorre um evento, são ocorrências de outros eventos e valores de variáveis aleatórias que estão sendo diretamente afetados (i.e., condicionados), decorrendo daí as modificações nas probabilidades, distribuições e esperanças através das respectivas definições. Portanto, prefiro "condicional", mas talvez seja mais uma questão de gosto.

(3) Primeiro, consideremos um exemplo do problema: na expressão "leite de vaca puro", é óbvio, pela concordância, que não estamos opinando sobre a vaca. Mas em espanhol, onde "leche" é feminina, a expressão seria altamente ambígua. Temos o mesmo problema em português com a expressão "Teorema do Limite Central", atualmente muito usado no Brasil. Assim como está, esta frase dá a nítida impressão de que é o limite que seja central, o que na realidade não faz sentido. Por isso, optei pelo uso de "Teorema Central do Limite", para afastar qualquer dúvida sobre o que seja central.

É interessante que a frase em inglês, "Central Limit Theorem", é também altamente ambígua, e acho que isto explica a tradução de uso corrente no Brasil. Ocorre que a origem da expressão foi, aparentemente, o alemão e não o inglês. De fato, a expressão é frequentemente atribuída a Pólya, que usou a frase "der zentrale Grenzwertsatz", i.e., o "central" refere-se ao "teorema do limite".

D. *Agradecimentos.*

Este livro surgiu de notas de aulas usadas no curso básico de Probabilidade do programa de Mestrado em Matemática Aplicada do IMPA. Um curso mais ou menos parecido com o presente livro foi dado pela primeira vez em 1976, e desde então as notas e a apostila resultante foram sendo modificadas de ano em ano, até chegarem a seu estado atual, que é este livro. Dei o curso três vezes neste período, e a apostila foi usada também por meus colegas Pedro J. Fernández e Ricardo Frischtak, no IMPA, e Annibal Parracho Sant'ana, no Instituto de Matemática da UFRJ. Sou muito grato a eles pelos seus comentários e sugestões.

As versões semifinais do manuscrito foram cuidadosamente lidas por Maria Eulália Vares e Sergio Wechsler. Fizeram muitas sugestões para melhorar a apresentação do texto e contribuíram ao livro com vários exercícios. Além disso, conseguiram corrigir grande número de erros de português, ao mesmo tempo ajeitando a minha versão do idioma. A eles, minha profunda gratidão. A propósito, devo dizer aqui que os erros de português ainda restantes são unica e exclusivamente da responsabilidade do autor.

Este livro não poderia ter acontecido sem o apoio e incentivo constantes dos meus colegas do IMPA. Estou especialmente obrigado a Djalma Pessoa, Elon Lima e Ruben Klein, não somente pelo apoio recebido, mas também pela insistência deles de que o livro saísse o mais rápido possível.

Grande contribuição foi dada pelas várias turmas de alunos do IMPA, através de suas perguntas, dúvidas, observações e, de modo geral, o "feedback" que deram em aula. Eles até emprestaram suas notas e cadernos para ajudar a escrever a apostila em português. Não dá para citar todos os nomes aqui, mas a todos os alunos — e aos assistentes (monitores) — gostaria de transmitir os meus agradecimentos especiais.

Finalmente, chego à peça chave do livro: a minha esposa e colega, Kang. Descobri nestes últimos meses que o processo de escrever um livro envolve muito mais tempo e trabalho que eu pensava. E foi ela que teve a paciência de agüentar tudo isso, inclusive abrindo mão de qualquer descanso durante os feriados de dezembro para me ajudar. Ela leu vários capítulos com cuidado, fez muitas sugestões boas, e passou a limpo a maior parte da apostila. Pela ajuda, pela paciência e pelo sacrifício, minha gratidão para com ela não tem limite.

Rio de Janeiro, maio de 1981.
BRJ



CAPÍTULO I

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
FACULTAD DE CIENCIAS MATEMÁTICAS
BIBLIOTECA

DEFINIÇÕES BÁSICAS

§1.1 Modelo matemático para um experimento (modelo probabilístico)

Suponhamos que um experimento seja realizado sob certas condições fixas. Seja Ω o conjunto de resultados possíveis, onde por "resultado possível" entende-se resultado elementar e indivisível do experimento. Ω será chamado *espaço amostral* do experimento. Por exemplo:

EXPERIMENTO 1. Jogar um dado equilibrado e observar o número da face superior. É claro que $\Omega = \{1, 2, 3, 4, 5, 6\}$, pois esses 6 resultados são os únicos possíveis. ("Número par", por exemplo, não é resultado elementar, pois consiste dos 3 resultados 2, 4, 6.)

Às vezes, o conjunto de resultados possíveis não é tão fácil de ser definido:

EXPERIMENTO 2. Selecionar ao acaso um habitante do Rio de Janeiro e medir sua altura em metros. Quais os resultados possíveis deste experimento? Números reais entre o e ? Supondo que não exista uma altura máxima, talvez seja razoável fazer $\Omega = (0, \infty)$. Mas é evidente que esse conjunto contém resultados impossíveis, tais como um milhão ou um bilhão de metros. Outros candidatos para Ω seriam, por exemplo, os intervalos limitados $(0, 3)$ e $[1/10, 3]$. Os dois intervalos contêm, aparentemente, todos os resultados possíveis do experimento. Esta propriedade já é suficiente para nossos propósitos, e podemos escolher qualquer desses intervalos (incluindo $(0, \infty)$) para o espaço amostral. De fato, a própria reta $\mathbb{R}$,

embora contenha muitíssimos resultados impossíveis, pode representar uma escolha muito conveniente para Ω, principalmente no caso em que desejemos atribuir uma distribuição normal ao resultado (veja o Capítulo 2).

O importante, então, é que Ω contenha todo resultado possível; por isso vamos supor:

(*i*) a todo resultado possível corresponde um, e somente um, ponto $\omega \in \Omega$; e

(*ii*) resultados distintos correspondem a pontos distintos em Ω, i. e., ω não pode representar mais de um resultado.

Ora, quando se realiza um experimento há certos *eventos* que ocorrem ou não. Por exemplo, no experimento 1 (jogar um dado e observar o resultado) alguns eventos são:

A = "observa-se um número par"
B = "observa-se o número 2"
C = "observa-se um número ≥ 4".

Notemos que cada um desses eventos pode ser identificado a um subconjunto de Ω, a saber: $A = \{2, 4, 6\}$, $B = \{2\}$, $C = \{4, 5, 6\}$. Esta identificação de eventos e subconjuntos costuma ser realizável no caso de um experimento qualquer.

Com efeito, seja Ω o espaço amostral e A um evento associado ao experimento, i. e., um evento que seguramente irá ou não ocorrer sempre que for realizado o experimento. Para fixarmos idéias, suponhamos que Ω consista exatamente nos resultados possíveis do experimento, de modo que Ω não contenha resultados impossíveis. Suponhamos, então, que ω seja o resultado do experimento. Se A ocorre, dizemos que ω é *favorável* a A. Se A não ocorre, dizemos que ω não é favorável a A (ou ainda, que ω é favorável ao evento "não A"). Identificaremos o evento A e o subconjunto de Ω que contém todo ω favorável a A. Por exemplo, consideremos o

EXPERIMENTO 3. Escolher, ao acaso, um ponto do círculo (disco) de raio 1 centrado na origem. Então

$$\Omega = \text{círculo unitário} = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}.$$

Aqui estão alguns eventos para esse experimento:

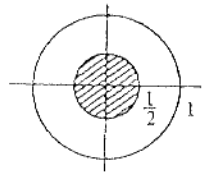
A = "distância entre o ponto escolhido e a origem é $\leq \frac{1}{2}$";

B = "distância entre o ponto escolhido e a origem é ≥ 15";

C = "1.ª coordenada do ponto escolhido é maior que a 2.ª".

Se $\omega = (x, y)$ for o resultado do experimento, ω será favorável ao evento A se, e somente se, $x^2 + y^2 \leq \frac{1}{4}$, e será favorável a C se, e só se, $x > y$. Nenhum resultado será favorável a B. Logo temos:

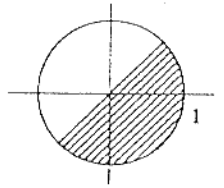
$$A = \left\{(x, y) \in \Omega : \sqrt{x^2 + y^2} \leq \frac{1}{2}\right\} =$$



;

$B = \varnothing$ = conjunto vazio;

$$C = \{(x, y)) \in \Omega : x > y\} =$$



.

Então, todo evento associado a este experimento pode ser identificado a um subconjunto do espaço amostral Ω. Reciprocamente, se A for um subconjunto qualquer de Ω, i. e., $A \subset \Omega$, então será conveniente identificar A e o evento "resultado do experimento pertence a A".

Chegamos à seguinte definição, que adotaremos no caso geral, inclusive nos casos em que utilizamos um espaço amostral maior que o estritamente necessário:

DEFINIÇÃO 1.1. Seja Ω o espaço amostral do experimento. Todo subconjunto $A \subset \Omega$ será chamado *evento*. Ω é o evento *certo*, $\varnothing$ o evento *impossível*. Se $\omega \in \Omega$, o evento $\{\omega\}$ é dito *elementar* (ou simples).

Observação. Às vezes, identificamos o evento $\{\omega\}$ (= "resultado do experimento é ω") e o ponto ω. Como, por exemplo, na indicação $P(\omega) = P(\{\omega\})$.

É bom saber traduzir a notação de conjuntos para a linguagem de eventos: $A \cup B$ é o evento "A ou B", $A \cap B =$ "A e B", $A^c =$ não A" (i. e., ocorre o evento A^c se, e somente se, não ocorre o evento A); $A \subset B$ significa: a ocorrência do evento A *implica* a ocorrência do evento B; $A \cap B = \varnothing$ significa: $A \in B$ são eventos *mutuamente exclusivos* ou *incompatíveis*. (Para um exercício sobre essa linguagem, veja o exercício 1.)

A esta altura é razoável perguntar: a que eventos vamos atribuir probabilidade? Consideremos novamente o experimento 1, e seja A um evento, i. e., $A \subset \Omega$. É evidente que podemos atribuir probabilidade a A pois estamos jogando um dado equilibrado. De fato, definimos:

$$P(A) = \frac{\#A}{6} = \frac{\text{número de resultados favoráveis a } A}{\text{número de resultados possíveis}}.$$

Esta é a definição clássica de probabilidade quando Ω é finito, e baseia-se no conceito de resultados equiprováveis, ou melhor, no princípio da indiferença (estamos "indiferentes" diante dos resultados 1, 2, 3, 4, 5, 6; logo definimos $P(i) = \frac{1}{6}\ \forall\ i \in \Omega$). Então para o experimento 1, nossa resposta é que todo evento terá uma probabilidade.

Consideremos agora o experimento 3 (escolher um ponto ao acaso no círculo unitário). Aqui "ao acaso" será interpretado assim: dois eventos têm a mesma probabilidade se, e somente se, eles têm a mesma área. (Essa probabilidade é chamada *geométrica*. Veja Gnedenko [11], §6.) Essa interpretação conduz à definição, para $A \subset \Omega$

$$P(A) = \frac{\text{área } A}{\text{área } \Omega} = \frac{\text{área } A}{\pi},$$

se área A estiver bem definida. Acontece que nem todo subconjunto de Ω tem uma área bem definida, i. e., nem todo evento tem uma probabilidade. (De fato, segundo um teorema profundo da Teoria da Medida: não se pode definir $P(A)$ para todo $A \subset \Omega$ de modo que $P(A) =$ $=$ (área A)/π para todo A cuja área está bem definida; i. e., não podemos estender a definição de $P(A)$ para todo evento de modo a satisfazer os axiomas usuais, que serão vistos mais adiante. A prova disto depende da Hipótese do Contínuo.) Vamos, então, atribuir probabilidade somente aos eventos cuja área estiver bem definida. Tais eventos serão chamados eventos aleatórios.

DEFINIÇÃO 1.2. Um evento A ao qual atribuímos uma probabilidade será chamado *evento aleatório*.

Na prática, o fato de não podermos atribuir probabilidade a todo evento não nos causará problemas. No experimento 3, por exemplo, obviamente é suficiente restringir a nossa atenção aos conjuntos com área bem definida, pois os conjuntos sem área definida nunca surgem na prática (de fato, é impossível visualizar um tal conjunto).

Vamos supor, contudo, que a classe dos eventos aleatórios possua certas propriedades básicos e intuitivas, que serão essenciais para o desenvolvimento posterior da teoria e do cálculo de probabilidades. Indicando com $\mathscr{A}$ a classe dos eventos aleatórios, vamos estipular as seguintes propriedades para $\mathscr{A}$:

A1. $\Omega \in \mathscr{A}$ (definiremos $P(\Omega) = 1$).

A2. *Se* $A \in \mathscr{A}$, *então* $A^c \in \mathscr{A}$ (é evidente que definiremos

$$P(A^c) = 1 - P(A)).$$

A3. *Se* $A \in \mathscr{A}$ *e* $B \in \mathscr{A}$, *então* $A \cup B \in \mathscr{A}$ (i. e., se atribuirmos uma probabilidade a A e outra a B, então atribuiremos uma probabilidade a "A ou B".)

Em outras palavras, vamos supor que $\mathscr{A}$ seja uma álgebra de eventos:

DEFINIÇÃO 1.3. Seja Ω um conjunto não-vazio. Uma classe $\mathscr{A}$ de subconjuntos de Ω satisfazendo A1, A2 e A3 é chamada *álgebra* de subconjuntos de Ω.

PROPOSIÇÃO 1.1. *Seja* $\mathscr{A}$ *uma álgebra de subconjuntos de* Ω. *Então valem as seguintes propriedades*:

A4. $\varnothing \in \mathscr{A}$ *e*

A5. $\forall\, n,\ \forall\, A_1, \ldots, A_n \in \mathscr{A}$, *temos* $\bigcup_{i=1}^{n} A_i \in \mathscr{A}$ *e* $\bigcap_{i=1}^{n} A_i \in \mathscr{A}$.

Esta proposição diz que uma álgebra é fechada para um número *finito* de aplicações das operações $\cup$, $\cap$, e ^{c}.

Prova. A1 e A2 implicam A4. Para A5, temos $A3 \Rightarrow A_1 \cup A_2 \in \mathscr{A} \Rightarrow$

$$\Rightarrow (A_1 \cup A_2) \cup A_3 \in \mathscr{A} \Rightarrow \ldots \Rightarrow \bigcup_{i=1}^{n} A_i \in \mathscr{A},$$ por indução.

Agora basta observar que

$$\bigcap_{i=1}^{n} A_i = \left(\bigcup_{i=1}^{n} A_i^c\right)^c$$

e aplicar sucessivamente $A2$, a parte já provada de $A5$ e, novamente, $A2$. □

(*Exercício.* Demonstre que $\mathscr{A}$ é também fechada para diferenças, i.e., se $A \in \mathscr{A}$ e $B \in \mathscr{A}$, então $A - B \in \mathscr{A}$, onde $A - B = A \cap B^c$.)

Sem perda de generalidade (veja a segunda observação a seguir), vamos supor que a classe dos eventos aleatórios também satisfaça:

$A3'$. *Se* $A_n \in \mathscr{A}$ *para* $n = 1, 2, 3, \ldots$, *então* $\bigcup_{n=1}^{\infty} A_n \in \mathscr{A}$.

DEFINIÇÃO 1.4. Uma classe $\mathscr{A}$ de subconjuntos de um conjunto não-vazio Ω satisfazendo A1, A2 e A3' é chamada *σ-álgebra* de subconjuntos de Ω.

Observações. (1) Uma σ-álgebra é sempre uma álgebra, pois $A3$ é conseqüência de $A3'$, já que $A \cup B = A \cup B \cup B \cup B \ldots \in \mathscr{A}$ se $\mathscr{A}$ é σ-álgebra.

(2) Podemos supor, sem perda de generalidade, que $\mathscr{A}$ é uma σ-álgebra em vez de álgebra, pelo Teorema da Extensão de Carathéodory. Este teorema da Teoria da Medida garante que uma probabilidade definida em uma álgebra, e de acordo com os axiomas usuais, pode ser estendida de uma única maneira para a σ-álgebra gerada pela álgebra. (Para entender o significado de "σ-álgebra gerada pela álgebra", veja o exercício 6.)

(3) Em inglês, usa-se às vezes o termo "field" (corpo) no lugar de "álgebra" e "σ-field" no lugar de "σ-álgebra". Para o tradutor de Gredenko (veja [11]), σ-álgebra é "Borel field". Em francês, é "tribu".

PROPOSIÇÃO 1.2. *Seja* $\mathscr{A}$ *uma* σ-*álgebra de subconjuntos de* Ω. *Se* $A_1, A_2 \ldots \in \mathscr{A}$, *então* $\bigcap_{n=1}^{\infty} A_n \in \mathscr{A}$.

Prova. $\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c\right)^c$. □

Podemos dizer, então, que uma σ-álgebra é fechada para um número *enumerável* de aplicações das operações $\cup$, $\cap$ e ^{c}.

EXEMPLOS (de σ-álgebra de eventos aleatórios).

(1) *Caso discreto.* Se Ω for finito ou enumerável, então $\mathscr{A}$ será (usualmente) a σ-álgebra de todas as partes de Ω, i.e., $\mathscr{A} = \mathscr{P}(\Omega)$. Por exemplo, no experimento 1, onde $\Omega = \{1, 2, 3, 4, 5, 6\}$, temos $\mathscr{A} = \mathscr{P}(\Omega) = \{\varnothing, \{1\}, \{2\}, \ldots \{6\}, \{1, 2\}, \ldots, \Omega\}$. A classe $\mathscr{A}$ tem $2^6 = 64$ elementos, de modo que há 64 eventos aleatórios associados a este experimento. No caso finito geral, se Ω tem n elementos, $\mathscr{P}(\Omega)$ tem 2^n. O leitor deveria se convencer do fato de $\mathscr{P}(\Omega)$ ser uma σ-álgebra, verificando $A1$, $A2$ e $A3'$.

(2) *Caso contínuo.* Consideremos o

EXPERIMENTO 4. Selecionar, ao acaso, um ponto do intervalo $[0, 1]$. Aqui, $\Omega = [0, 1]$ e $\mathscr{A}$ = todos os subconjuntos cujo comprimento esteja bem definido. Quem são esses conjuntos? Consideremos primeiro uniões finitas de intervalos e seja $\mathscr{A}_0 = \{A \subset [0, 1]: A$ é união finita de intervalos$\}$. Notemos que $\mathscr{A}_0$ é álgebra, pois $\Omega \in \mathscr{A}_0$, se $A \in \mathscr{A}_0$ então A^c também é união finita de intervalos, A3 é trivial. O conjunto vazio $\varnothing$ e o evento elementar $\{\omega\}$, onde $\omega \in [0, 1]$, serão interpretados como intervalos degenerados de comprimento 0, portanto serão elementos de $\mathscr{A}_0$. Mas $\mathscr{A}_0$ não é σ-álgebra, pois não contém toda união *enumerável* de intervalos, como teria que conter se fosse σ-álgebra. Por exemplo, o evento

$$A = \left(0, \frac{1}{2}\right) \cup \left(\frac{1}{2}, \frac{3}{4}\right) \cup \left(\frac{3}{4}, \frac{7}{8}\right) \cup \ldots \cup \left(1 - \frac{1}{2^n}, 1 - \frac{1}{2^{n+1}}\right) \cup \ldots$$

é união enumerável de intervalos mas óbviamente não é união finita. Então a classe de eventos aleatórias será maior e mais complicada que $\mathscr{A}_0$. Aliás, é bem claro que A é um evento aleatório, pois seu comprimento pode ser definido somando-se os comprimentos dos intervalos componentes, fazendo com que comprimento $(A) = 1 = P(A)$.

Outro evento que não pertence a $\mathscr{A}_0$, mas cuja probabilidade pode ser definida, é o conjunto dos racionais, $\{r \in [0, 1]: r$ racional$\}$. Qual a probabilidade de selecionar um número racional? É claro que 0 é o único candidato para o comprimento do conjunto dos racionais, pelo seguinte argumento:

Sejam $r_1, r_2, \ldots$ os racionais em $[0, 1]$, e seja A_n o intervalo aberto de centro r_n e comprimento $\varepsilon/2^n$, onde $\varepsilon > 0$. Então,

$$B = \{r_n : n = 1, 2, \ldots\} \subset \bigcup_{n=1}^{\infty} A_n,$$

e o comprimento do conjunto dos racionais satisfaz

$$\text{comprimento } (B) \leq \text{comprimento} \left(\bigcup_{n=1}^{\infty} A_n \right)$$
$$\leq \sum_{n=1}^{\infty} \text{comprimento } (A_n) = \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \varepsilon.$$

Como o comprimento de B é menor que ou igual a ε, para todo $\varepsilon > 0$, ele é igual a zero. (Um argumento alternativo: como B é união enumerável dos intervalos degenerados disjuntos $\{r_n\}$, todos de comprimento zero, o comprimento de B é a soma dos comprimentos dos componentes, ou seja, comprimento $(B) = 0$.)

Os eventos A e B acima são uniões enumeráveis de intervalos e, portanto, pertencem a toda σ-álgebra que contém os intervalos. Neste livro, nossa σ-álgebra de eventos aleatórios para o experimento 4 será a σ-álgebra gerada pelos intervalos, i. e., a menor σ-álgebra que contém todos os intervalos (veja o exercício 6). Esta σ-álgebra é chamada *σ-álgebra de Borel* em $[0, 1]$ e seus elementos são chamados *borelianos*. Notação: $\mathscr{B}_{[0,1]} = \{A \subset [0, 1]: A \text{ boreliano}\}$.

Observação. Vamos indicar com $\mathscr{B}$ a σ-álgebra de Borel na reta, i. e., a menor σ-álgebra contendo todos os intervalos. Os elementos desta σ-álgebra são os borelianos da reta. Em termos intuitivos, um boreliano é um conjunto que pode ser obtido de um número enumerável de intervalos aplicando-se as operações $\cup$, $\cap$ e ^{c} um número enumerável de vezes. O conjunto dos racionais, por exemplo, é boreliano por ser união enumerável de intervalos degenerados (pontos). O conjunto dos irracionais também o é, pois é complementar de união enumerável.

Definições e notações análogas valem para dimensões maiores que um. Por exemplo, $\mathscr{B}^2$ é a σ-álgebra de Borel no plano $\mathbb{R}^2$, i. e., a menor σ-álgebra contendo todos os retângulos. A idéia intuitiva de boreliano no plano é a de um conjunto que pode ser obtido partindo-se de um número enumerável de retângulos e aplicando-se as operações $\cup$, $\cap$ e ^{c} um número enumerável de vezes. Entre os borelianos do plano encontram-se as regiões abertas, porque toda região aberta pode ser descrita como união enumerável de retângulos. Aliás, todo subconjunto do plano que pode ser desenhado ou visualizado é boreliano, e podemos dizer a mesma coisa sobre os borelianos da reta, do espaço, etc.

No caso de n geral, $\mathscr{B}^n$ é a σ-álgebra de Borel no $\mathbb{R}^n$, i. e., a menor σ-álgebra contendo todos os retângulos n-dimensionais (os retângulos tridimensionais, por exemplo, são os paralelepípedos retângulos no $\mathbb{R}^3$).

Até agora, temos definido a probabilidade de um evento aleatório utilizando algumas definições clássicas (resultados equiprováveis, probabilidade geométrica). Outro método de definir probabilidade é o da freqüência relativa: poderíamos definir $P(A)$ como o limite da freqüência relativa da ocorrência de A em n repetições independentes do experimento, com n tendendo ao infinito, i. e.,

$$P(A) = \lim_{n \to \infty} \frac{1}{n} \times (\text{número de ocorrências de } A \text{ em } n \text{ "ensaios" independentes do experimento}).$$

Essa é a definição "freqüentista" ou "estatística" de probabilidade. Baseia-se na experiência, comum a todos nós, da estabilidade da freqüência relativa de ocorrência de eventos, quando realizamos muitas repetições do experimento. Essa definição foi usada por von Mises na construção de uma teoria de probabilidade. Não é, contudo, a definição que adotaremos neste livro, mas será obtida no Capítulo 5 como conseqüência da construção axiomática de probabilidade que começaremos agora.

Não vamos nos preocupar, doravante, com o problema de como definir probabilidade para cada experimento. Simplesmente, vamos admitir que *existem* as probabilidades em uma certa σ-álgebra $\mathscr{A}$ de eventos, chamados eventos aleatórios; vamos supor que a todo $A \in \mathscr{A}$ seja associado um número real $P(A)$, chamado *probabilidade de A*, de modo que os axiomas a seguir sejam satisfeitos. (Essa construção axiomática de probabilidade se deve a Kolmogorov – veja a referência [12] – e conseguiu proporcional à Teoria da Probabilidade uma base matemática firme.)

AXIOMA 1. $P(A) \geq 0$.
AXIOMA 2. $P(\Omega) = 1$.

AXIOMA 3. (*Aditividade finita*). *Se* $A_1, \ldots, A_n \in \mathscr{A}$ *são disjuntos* (2 *a* 2), *então* $P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k)$. (Os eventos são disjuntos, ou disjuntos 2 a 2, se são mutuamente exclusivos, i.e., $A_i \cap A_j = \varnothing$ se $i \neq j$.)

Observações. Como $A_1 \cup A_2 \cup A_3 = (A_1 \cup A_2) \cup A_3$, podemos usar indução para mostrar que o Axioma 3 está satisfeito (para todo n) quando está satisfeito para $n = 2$.

Uma função P satisfazendo Axiomas 1, 2 e 3 é chamada *probabilidade finitamente aditiva*. Embora alguma coisa tenha sido feita com tais probabilidades (veja, por exemplo, Dubins e Savage: *How to Gamble if You Must*), é matematicamente mais conveniente supor σ-aditividade:

AXIOMA 3'. (*σ-aditividade*). *Se* $A_1, A_2, \ldots \in \mathscr{A}$ *são disjuntos* (*i.e., mutuamente exclusivos*), *então*

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

PROPOSIÇÃO 1.3. *O Axioma 3' implica o Axioma 3, i.e., se P é σ-aditiva, então é finitamente aditiva.*

Prova. Suponhamos satisfeito o Axioma 3', e sejam $A_1, \ldots, A_n \in \mathscr{A}$ disjuntos. Notemos inicialmente que $P(\varnothing) = 0$, já que

$$P(\Omega) = P(\Omega \cup \varnothing \cup \varnothing \cup \ldots) = P(\Omega) + P(\varnothing) + P(\varnothing) + \ldots.$$

Definamos $A_k = \varnothing$ para $k = n+1, n+2, \ldots$. Então $A_1, A_2, \ldots$ são disjuntos, logo

$$P\left(\bigcup_{k=1}^{n} A_k\right) = P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) = \sum_{k=1}^{n} P(A_k) +$$

$$+ P(\varnothing) + P(\varnothing) + \ldots = \sum_{k=1}^{n} P(A_k). \square$$

DEFINIÇÃO 1.5. Uma função P definida numa σ-álgebra $\mathscr{A}$ e satisfazendo os Axiomas 1, 2 e 3' chama-se uma *medida de probabilidade em* $\mathscr{A}$ ou simplesmente uma *probabilidade em* $\mathscr{A}$.

Ocorre que, dados os Axiomas 1, 2, 3, o Axioma 3' é equivalente ao:

AXIOMA 4. ("*Continuidade no vazio*"). *Se a sequência* $(A_n)_{n \geq 1}$, *onde* $A_n \in \mathscr{A}$ $\forall n$, *decrescer para o vazio, então* $P(A_n) \to 0$ *quando* $n \to \infty$.

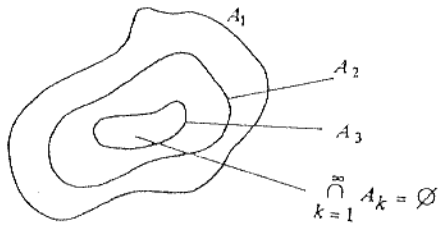
Observação. $(A_n)_{n \geq 1}$ *decresce para o vazio* ($A_n \downarrow \varnothing$) significa $A_n \supset A_{n+1}$ $\forall n$, ou seja, $(A_n)_{n \geq 1}$ decresce, e $\bigcap_{n \geq 1} A_n = \varnothing$.

PROPOSIÇÃO 1.4. *Dados os Axiomas 1, 2, 3, o Axioma 4 é equivalente ao Axioma 3' (i.e., uma probabilidade finitamente aditiva é uma probabilidade se, e só se, é contínua no vazio).*

Prova. (i) Suponhamos o Axioma 3'. Sejam $A_1, A_2, \ldots \in \mathscr{A}$ tais que $A_n \downarrow \varnothing$. Queremos provar que $P(A_n) \to 0$. Temos

$$A_1 = (A_1 - A_2) \cup (A_2 - A_3) \cup \ldots = \bigcup_{k=1}^{\infty} (A_k - A_{k+1}),$$

pelo diagrama:



Os "anéis" $A_k - A_{k+1}$ são disjuntos, porque a seqüência é decrescente, e pertencem a $\mathscr{A}$, já que $\mathscr{A}$ é fechada para diferenças (veja o exercício seguinte à Proposição 1.1.). Pelo Axioma 3',

$$P(A_1) = \sum_{k=1}^{\infty} P(A_k - A_{k+1}),$$

portanto a série é convergente e

$$\sum_{k=1}^{n-1} P(A_k - A_{k+1}) \xrightarrow[n \to \infty]{} P(A_1).$$

Pela aditividade finita,

$$P(A_k - A_{k+1}) = P(A_k) - P(A_{k-1}),$$

logo

$$P(A_1) = \lim_{n\to\infty} \sum_{k=1}^{n-1} (P(A_k) - P(A_{k+1})) = \lim_{n\to\infty} (P(A_1) - P(A_n)),$$

e então $P(A_n)\to 0$.

(ii) Suponhamos o Axioma 4 e sejam $A_1, A_2, \ldots \in \mathscr{A}$ disjuntos. Queremos provar que $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$. Seja $A = \bigcup_{n=1}^{\infty} A_n$, então

$$A = \left(\bigcup_{n=1}^{k} A_n\right) \cup \left(\bigcup_{n=k+1}^{\infty} A_n\right)$$

e pela aditividade finita,

$$P(A) = \sum_{n=1}^{k} P(A_n) + P\left(\bigcup_{n=k+1}^{\infty} A_n\right).$$

Seja $B_k = \bigcup_{n=k+1}^{\infty} A_n$, então $B_k \downarrow \varnothing$ e portanto $P(B_k)\to 0$ (pelo Axioma 4). Logo

$$\sum_{n=1}^{k} P(A_n) \xrightarrow[k\to\infty]{} P(A),$$

i.e., $P(A) = \sum_{n=1}^{\infty} P(A_n)$. □

COROLÁRIO. *Os dois seguintes sistemas de axiomas são equivalentes:*

Sistema I: *Axiomas* 1, 2, 3′.
Sistema II: *Axiomas* 1, 2, 3, 4.

Prova. O sistema I é equivalente aos Axiomas 1, 2, 3, 3′, pois já vimos que o Axioma 3′ implica o Axioma 3. Agora basta aplicar a Proposição 1.4. □

Observação. Então para verificar se P é probabilidade em $\mathscr{A}$, basta verificar os axiomas do sistema I ou os axiomas do sistema II.

Propriedades de probabilidade. Seja P uma probabilidade em uma σ-álgebra $\mathscr{A}$. Suponhamos que todo A abaixo pertença a $\mathscr{A}$. Então as seguintes propriedades são conseqüências dos axiomas:

P1. $P(A^c) = 1 - P(A)$. (Conseqüência dos axiomas 2 e 3.)

Caso particular importante: $P(\varnothing) = 1 - P(\Omega) = 0$.

P2. $0 \le P(A) \le 1$. (Conseqüência do Axioma 1 e P1.)

P3. $A_1 \subset A_2 \Rightarrow P(A_1) \le P(A_2)$. (Pela aditividade finita, $P(A_2) = P(A_1) + P(A_2 - A_1) \ge P(A_1)$, pelo Axioma 1.)

P4. $P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i)$. (Pela aditividade finita,

$$P(A_1 \cup A_2) = P(A_1) + P(A_2 \cap A_1^c) \le P(A_1) + P(A_2),$$

por P3, já que $A_2 \cap A_1^c \subset A_2$. Completa-se a prova por indução.)

P5. $P\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} P(A_i)$. (Exercício 2.)

P6. (*Continuidade de probabilidade*). *Se* $A_n \uparrow A$, *então* $P(A_n) \uparrow P(A)$. *Se* $A_n \downarrow A$, *então* $P(A_n) \downarrow P(A)$.

Prova de P6. Vamos supor que $A_n \downarrow A$, i.e., que $A_n \supset A_{n+1}\ \forall n$ e $\bigcap_{n\ge 1} A_n = A$. Então, $P(A_n) \ge P(A_{n+1})$, por P3, e $(A_n - A) \downarrow \varnothing \Rightarrow P(A_n - A) \to 0$, pela continuidade no vazio. A aditividade finita implica $P(A_n - A) = P(A_n) - P(A)$, pois $A \subset A_n$. Resumindo, temos $P(A_n) - P(A) \to 0$ e $\{P(A_n)\}_{n\ge 1}$ decrescente, logo $P(A_n) \downarrow P(A)$.

Se $A_n \uparrow A$ (i.e., $A_n \subset A_{n+1}\ \forall n$ e $\bigcup_{n\ge 1} A_n = A$), então $A_n^c \downarrow A^c$. Logo $P(A_n^c) \downarrow P(A^c)$, ou seja, $1 - P(A_n) \downarrow 1 - P(A)$; portanto $P(A_n) \uparrow P(A)$. □

Modelo probabilístico. Terminamos a formulação do modelo matemático para um experimento, ou modelo probabilístico. É constituído de:

(a) Um conjunto não-vazio Ω, de resultados possíveis, o *espaço amostral.*
(b) Uma σ-álgebra $\mathscr{A}$ de *eventos aleatórios.*
(c) Uma *probabilidade* P definida em $\mathscr{A}$.

Agora vamos retirar nosso modelo do contexto de um experimento e reformulá-lo como um conceito matemático abstrato.

DEFINIÇÃO 1.6. Um *espaço de probabilidade* é um trio $(\Omega, \mathscr{A}, P)$, onde (a) Ω é um conjunto não-vazio,
(b) $\mathscr{A}$ é uma σ-álgebra de subconjuntos de Ω e
(c) P é uma probabilidade em $\mathscr{A}$.

A partir de agora, tudo será estudado em espaços de probabilidade, apesar de mantermos a linguagem de experimentos e eventos. (Já vimos

que todo modelo probabilístico é um espaço de probabilidade. Reciprocamente, o espaço de probabilidade $(\Omega, \mathscr{A}, P)$ pode ser considerado um modelo para o experimento "selecionar um ponto de Ω conforme a probabilidade P". Se o leitor quiser, poderá continuar considerando um espaço de probabilidade como um modelo probabilístico.)

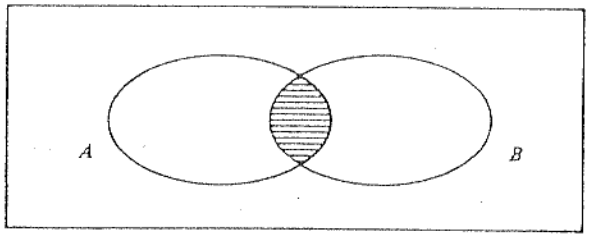
§1.2 Probabilidade condicional

DEFINIÇÃO 1.7. Seja $(\Omega, \mathscr{A}, P)$ um espaço de probabilidade. Se $B \in \mathscr{A}$ e $P(B) > 0$, a *probabilidade condicional de A dado B* é definida por

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad A \in \mathscr{A}.$$

Observação. Se $P(B) = 0$, $P(A \mid B)$ pode ser arbitrariamente definida. A maioria dos livros faz $P(A \mid B) = 0$, mas é mais interessante fazer $P(A \mid B) = P(A)$ para que $P(A \mid B)$ seja uma probabilidade em $\mathscr{A}$ (como função de A). É também conveniente, por independência, fazer $P(A \mid B) = P(A)$ – veja § 1.3.

Consideremos um diagrama de Venn:



Se A e B são desenhados de modo que as áreas de A, B e $A \cap B$ sejam proporcionais às suas probabilidades, então $P(A \mid B)$ é a proporção do evento B ocupada pelo evento A. Note que $P(A \mid B)$, $A \in \mathscr{A}$, é realmente uma probabilidade em $\mathscr{A}$ (verifique os axiomas!). Conseqüentemente as propriedades de probabilidade são mantidas, por exemplo: $P(A^c \mid B) = 1 - P(A \mid B)$.

Probabilidade condicional possui uma interpretação intuitiva em termos de freqüências relativas. Pensando em probabilidade como limite de freqüência relativa, temos

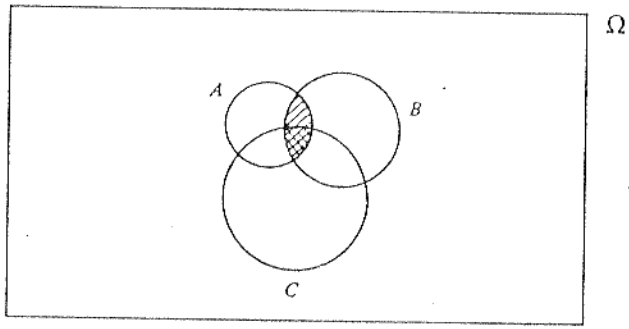
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} =$$

$$= \frac{\lim_{n \to \infty} \frac{1}{n} \times (\text{número de ocorrências de "}A \text{ e } B\text{" em } n \text{ ensaios independentes do experimento})}{\lim_{n \to \infty} \frac{1}{n} \times (\text{número de ocorrências de } B \text{ em } n \text{ ensaios independentes do experimento})}$$

$$= \lim_{n \to \infty} \frac{\text{número de ocorrências de } A \cap B \text{ em } n \text{ ensaios}}{\text{número de ocorrências de } B \text{ nos mesmos } n \text{ ensaios}}$$

Então, quando n é grande, $P(A \mid B)$ é aproximadamente igual ao quociente do número de ocorrências de A e B sobre o número de ocorrências de B em n ensaios independentes do experimento, i. e., $P(A \mid B)$ é aproximadamente a proporção, entre os experimentos em que ocorre o evento B, daqueles em que o evento A também ocorre. (Para uma aplicação desta interpretação a um caso específico, veja o exemplo 4 adiante.)

Decorre da definição que $P(A \cap B) = P(B)\,P(A \mid B)$, e esta igualdade é válida também quando $P(B) = 0$. Esta igualdade se generaliza: sendo A, B, C eventos aleatórios, temos $P(A \cap B \cap C) = P(A)\,P(B \mid A) \cdot P(C \mid A \cap B)$. Isto pode ser visto pelo seguinte diagrama, pensando nas probabilidades de todos os eventos como proporcionais às suas áreas.



(Prova formal: $P(A \cap B \cap C) = P(A \cap B)P(C \mid A \cap B) =$
$= P(A)P(B \mid A)P(C \mid A \cap B)$.)

Por indução, temos o seguinte

TEOREMA 1.1. (*Teorema da Multiplicação ou Teorema da Probabilidade Composta*). *Seja* $(\Omega, \mathscr{A}, P)$ *um espaço de probabilidade. Então*

(*i*) $P(A \cap B) = P(A)P(B \mid A) = P(B)P(A \mid B)$, $\forall\, A,\ B \in \mathscr{A}$,

(*ii*) $P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \ldots P(A_n \mid A_1 \cap \ldots \cap A_{n-1})$, $\forall\, A_1, \ldots, A_n \in \mathscr{A}$, $\forall\, n = 2, 3, \ldots$.

EXEMPLO 3. Selecionar três cartas de um baralho, ao acaso e sem reposição. Qual a probabilidade de tirar 3 reis?

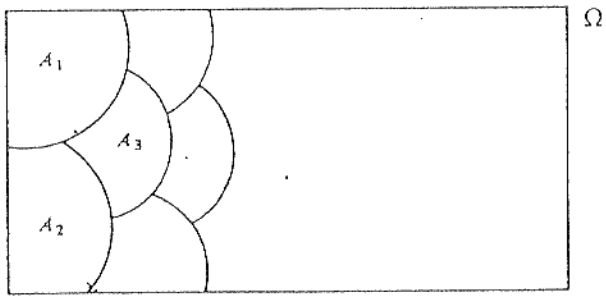
Seja A_i o evento "tirar rei na i-ésima extração". Então (com $A =$ = "tirar 3 reis") temos

$$P(A) = P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) = \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50}.$$

Verifiquemos através da distribuição hipergeométrica:

$$P(A) = \frac{\binom{4}{3}\binom{48}{0}}{\binom{52}{3}} = \frac{4 \cdot 1 \cdot 3! \cdot 49!}{52!} = \frac{4 \cdot 3 \cdot 2}{52 \cdot 51 \cdot 50}.$$

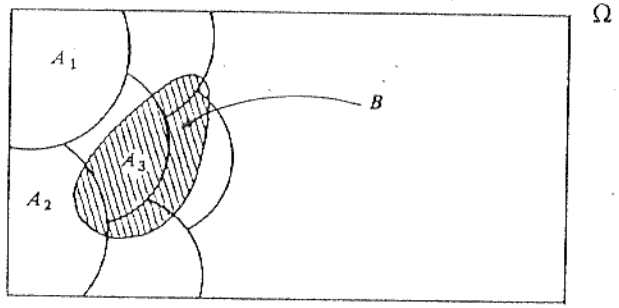
Agora suponhamos que $A_1, A_2, \ldots$ sejam eventos aleatórios mutuamente exclusivos e exaustivos (i. e., que os A_i sejam disjuntos = mutuamente exclusivos e $\cup A_i = \Omega$). Então os A_i formam uma *partição* do espaço amostral Ω:



Vamos admitir que a seqüência $A_1, A_2, \ldots$ seja finita ou enumerável — então, por exemplo, A e A^c formam uma partição, $\forall\, A \in \mathscr{A}$.

Para todo evento $B \in \mathscr{A}$, temos $B = \bigcup_i (A_i \cap B)$. Como os A_i são disjuntos, então os $B \cap A_i$ são disjuntos e

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



Logo temos o seguinte

TEOREMA 1.2. (*Teorema da Probabilidade Total (ou Absoluta)*). *Se a sequência (finita ou enumerável) de evêntos aleatórios* $A_1, A_2, \ldots$ *formar uma partição de* Ω, *então*

$$P(B) = \sum_i P(A_i)\, P(B \mid A_i), \quad \forall\ B \in \mathscr{A}.$$

Usando esse teorema, podemos calcular a probabilidade de A_i dada a ocorrência de B:

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)} \Rightarrow P(A_i \mid B) = \frac{P(A_i)\, P(B \mid A_i)}{\sum_j P(A_j)\, P(B \mid A_j)}.$$

Esta é a *fórmula de Bayes*. Ela é útil quando conhecemos as probabilidades dos A_i e a probabilidade condicional de B dado A_i, mas não conhecemos *diretamente* a probabilidade de B:

EXEMPLO 4. Experimento de duas etapas (experimento composto). Supor que uma caixa contenha três moedas: duas honestas e uma de duas caras. Retirar uma moeda ao acaso e jogá-la. Pergunta: qual a probabilidade condicional da moeda ter sido a de duas caras, dado que o resultado final foi cara?

Este é um experimento de duas etapas, e queremos calcular a probabilidade de um evento determinado pela primeira etapa dado um evento determinado pela segunda. Sejam, então, A_1 = "moeda re-

tirada é honesta", A_2 = "moeda retirada é a de duas caras", e B = "resultado final é cara".

Aplicando a fórmula de Bayes, temos

$$P(A_2 \mid B) = \frac{P(A_2)\,P(B \mid A_2)}{P(A_1)\,P(B \mid A_1) + P(A_2)\,P(B \mid A_2)} = \frac{\frac{1}{3} \cdot 1}{\frac{2}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot 1} = \frac{1}{2}.$$

Podemos interpretar este resultado da seguinte maneira, em termos de freqüência relativa: se o experimento fosse repetido independentemente um grande número de vezes, então a moeda de duas caras seria a escolhida na primeira etapa de aproximadamente metade dos experimentos em que o resultado final fosse cara.

Observação. A fórmula de Bayes é, às vezes, chamada de fórmula de probabilidades "posteriores". Com efeito, as probabilidades $P(A_i)$ podem ser chamadas probabilidades "a priori" e as $P(A_i \mid B)$, probabilidades "a posteriori".

§1.3 Independência

DEFINIÇÃO 1.8. Seja $(\Omega, \mathscr{A}, P)$ um espaço de probabilidade. Os eventos aleatórios A e B são (estocasticamente) *independentes* se

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

Observação. Eventos de probabilidade zero ou um são independentes de qualquer outro: se $P(A) = 0$, então $P(A \cap B) = 0$ e A e B são independentes, $\forall\, B \in \mathscr{A}$. Se $P(B) = 1$, então $P(A \cap B) = P(A) - P(A \cap B^c)$ e, como $A \cap B^c \subset B^c$ implica $P(A \cap B^c) \leq P(B^c) = 0$, temos $P(A \cap B^c) = 0$ e $P(A \cap B) = P(A) = P(A)\,P(B)$. Logo A e B são independentes, $\forall\, A \in \mathscr{A}$.

PROPOSIÇÃO 1.5. *A é independente de si mesmo se, e somente se, $P(A) = 0$ ou 1.*

Prova. $P(A) = P(A \cap A) = P(A)P(A) \Leftrightarrow P(A) = 0$ ou 1. □

PROPOSIÇÃO 1.6. *Se A e B são independentes, então A e B^c também são independentes (e também A^c e B, e ainda A^c e B^c).*

Prova. Supor A, B independentes. Então $P(A \cap B^c) = P(A) - P(A \cap B) =$ $= $ (pela independencia) $= P(A) - P(A)\,P(B) = P(A)\,(1 - P(B)) =$ $= P(A) \cdot P(B^c)$. □

Aqui está uma justificação intuitiva da Definição 1.8: B é independente de A se tanto a ocorrência quanto a não ocorrência de A não afetam a probabilidade de B ocorrer, i. e., $P(B \mid A) = P(B)$ e $P(B \mid A^c) = P(B)$. Estas duas equações significam que

$$P(A \cap B) = P(A)\,P(B \mid A) = P(A)\,P(B) \text{ e}$$
$$P(A^c \cap B) = P(A^c)\,P(B \mid A^c) = P(A^c)\,P(B);$$

pela Proposição 1.6, basta uma destas últimas equações para a definição.

Observação. Se $A \cap B = \varnothing$, então A e B não são independentes (a menos que um deles tenha probabilidade zero).

EXEMPLO 5. No experimento 1, os eventos A = "observa-se um número par" e A^c = "observa-se um número ímpar" não são independentes. Intuitivamente, porque não são compatíveis, e formalmente, porque

$$P(A \cap A^c) = P(\varnothing) = 0 \neq \frac{1}{4} = P(A)\,P(A^c).$$

EXEMPLO 6. No experimento 3, os eventos

A = "distância entre o ponto escolhido e a origem é $\leq \frac{1}{2}$"

e

C = "1.ª coordenada do ponto escolhido é maior que a 2.ª"

são independentes, pois o evento C ocupa metade da área do evento A, fazendo com que (veja os desenhos no §1.1)

$$P(A \cap C) = \frac{\text{área } (A \cap C)}{\pi} = \frac{1}{8} = \frac{1}{4} \cdot \frac{1}{2} = P(A)\,P(C).$$

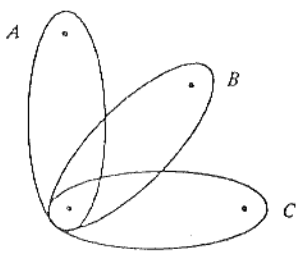
Como vamos definir a independência coletiva de três eventos aleatórios A, B e C? Queremos não somente que C seja independente

de A, de B, e que A e B sejam independentes (i. e., que os três eventos sejam independentes 2 a 2), mas também que C seja independente de $A \cap B$, $A \cap B^c$, etc. Isto é, queremos que a ocorrência do evento "A e B" não afete a probabilidade de ocorrência de C, etc. Por exemplo, queremos que $P(A \cap B \cap C) = P(A \cap B)\,P(C) = P(A)\,P(B)\,P(C)$, o que não é uma conseqüência da independência 2 a 2:

DEFINIÇÃO 1.9. Os eventos aleatórios A_i, $i \in I$ (I um conjunto de índices), são *independentes* 2 *a* 2 (ou *a pares*) se

$$P(A_i \cap A_j) = P(A_i)\,P(A_j) \quad \forall\, i,\ j \in I,\ i \neq j.$$

EXEMPLO 7. Independência a pares não implica independência coletiva. Seja Ω um conjunto de quatro pontos, com os eventos A, B, C assim definidos:



Seja $P(\omega) = \frac{1}{4}$, $\forall\, \omega \in \Omega$. Então $P(A) = P(B) = P(C) = \frac{1}{2}$ e $P(A \cap B) = \frac{1}{4} = P(A \cap C) = P(B \cap C)$. Logo A, B, C são independentes 2 a 2. Mas

$$P(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8} = P(A)\,P(B)\,P(C).$$

(Notemos que Ω suporta no máximo 2 eventos independentes de probabilidade $\frac{1}{2}$ cada, pois $\#\,\Omega = 4$. Para que existissem três eventos independentes de probabilidade $\frac{1}{2}$, Ω precisaria conter pelo menos 8 pontos, pois haveria 8 eventos incompatíveis de probabilidade $\frac{1}{8}$ cada: $A \cap B \cap C$, $A^c \cap B \cap C$, etc.)

Outro exemplo ainda mais intuitivo pode ser encontrado em Feller (Vol. I, 2.ª edição, §V. 3): no lançamento de dois dados honestos, sejam os eventos A = "face ímpar no primeiro dado", B = "face ímpar no segundo dado", C = "soma ímpar das duas faces". É fácil ver que A, B, C têm, cada um, probabilidade 1/2 e são independentes 2 a 2. Mas eles não podem ocorrer simultaneamente, de modo que $A \cap B \cap C = \varnothing$ e

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

DEFINIÇÃO 1.10. (a) Os eventos $A_1, \ldots, A_n$ ($n \geq 2$) são chamados (coletivamente ou estocasticamente) *independentes* se

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_m}) = P(A_{i_1})\,P(A_{i_2}) \ldots P(A_{i_m})$$

$\forall\, 1 \leq i_1 < i_2 < \ldots < i_m \leq n$, $\forall\, m = 2, 3, \ldots, n$ (i. e., se todas as combinações satisfazem a regra produto).

(b) Os eventos $A_1, A_2 \ldots$ são *independentes* se $\forall\, n \geq 2$, $A_1, A_2, \ldots, A_n$ são independentes.

(c) Os eventos A_i, $i \in I$ (onde I é um conjunto de índices tal que $\#\, I \geq 2$) são *independentes* se toda subfamília finita deles é de eventos independentes, i. e., se $A_{i_1}, A_{i_2}, A_{i_3}, \ldots, A_{i_m}$ são independentes para toda combinação $\{i_1, \ldots, i_m\}$ de elementos de I, $\forall\, m = 2, 3, \ldots$.

Observações. (1) Tais eventos são chamados, às vezes, *estatisticamente* ou *mutuamente* independentes.

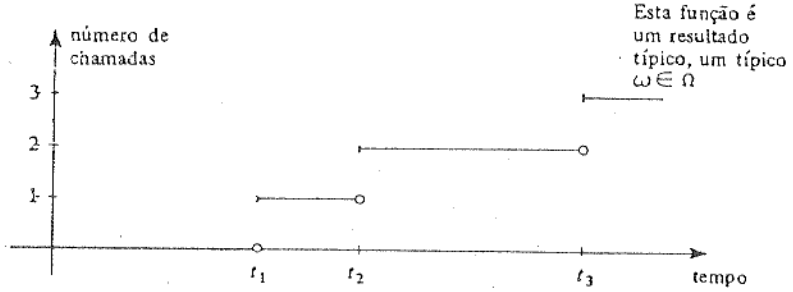
(2) Vemos pelo item (c) que toda subfamília de uma família de eventos independentes é de eventos independentes.

Vamos ver agora que a Definição 1.10 é consistente com nossa idéia intuitiva de independência (por exemplo, no caso de três eventos A, B, C, o evento C é independente de $A \cap B$, de $A \cap B^c$, de $A^c \cap B$, de $A^c \cap B^c$; especificamente, $P(A \cap B^c \cap C) = P(A)\,P(B^c)\,P(C)$, etc.).

PROPOSIÇÃO 1.7. *Se os eventos A_i, $i \in I$, são independentes, então os eventos B_i, $i \in I$, também são independentes, onde cada B_i é igual a A_i, ou A_i^c (ou um ou outro).*

Prova. Pelos itens (a) e (c) da definição, basta provar que toda subfamília finita dos B_i satisfaz a regra produto. Para tanto, é suficiente provar que se $A_1, \ldots, A_n$ são independentes, então $P(B_1 \cap \ldots \cap B_n) = \prod_{i=1}^{n} P(B_i)$, onde $B_i = A_i$ ou $B_i = A_i^c$. Esta prova é semelhante à prova da Proposição 1.6, usando indução finita. (*Exercício*: complete a prova. Se quiser, pode ver Fernandez [10], Lema 4.3.1).□

EXEMPLO 8. *O processo de Poisson.* Consideremos o número de telefonemas que chegam em uma central telefônica. Vamos contar o número de chamadas que chegam até o tempo t, para todo $t \geq 0$. Podemos representar um resultado possível deste experimento por meio de uma função-escada:



A cada tal função ω corresponde um resultado possível do experimento (chamadas chegam em $t_1, t_2, t_3, \ldots$), e cada resultado do experimento gera uma função deste tipo (sob certas suposições que estão adiante). Então podemos fazer Ω = conjunto de todas as funções-escada com gráfico do tipo que aparece acima $= \{\omega: [0, \infty) \to \{0, 1, 2, \ldots\} \mid \exists\ 0 < t_1 < t_2 < \ldots (t_n \uparrow \infty)$ tal que $\omega(t) = 0$ para $t \in [0, t_1)$, $\omega(t) = 1$ para $t \in [t_1, t_2), \ldots, \omega(t) = n$ em $[t_n, t_{n+1}), \ldots\}$.

Agora, seja o evento $A_{s,t}^k$ = "chegam exatamente k chamadas no intervalo $(s, s+t]$", para $s, t \geq 0$; $k = 0, 1, 2, \ldots$. Então,

$$A_{s,t}^k = \{\omega \in \Omega: \omega(s+t) - \omega(s) = k\},\ s, t \geq 0,\ k = 0, 1, 2, \ldots.$$

Vamos supor que a σ-álgebra $\mathcal{A}$ contenha todos os eventos $A_{s,t}^k$ (daqui a pouco calcularemos a probabilidade destes eventos). Vamos fazer as seguintes hipóteses:

HIPÓTESE 1. (*Incrementos estacionários*). *A probabilidade de chegada de k telefonemas no intervalo $(s, s+t]$ depende somente de t e não de s* (i. e., a probabilidade de exatamente k telefonemas

chegarem durante um período de duração t depende apenas de t e não da hora e nem do dia. Esta hipótese não é satisfeita na prática, mas é uma boa aproximação durante curtos períodos de tempo, por exemplo, durante o horário do pique.). Esta hipótese implica

$$P(A_{s,t}^k) = P(A_{0,t}^k) \overset{def}{=} P_k(t).$$

HIPÓTESE 2. (*Incrementos independentes.*) *Os números de chegadas durante intervalos disjuntos de tempo são independentes* (ou seja, $A_{s,t}^k$ e $A_{u,v}^j$ são independentes para toda escolha de k e j se $(s, s+t] \cap (u, u+v] = \varnothing$, e temos independência também no caso de 3, 4, 5, ... intervalos disjuntos).

HIPÓTESE 3. *As chamadas chegam sozinhas e não simultaneamente.* Isto será interpretado em termos de probabilidades condicionais da seguinte maneira: a probabilidade condicional de terem chegado duas ou mais chamadas em $(0, t]$, dado que chegaram uma ou mais chamadas em $(0, t]$, tende a zero quando $t \to 0$. Isto quer dizer que

$$\frac{\text{probabilidade de chegada de duas ou mais chamadas em } (0, t]}{\text{probabilidade de chegada de uma ou mais chamadas em } (0, t]} \xrightarrow{t \to 0} 0,$$

i. e.,

$$\frac{1 - P_0(t) - P_1(t)}{1 - P_0(t)} \to 0, \text{ quando } t \to 0,$$

ou equivalentemente,

$$\frac{P_1(t)}{1 - P_0(t)} \to 1, \text{ quando } t \to 0.$$

Podemos calcular agora as probabilidades $P_k(t)$. Vamos começar com $P_0(t) = P(A_{0,t}^0)$, e mostraremos que é uma função exponencial do tipo $e^{-\lambda t}$.

Como não chega telefonema algum no intervalo $(0, t]$ se, e somente se, nenhum telefonema chega nos n intervalos

$$\left(0, \frac{t}{n}\right], \left(\frac{t}{n}, \frac{2t}{n}\right], \ldots, \left(\frac{(n-1)t}{n}, t\right],$$

temos

$$A_{0,t}^0 = A_{0,t/n}^0 \cap A_{t/n,\,t/n}^0 \cap \ldots \cap A_{t-t/n,\,t/n}^0.$$

Pela hipótese 2 (os intervalos são disjuntos), estes eventos são independentes, logo

$$P_0(t) = \prod_{i=0}^{n-1} P(A^0_{it/n,\ t/n}) = \text{(pela hipótese 1)} = P_0^n\left(\frac{t}{n}\right), \forall\, t > 0, \forall\, n.$$

Então,

$$P_0(mt) = P_0^m(t)$$

e

$$P_0\left(\frac{m}{n}t\right) = P_0^m\left(\frac{t}{n}\right) = P_0^{m/n}(t),$$

para todo m e n $(= 1, 2, 3, \ldots)$.

Em outras palavras, se $r > 0$ é racional, $P_0(r) = P_0^r(1)$. Ora, $P_0(t)$ é uma função decrescente, pois

$$t \le s \Rightarrow A^0_{0,\,s} \subset A^0_{0,\,t} \Rightarrow P_0(t) \ge P_0(s).$$

Logo para $t > 0$ fixo e r_1, r_2 racionais tais que $r_1 \le t \le r_2$, temos

$$P_0^{r_1}(1) = P_0(r_1) \ge P_0(t) \ge P_0(r_2) = P_0^{r_2}(1).$$

Se $r_1 \uparrow t$ e $r_2 \downarrow t$, então $P_0^{r_1}(1) \downarrow P_0^t(1)$ e $P_0^{r_2}(1) \uparrow P_0^t(1)$, logo

$$P_0(t) = P_0^t(1) \quad \forall\, t > 0.$$

Podemos supor $0 < P_0(1) < 1$, para evitar um caso trivial ($P_0(1) = 1$) e outro que contradiz as hipóteses ($P_0(1) = 0$). Com efeito, se $P_0(1)$ fosse igual a um, teríamos $P_0(t) = 1$ para todo $t > 0$, i.e., com probabilidade 1, nunca chegaria nada. Esse é um caso trivial que não é de maior interesse na prática. Por outro lado, se $P_0(1)$ fosse igual a zero, teríamos $P_0(t) = 0$ para todo $t > 0$, i.e., para cada $t > 0$, haveria probabilidade um de chegar pelo menos um telefonema em $(0, t]$. Portanto, teriam que chegar pelo menos dois telefonemas em $(0, t]$, com probabilidade um, pois a chegada de pelo menos um em $\left(0, \frac{t}{2}\right]$ e pelo menos um em $\left(\frac{t}{2}, t\right]$ (este evento também seria de probabilidade um, pela hipótese 1), implica a chegada de pelo menos dois em $(0, t]$. Em conseqüência disto, teríamos $1 - P_0(t) = 1$ e $1 - P_0(t) - P_1(t) = 1$ para todo $t > 0$, contradizendo assim a hipótese 3.

Definindo $\lambda = -\log P_0(1)$, temos o resultado enunciado, i.e.,

$$P_0(t) = e^{-\lambda t},\ t > 0.$$

Observação. É claro que $P_0(0) = 1$, pois o evento "nenhuma chegada em um intervalo vazio de tempo" é o evento certo. Formalmente,

$$A^0_{s,0} = \{\omega \in \Omega:\ \omega(s) - \omega(s) = 0\} = \Omega.$$

Conseqüentemente, $P_0(t) = e^{-\lambda t}$ para $t \ge 0$ e P_0 é contínua em $[0, \infty)$.

Obteremos agora as probabilidades $P_k(t)$, para todo valor de k. O método que utilizaremos consiste na aplicação das hipóteses para obter equações diferenciais satisfeitas pelas funções P_k, com a subseqüente solução destas equações. A derivação a seguir, até a fórmula (1.2), pode ser omitida em uma primeira leitura.

Sejam $k \ge 1$, $s \ge 0$ e $t > 0$. Então chegam k telefonemas em $(0, s+t]$ se, é somente se, *ou* chega nenhum em $(0, s]$ e chegam k em $(s, s+t]$, *ou* chega um em $(0, s]$ e chegam $k-1$ em $(s, s+t]$, *ou* .... Isto é,

$$A^k_{0,\,s+t} = (A^0_{0,\,s} \cap A^k_{s,\,t}) \cup (A^1_{0,\,s} \cap A^{k-1}_{s,\,t}) \cup \ldots \cup (A^k_{0,\,s} \cap A^0_{s,\,t}).$$

Os eventos $A^i_{0,\,s} \cap A^{k-i}_{s,\,t}$ são disjuntos em i, e para todo i os eventos $A^i_{0,\,s}$ e $A^{k-i}_{s,\,t}$ são independentes (pela hipótese 2, pois os intervalos $(0, s]$ e $(s, s+t]$ são disjuntos). Logo

$$P_k(s+t) = \sum_{i=0}^{k} P(A^i_{0,\,s})\, P(A^{k-i}_{s,\,t}) = \sum_{i=0}^{k} P_i(s)\, P_{k-i}(t) =$$

$$= \sum_{i=0}^{k-2} P_i(s)\, P_{k-i}(t) + P_{k-1}(s)\, P_1(t) + P_k(s)\, e^{-\lambda t}.$$

Como $P_0(t) = e^{-\lambda t}$, a regra de L'Hôpital implica que

$$\lim_{t \to 0} \frac{1 - P_0(t)}{t} = \lim_{t \to 0} \frac{1 - e^{-\lambda t}}{t} = \lambda.$$

Pela hipótese 3 temos, então,

$$\lim_{t \to 0} \frac{P_1(t)}{t} = \lim_{t \to 0} \left\{ \frac{P_1(t)}{(1 - P_0(t))} \cdot \frac{(1 - P_0(t))}{t} \right\} = \lambda$$

e

$$\lim_{t \to 0} \frac{1 - P_0(t) - P_1(t)}{t} = \lim_{t \to 0} \left\{ \frac{(1 - P_0(t) - P_1(t))}{(1 - P_0(t))} \cdot \frac{(1 - P_0(t))}{t} \right\} = 0.$$

Agora, indicaremos com $P_k'(s)$ a derivada *a direita* de P_k em s:

$$P_k'(s) = \lim_{t \to 0} \frac{P_k(s+t) - P_k(s)}{t} =$$

$$= \lim_{t \to 0} \left[\frac{\sum_{i=0}^{k-2} P_i(s)\, P_{k-i}(t)}{t} + \frac{P_{k-1}(s)\, P_1(t)}{t} + \frac{P_k(s)\,(e^{-\lambda t} - 1)}{t} \right].$$

Aplicando as três fórmulas acima, temos

$$\lim_{t \to 0} \frac{P_k(s)\,(e^{-\lambda t} - 1)}{t} = -\lambda\, P_k(s),$$

$$\lim_{t \to 0} \frac{P_{k-1}(s)\, P_1(t)}{t} = \lambda\, P_{k-1}(s),$$

$$0 \leq \frac{\sum_{i=0}^{k-2} P_i(s)\, P_{k-i}(t)}{t} \leq \frac{\sum_{i=0}^{k-2} P_{k-i}(t)}{t} = \text{(fazendo } j = k - i) =$$

$$= \frac{\sum_{j=2}^{k} P_j(t)}{t} \leq \frac{1 - P_0(t) - P_1(t)}{t} \underset{t \to 0}{\to} 0.$$

Portanto, a derivada à direita é (trocando s por t)

$$P_k'(t) = \lambda\, P_{k-1}(t) - \lambda\, P_k(t), \tag{1.1}$$

para $t \geq 0$ e $k = 1, 2, \ldots$. Podemos provar que, para $t > 0$, a derivada à esquerda é a mesma, usando a equação

$$P_k(s) = \sum_{i=0}^{k} P_i(s - t)\, P_{k-i}(t).$$

Resta, então, resolver as equações diferenciais (1.1), sujeitas às condições iniciais

$$P_k(0) = P(A_{0,0}^k) = 0, \quad k \geq 1,$$

e levando-se em conta que

$$P_0(t) = e^{-\lambda t}, \quad t \geq 0.$$

A solução pode ser obtida por indução. Para ilustrar o método, obteremos $P_1(t)$, que satisfaz a equação

$$P_1'(t) = \lambda\, P_0(t) - \lambda\, P_1(t) = \lambda\, e^{-\lambda t} - \lambda\, P_1(t).$$

Fazendo

$$P_1(t) = e^{-\lambda t}\, Q(t),$$

temos $Q(0) = 0$ e $Q'(t) = \lambda$, de modo que $Q(t) = \lambda t$ e

$$P_1(t) = \lambda t\, e^{-\lambda t}, \quad t \geq 0.$$

A solução geral, para $k = 0, 1, 2, \ldots$, é

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad t \geq 0. \tag{1.2}$$

(*Exercício.* Verifique essa solução indutivamente, através da solução direta das equações diferenciais, ou através da substituição nas equações da solução proposta.)

Observações. Provamos, então, que o número de chegadas até o tempo t possui distribuição de Poisson com parâmetro λt (veja o Capítulo 2 para a definição de distribuição). λt é o número médio de chegadas durante um período de duração t; λ é o número médio de chegadas durante um intervalo *unitário* de tempo (veja o Capítulo 3). λ é chamado o *parâmetro* do processo de Poisson. Como representa, neste exemplo, uma taxa média de chegadas, é também chamada *taxa* ou *intensidade* do processo.

Poderíamos mostrar que nossas três hipóteses determinam uma probabilidade na σ-álgebra gerada pelos eventos $A_{s,t}^k$, completando assim um modelo probabilístico para o processo de Poisson.

EXERCÍCIOS DO CAPÍTULO 1

§1.1

1. Sejam A, B e C eventos aleatórios. Identifique as seguintes equações e frases, casando cada equação expressa na notação de conjuntos com a correspondente frase na linguagem de eventos:

   (a) $A \cap B \cap C = A \cup B \cup C$ — (*i*) A e "B ou C" são incompatíveis.

   (b) $A \cap B \cap C = A$ — (ii) Os eventos A, B, C são idênticos.

(c) $A \cup B \cup C = A$ (iii) A ocorrência de A implica a de "B e C".

(d) $(A \cup B \cup C) - (B \cup C) = A$ (iv) A ocorrência de A decorre da de "B ou C".

2. A partir dos axiomas, prove a propriedade P5:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n).$$

3. Sejam $A_1, A_2, \ldots$ eventos aleatórios. Mostre que:

(a) $P\left(\bigcap_{k=1}^{n} A_k\right) \geq 1 - \sum_{k=1}^{n} P(A_k^c)$.

(b) Se $P(A_k) \geq 1 - \varepsilon$ para $k = 1, \ldots, n$, então $P\left(\bigcap_{k=1}^{n} A_k\right) \geq 1 - n\varepsilon$.

(c) $P\left(\bigcap_{k=1}^{\infty} A_k\right) \geq 1 - \sum_{k=1}^{\infty} P(A_k^c)$.

4. Demonstre as seguintes propriedades:

(a) Se $P(A_n) = 0$ para $n = 1, 2, \ldots$, então $P\left(\bigcup_{n=1}^{\infty} A_n\right) = 0$.

(b) Se $P(A_n) = 1$ para $n = 1, 2, \ldots$, então $P\left(\bigcap_{n=1}^{\infty} A_n\right) = 1$.

5. Demonstre: se $A_1, A_2, \ldots$ e $B_1, B_2 \ldots$ são eventos aleatórios do mesmo espaço de probabilidade tais que $P(A_n) \to 1$ e $P(B_n) \to p$, quando $n \to \infty$, então $P(A_n \cap B_n) \to p$.

6. Seja Ω um conjunto não-vazio.

(a) Prove: se $\mathscr{A}$ e $\mathscr{B}$ são σ-álgebras de subconjuntos de Ω, então $\mathscr{A} \cap \mathscr{B}$ também é uma σ-álgebra.

(b) Generalize o item (a): se $\mathscr{A}_i$, $i \in I$, são σ-álgebras de partes de Ω, onde I é um conjunto não-vazio de índices, então $\bigcap_{i \in I} \mathscr{A}_i$ também é uma σ-álgebra.

(c) Seja $\mathscr{C}$ uma classe de subconjuntos de Ω. Mostre que existe *pelo menos uma* σ-álgebra que contém $\mathscr{C}$. (*Sugestão.* Qual a "maior" classe de subconjuntos de Ω?)

(d) Visando a plena utilização dos itens (b) e (c), como você definiria "a menor σ-álgebra contendo $\mathscr{C}$", onde $\mathscr{C}$ é uma classe de subconjuntos de Ω?

7. Sejam $A_1, A_2, \ldots$ eventos aleatórios em um espaço de probabilidade $(\Omega, \mathscr{A}, P)$, e definam-se

$$\limsup_{n \to \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k,$$

$$\liminf_{n \to \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k.$$

(Veremos uma interpretação intuitiva desses eventos no §5.2.) Se $\limsup_{n \to \infty} A_n = \liminf_{n \to \infty} A_n = A$, chamamos o evento A de $\lim_{n \to \infty} A_n$ (limite de A_n). Demonstre que se $A = \lim A_n$, então $P(A_n) \to P(A)$ quando $n \to \infty$.

8. No jogo de "Craps" dois dados são jogados. Se o jogador tira 7 ou 11 pontos ele ganha. Se ele tira 2, 3 ou 12 ele perde. Nos outros casos ele continua jogando os dois dados até sair 7, caso em que ele perde, ou então sair o primeiro resultado, caso em que ele ganha. Descreva o espaço amostral. Qual é a probabilidade dele ganhar?

9. Uma caixa contém $2n$ sorvetes, n do sabor A e n do sabor B. De um grupo de $2n$ pessoas, $a < n$ preferem o sabor A, $b < n$ o sabor B e $2n - (a + b)$ não têm preferência. Demonstre: se os sorvetes são distribuídos ao acaso, a proabilidade de que a preferência de todas as pessoas seja respeitada é de $\binom{2n-a-b}{n-a} \Big/ \binom{2n}{n}$.

10. Suponhamos que dez cartas estejam numeradas de 1 até 10. Das dez cartas, retira-se uma de cada vez, ao acaso e sem reposição, até retirar-se o primeiro número par. Conta-se o número de retiradas necessárias. Exiba um bom modelo probabilístico para este experimento.

11. Para cada um dos seguintes experimentos, descreva um espaço de probabilidade que sirva de modelo.

(a) Seleciona-se um ponto, ao acaso, do quadrado unitário

$$\{(x, y): 0 \leq x \leq 1, 0 \leq y \leq 1\}.$$

(b) Retiram-se cartas sucessivamente de um baralho de 52 cartas, ao acaso e *com* reposição, até retirar-se o primeiro rei. Registra-se o número total de retiradas.

(c) Quinze bolas são retiradas, ao acaso e *com* reposição, de uma urna contendo 5 bolas vermelhas, 9 bolas pretas, e uma bola branca. Observa-se o número de vezes que ocorra cada cor.
(d) O experimento (c) é realizado *sem* reposição.

12. Retiram-se 4 cartas, ao acaso, de um baralho de 52 cartas. Registra-se o número de reis na amostra. Exiba um bom modelo probabilístico para este experimento se:

(a) As retiradas são feitas *sem* reposição.
(b) As retiradas são feitas *com* reposição.
(c) Determine em que caso, (a) ou (b), é mais provável obter 4 reis.

13. (a) Sejam A, B e C eventos aleatórios em um espaço de probabilidade $(\Omega, \mathscr{A}, P)$. Mostre que $P(A\cup B) = P(A) + P(B) - P(A\cap B)$ e $P(A\cup B\cup C) = P(A) + P(B) + P(C) - P(A\cap B) - P(A\cap C) - P(B\cap C) + P(A\cap B\cap C)$.

(b) Enuncie a generalização do item (a) para o caso da união de n eventos aleatórios.

(c) Prove as seguintes *desigualdades de Bonferroni*:

(i) $$\sum_{i=1}^{n} P(A_i) - \sum_{1\le i<j\le n} P(A_i\cap A_j) \le P\left(\bigcup_{i=1}^{n} A_i\right) \le$$
$$\le \sum_{i=1}^{n} P(A_i) - \sum_{1\le i<j\le n} P(A_i\cap A_j) + \sum_{1\le i<j<k\le n} P(A_i\cap A_j\cap A_k).$$

(ii) Se k é ímpar, $k\le n$, então
$$P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i) - \sum_{1\le i_1<i_2\le n} P(A_{i_1}\cap A_{i_2}) +$$
$$+ \ldots (-1)^{k-1} \sum_{1\le i_1<\cdots<i_k\le n} P(A_{i_1}\cap\ldots\cap A_{i_k});$$

se k é par, $k\le n$, vale $\ge$ nesta última desigualdade.

14. (Problema de casamentos.)

(a) No exercício 1, se você identificasse ao acaso as equações e frases, qual a probabilidade de que você acertaria pelo menos um casamento?
(b) Resolva o item (a) para o caso em que há n equações e n frases para serem identificadas. (*Sugestão.* Use o exercício 13 (b).)
(c) Mostre que a probabilidade considerada no item (b) converge para $p = \frac{1}{e}$ quando $n\to\infty$.

15. Suponha que n cartas numeradas de 1 até n sejam embaralhadas e retiradas uma por uma, sem reposição, até todas as cartas serem retiradas. Qual a probabilidade de que para pelo menos uma carta, o número da carta coincida com o número da retirada? (*Observação.* A resposta é igual à do exercício 14 (b). Por quê?).

§1.2

16. Seja $(\Omega, \mathscr{A}, P)$ um espaço de probabilidade e suponha que todos os conjuntos abaixo pertençam a $\mathscr{A}$. Prove:

(a) Se os A_n são disjuntos e $P(B\,|\,A_n)\ge c$ para todo n, então $P(B\,|\cup A_n)\ge c$ (pode supor $P(A_n)>0$ para todo n).
(b) O item (a) com "=" no lugar de "$\ge$".
(c) Se $A_n\supset A_{n+1}$ e $P(A_{n+1}\,|\,A_n)\le \frac{1}{2}$ para todo n, então $P(A_n)\to 0$ quando $n\to\infty$.
(d) Se os A_n são disjuntos e $P(B\,|\,A_n) = P(C\,|\,A_n)\ \forall\, n$, então
$$P(B\,|\cup A_n) = P(C\,|\cup A_n).$$
(e) Se $A_1, A_2, \ldots$ são disjuntos e $\cup A_n = \Omega$, então
$$P(B\,|\,C) = \sum_n P(A_n\,|\,C)\,P(B\,|\,A_n\cap C).$$

17. Suponha que a ocorrência ou não de chuva dependa das condições do tempo no dia imediatamente anterior. Admita-se que se chove hoje, choverá amanhã com probabilidade 0,7 e que se não chove hoje choverá amanhã com probabilidade 0,4. Sabendo-se que choveu hoje, calcule a probabilidade de que choverá depois de amanhã.

18. 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, e
(b) a soma dos números ser 8?

19. Em teste de múltipla escolha, a probabilidade do aluno saber a resposta é p. Havendo m escolhas, se êle sabe a resposta êle responde corretamente com probabilidade 1; se não sabe êle responde corretamente com probabilidade $\frac{1}{m}$. Qual a probabilidade que êle sabia a resposta dado que a pergunta foi respondida corretamente?

Calcule o limite desta probabilidade quando (i) $m \to \infty$ com p fixo e (ii) $p \to 0$ com m fixo.

20. (De Fernandez [10].) Durante o mês de novembro a probabilidade de chuva é de 0,3. O Fluminense ganha um jogo em um dia com chuva com a probabilidade de 0,4; em um dia sem chuva com a probabilidade de 0,6. Se ganhou um jogo em novembro, qual é a probabilidade de que choveu nesse dia?

21. (De Fernandez [10].) Pedro quer enviar uma carta a Marina. A probabilidade de que Pedro escreva a carta é de 0,80. A probabilidade de que o correio não a perca é de 0,9. A probabilidade de que o carteiro a entregue é de 0,9. Dado que Marina não recebeu a carta, qual é a probabilidade condicional de que Pedro não a tenha escrito?

§1.3

22. Sejam $A_1, \ldots, A_n$ eventos aleatórios independentes, com $p_k = P(A_k)$, $k = 1, \ldots, n$. Obtenha a probabilidade de ocorrência dos seguintes eventos, em termos das probabilidades p_k:

(a) A ocorrência de nenhum dos A_k.
(b) A ocorrência de pelo menos um dos A_k.
(c) A ocorrência de exatamente um dos A_k.
(d) A ocorrência de exatamente dois dos A_k.
(e) A ocorrência de todos os A_k.
(f) A ocorrência de, no máximo, $n-1$ dos A_k.

23. Sejam $A_1, \ldots, A_n$ eventos aleatórios independentes, com $p_k = P(A_k)$, $k = 1, \ldots, n$. Faça uma adaptação das desigualdades de Bonferroni (exercício 13 (c)) para este caso, expressando-as em termos das p_k.

24. Em certa rodovia, a intensidade média do fluxo de tráfego é de 30 carros por minuto. Um medidor é colocado na rua para registrar o número de carros passando por cima. Suponha válidas as três hipóteses do processo de Poisson, adaptadas para a contagem de carros em vez de telefonemas, e calcule:

(a) A probabilidade de que dois ou mais carros sejam registrados durante determinado intervalo de dois segundos.
(b) A probabilidade de passar mais de um minuto até registrar o primeiro carro.

25. Consideremos um experimento em que será contado o número de estrelas em uma região longíqua do espaço, a região sendo de volume V. Façamos as seguintes três hipóteses, que são análogas espaciais das hipóteses do processo de Poisson:

(H1) A probabilidade de achar k estrelas na região depende somente de V.
(H2) Os números de estrelas contadas em regiões disjuntas do espaço são independentes.
(H3) Duas estrelas não ocupam o mesmo lugar.

Interpretando estas hipóteses de maneira semelhante à do processo do Poisson, obtenha o valor de $P_k(V) =$ probabilidade de achar exatamente k estrelas na região de volume V. Aqui, o parâmetro λ é a *densidade estelar* na vizinhança da região sendo considerada.

26. N pontos são escolhidos, independentemente e ao acaso, de uma esfera (bola) de raio R.

(a) Calcule a probabilidade da distância entre o centro da esfera e o ponto mais próximo ser maior que r.
(b) Qual o limite da probabilidade obtida no item (a) quando $R \to \infty$ e $\frac{N}{R^3} \to \frac{4}{3}\pi\lambda$? (Observação: este λ é o mesmo do exercício anterior.)

27. Acende-se uma lâmpada no instante $t = 0$. Para $t > 0$, seja $Q(t + \Delta t | t)$ a probabilidade condicional da lâmpada queimar até o instante $t + \Delta t$, dado que ficou acesa até o instante t. Suponha que $\forall\, t > 0$, $\frac{Q(t + \Delta t | t)}{\Delta t} \to \lambda t$ quando $\Delta t \to 0$, onde $\lambda > 0$ não depende de t. (Este limite é chamada *taxa de falha* da lâmpada. Neste exemplo, a taxa de falha, λt, é proporcional à idade.)

(a) Ache a equação diferencial satisfeita pela função $P(t) =$ probabilidade da lâmpada ficar acesa até o instante t. Você pode supor que a função P seja contínua, com $P(0) = 1$, e que as derivadas à direita e à esquerda sejam iguais.
(b) Resolva a equação diferencial do item (a).
(c) Obtenha e resolva a equação diferencial satisfeita por $P(t)$ quando a taxa de falha é constante ($= \lambda$).

28. Uma lâmpada está acesa no tempo $t = 0$. Sempre que a lâmpada queimar, é substituída por uma lâmpada nova, embora isso não seja feito imediatamente. Suponha que para todo $t > 0$:

(H1) dado que a lâmpada esteja acesa no instante t, a probabilidade dela estar queimada no instante $t + \Delta t$, dividida por Δt, converge para λ quando $\Delta t \to 0$; e

(H2) dado que a lâmpada esteja queimada no instante t, a probabilidade dela estar novamente acesa em $t + \Delta t$, dividida por Δt, converge para ξ quando $\Delta t \to 0$. ($\lambda, \xi > 0$.)

(a) Seja $P(t)$ a probabilidade da lâmpada estar acesa no instante t, $t \geq 0$. Ache a equação diferencial satisfeita por $P(t)$.

(b) Resolva a equação diferencial do item (a). Determine $\lim_{t \to \infty} P(t)$.

Esse resultado tem sentido intuitivo?

29. Suponhamos que cada elemento de certa população ou morre ou se divide. (Exemplo: uma colônia de bactérias.) Façamos três hipóteses:

(H1) A probabilidade de que um elemento, vivo no instante t, venha a morrer até o instante $t + \Delta t$, é assintoticamente equivalente a $\mu \Delta t$ (i. e., a razão dos dois converge para 1 quando $\Delta t \to 0$).

(H2) Um elemento vivo no instante t se divide até o instante $t + \Delta t$ com probabilidade assintoticamente equivalente a $\lambda \Delta t$, e produz "netos" (i. e., se divide ao menos duas vezes) com probabilidade que, dividida por Δt, converge para 0 quando $\Delta t \to 0$.

(H3) Não há interação entre os elementos, e eles morrem ou se dividem independentemente.

(a) Ache as equações diferenciais satisfeitas pelas probabilidades $P_n(t)$ = probabilidade da população conter exatamente n elementos no instante t ($n = 0, 1, 2, \ldots$; $t \geq 0$).

(b) Mostre que se $\lambda = \mu = 1$ e $P_1(0) = 1$, uma solução será

$$P_0(t) = \frac{t}{1+t}; \; P_n(t) = \frac{t^{n-1}}{(1+t)^{n+1}}, \; n = 1, 2, \ldots .$$

(c) Supondo que a solução do item (b) seja a única, qual a probabilidade da população mais cedo ou mais tarde ficar extinta?



CAPÍTULO II

VARIÁVEIS ALEATÓRIAS

§2.1 Variáveis aleatórias e funções de distribuição

Informalmente, uma variável aleatória é um *característico numérico* do resultado de um experimento. Por exemplo:

EXEMPLO 1. Lançar uma moeda n vezes e observar a seqüência de caras (c) e coroas ($\tilde{c}$) obtidas. Os resultados possíveis aqui são seqüências de extensão n de caras e coroas e podemos definir

$$\Omega = \{(\omega_1, \ldots, \omega_n): \; \omega_i = c \text{ ou } \tilde{c}; \; i = 1, \ldots, n\}.$$

O número de caras observadas nos n lançamentos é um característico numérico da seqüência de caras e coroas. De fato, se definimos X = número de caras observadas, vemos que o valor de X depende do resultado do experimento e podemos definir

$$\begin{aligned} X(\omega) &= \text{número de } c\text{'s em } \omega = (\omega_1, \ldots, \omega_n) \\ &= \#\{i: \; \omega_i = c, \; 1 \leq i \leq n\}. \end{aligned}$$

EXEMPLO 2. Escolher um ponto ao acaso em $[0, 1]$. Seja X o quadrado do valor obtido. Então

$$\Omega = [0, 1]$$

e

$$X(\omega) = \omega^2.$$

EXEMPLO 3. Escolher um ponto ao acaso no círculo unitário. Seja X a distância entre o ponto escolhido e a origem. Então

$$\Omega = \{(x, y): \; x^2 + y^2 \leq 1\}$$

e, com $\omega = (x, y)$,

$$X(\omega) = \sqrt{x^2 + y^2}.$$

Agora vemos que quando o resultado de experimento for um número real, o próprio resultado será o valor de uma variável aleatória, definida por $X(\omega) = \omega$:

EXEMPLO 4. Escolher um ponto ao acaso em $[0, 1]$, e seja X o valor de resultado. Então

$$\Omega = [0, 1], \ X(\omega) = \omega.$$

Quando o resultado for um ponto no plano, poderá ser considerado como valor de um par de variáveis aleatórias:

EXEMPLO 5. Escolher um ponto ao acaso no círculo unitário, e sejam X e Y as coordenadas do resultado. Então

$$\Omega = \{(x, y): \ x^2 + y^2 \leq 1\}$$

e, com $\omega = (x, y)$, temos $X(\omega) = x$, $Y(\omega) = y$, e $(X(\omega), \ Y(\omega)) = (x, y) = \omega$.

Nestes exemplos, cada variável aleatória é uma função real do resultado de um experimento. No exemplo 4, X é a *função identidade*. No exemplo 5, X e Y são as (funções) *coordenadas*. Não vamos admitir, contudo, que *toda* função de ω seja uma variável aleatória. Por razões técnicas, diremos que $X(\omega)$ é variável aleatória se, e somente se, o evento $[X \leq x] \stackrel{def}{=} \{\omega \in \Omega: \ X(\omega) \leq x\}$ for aleatório para todo x real. (Com efeito, adotaremos essa definição para poder definir a função de distribuição de X e daí a esperança de X, etc. Definiremos a função de distribuição de X, na Definição 2.2, como $F(x) = P(X \leq x)$, que não terá sentido exceto quando $[X \leq x]$ pertencer a $\mathscr{A}$.)

DEFINIÇÃO 2.1. Uma *variável aleatória* X em um espaço de probabilidade $(\Omega, \mathscr{A}, P)$ é uma função real definida no espaço Ω tal que $[X \leq x]$ é evento aleatório para todo $x \in \mathbb{R}$; i.e., $X: \Omega \to \mathbb{R}$ é variável aleatória se $[X \leq x] \in \mathscr{A} \ \forall x \in \mathbb{R}$.

Observação. Na linguagem da Teoria da Medida, $[X \leq x] \in \mathscr{A} \ \forall x \in \mathbb{R}$ significa que X é uma função mensurável a $\mathscr{A}$. Não é fácil encontrar exemplos de funções não-mensuráveis. De fato, as funções encontradas na prática são sempre variáveis aleatórias, e não nos preocuparemos com esta questão. Mas para ilustrar o conceito, consideremos o seguinte:

EXEMPLO 6. O processo de Poisson. Se nosso experimento for a observação do desenrolar do processo de Poisson, então um resultado típico ω será uma função-escada (veja §1.3). Para todo $t \geq 0$, definimos X_t = número de chegadas até o tempo t (inclusive). Então X_t é um característico numérico do resultado do experimento, e $X_t(\omega) = \omega(t)$. É fácil ver que $[X_t = k] = A^k_{0,t}$, que é evento aleatório por suposição (de fato, a esse evento foi atribuída a probabilidade $\frac{(\lambda t)^k}{k!} e^{-\lambda t}$). Isso torna X_t uma variável aleatória, já que $[X_t \leq x] = \bigcup_{0 \leq k \leq x} [X_t = k] \in \mathscr{A}$, para $x \geq 0$ (se $x < 0$, $[X_t \leq x] = \varnothing \in \mathscr{A}$).

Agora, seja T_1 o tempo até a primeira chegada. Então T_1 é função do resultado do experimento, pois

$$T_1(\omega) = \sup \{t: \ \omega(t) = 0\} = \min \{t: \ \omega(t) = 1\}.$$

É claro que o primeiro telefonema chega depois do instante $t \geq 0$ se, e somente se, não chega telefonema algum até o instante t, inclusive. Logo, $[T_1 \leq t] = [T_1 > t]^c = (A^0_{0,t})^c$ para $t \geq 0$, e T_1 é variável aleatória (se $t < 0$, $[T_1 \leq t] = \varnothing$).

Seja T_2 o tempo entre a primeira e a segunda chegada. É T_2 uma variável aleatória? Isso é mais difícil de provar. Consideremos o seguinte argumento:

Seja $Z = T_1 + T_2$ = tempo da segunda chegada. Então

$$[Z \leq z] = [Z > z]^c = (A^0_{0,z} \cup A^1_{0,z})^c \in \mathscr{A},$$

logo Z é variável aleatória. Por isso, $T_2 = Z - T_1$ é variável aleatória. (*Nota*: funções contínuas de variáveis aleatórias são variáveis aleatórias! Esse fato não será provado, mas sendo ele bem razoável, espera-se que o leitor o aceite).

DEFINIÇÃO 2.2. A *função de distribuição* da variável aleatória X, representada por F_X ou simplesmente por F, é definida por

$$F_X(x) = P(X \leq x), \ x \in \mathbb{R}.$$

Observação. Na literatura, a função de distribuição de X é freqüentemente chamada de *função de distribuição acumulada* de X. Muitos autores, entre os quais se encontram Gnedenko [11] e Breiman [6] (e as escolas russa e francesa), definem $F_X(x) = P(X < x)$, fazendo com que a função de distribuição seja contínua à esquerda, em vez de contínua à direita.

Propriedades. Se X é uma variável aleatória, sua função de distribuição F goza das seguintes propriedades:

F1. $x \le y \Rightarrow F(x) \le F(y)$, *i.e., F é não-decrescente.*

F2. *Se $x_n \downarrow x$, então $F(x_n) \downarrow F(x)$, i.e., F é contínua à direita.*

F3. *Se $x_n \downarrow -\infty$, então $F(x_n) \downarrow 0$. Se $x_n \uparrow +\infty$, então $F(x_n) \uparrow 1$.* (Logo podemos escrever $F(-\infty) = 0$, $F(+\infty) = 1$.)

Prova. F1. $$x \le y \Rightarrow [X \le x] \subset [X \le y]$$
$$\Rightarrow F(x) = P(X \le x) \le P(X \le y) = F(y).$$

F2. Se $x_n \downarrow x$, então $[X \le x_n]$ é uma seqüência decrescente de eventos aleatórios e $\bigcap_{n \ge 1} [X \le x_n] = [X \le x]$ (porque $X \le x$ se, e somente se, $X \le x_n \, \forall n$). Em outras palavras, $[X \le x_n] \downarrow [X \le x]$ e, pela continuidade de probabilidade, $F(x_n) = P(X \le x_n) \downarrow P(X \le x) = F(x)$.

F3. Se $x_n \downarrow -\infty$, então $[X \le x_n] \downarrow \varnothing$ e $F(x_n) = P(X \le x_n) \downarrow 0$. Se $x_n \uparrow +\infty$, então $[X \le x_n] \uparrow \Omega$ e $F(x_n) = P(X \le x_n) \uparrow 1$. □

Observação. Uma função de distribuição é monótona não-decrescente e portanto tem um número finito ou enumerável de pontos de descontinuidade. Além disto, todas as descontinuidades são do tipo salto. Pela continuidade à direita, o salto no ponto x é igual a

$$F(x) - F(x-) = F(x) - \lim_{n \to \infty} F\left(x - \frac{1}{n}\right) = P(X \le x) -$$

$$- \lim_{n \to \infty} P\left(X \le x - \frac{1}{n}\right) = \lim_{n \to \infty} \left(P(X \le x) - P\left(X \le x - \frac{1}{n}\right)\right).$$

Como $\left[X \le x - \frac{1}{n}\right] \subset [X \le x]$, temos

$$\left[x - \frac{1}{n} < X \le x\right] = [X \le x] - \left[X \le x - \frac{1}{n}\right] \in \mathcal{A}$$

e

$$P\left(x - \frac{1}{n} < X \le x\right) = P(X \le x) - P\left(X \le x - \frac{1}{n}\right).$$

Mas a seqüência de eventos $\left[x - \frac{1}{n} < X \le x\right]$ é decrescente, com

$$\bigcap_n \left[x - \frac{1}{n} < X \le x\right] = [X = x],$$

de modo que o evento $[X = x]$ é aleatório e

$$P(X = x) = \lim_{n \to \infty} P\left(x - \frac{1}{n} < X \le x\right)$$
$$= F(x) - F(x-).$$

Portanto, o salto de F no ponto x é igual a $P(X = x)$, e F é contínua no ponto x se, e somente se, $P(X = x) = 0$. Em outras palavras, a função de distribuição de X é contínua se, e somente se, para todo $x \in \mathbb{R}$ a probabilidade de X tomar o valor x é zero.

(*Exercício.* Prove que o número de saltos de F de tamanho $\ge \frac{1}{n}$ é $\le n$. Utilize esse resultado para provar diretamente que o número de saltos de F é finito ou enumerável.)

Questão: É toda função F satisfazendo F1, F2 e F3 a função de distribuição de alguma variável aleatória? *Resposta:* sim. A prova deste fato envolve conceitos da Teoria da Medida. Mas para termos uma idéia da prova, suponhamos que a função F satisfaça F_1, F_2 e F_3, e consideremos o problema de construir uma variável aleatória X tal que $F_X = F$. Se pudermos definir uma probabilidade P nos borelianos da reta tal que $P((-\infty, x]) = F(x) \, \forall x \in \mathbb{R}$, então bastará definir $X(\omega) = \omega \, \forall \omega \in \mathbb{R}$ (i.e., X será a função identidade). Pois neste caso, $F_X(x) = P(X \le x) = P((-\infty, x]) = F(x)$. O método para a construção de tal P é o seguinte:

Definir

$$P((-\infty, x]) = F(x),$$
$$P((x, \infty)) = 1 - F(x),$$
$$P((a, b]) = F(b) - F(a),$$

e definir P na álgebra de uniões finitas de tais intervalos por aditividade finita, por exemplo,

$$P((a, b] \cup (c, d]) = F(b) - F(a) + F(d) - F(c),$$

para $a < b \le c < d$. Depois de verificar que P é σ-aditiva nesta álgebra, pode-se estender P para $\mathcal{B} = \sigma$-álgebra dos borelianos, pelo Teorema da Extensão de Carathéodory.

Portanto, toda função F que satisfaça F1, F2 e F3 será chamada *função de distribuição.*

Observação. Uma função de distribuição pode corresponder a várias variáveis aleatórias no mesmo espaço de probabilidade $(\Omega, \mathscr{A}, P)$. Por exemplo, se $X \sim N(0,1)$, i.e., se X tem distribuição normal com parâmetros 0 e 1 (veja o exemplo 9 adiante), então $-X \sim N(0,1)$. Conseqüentemente, $F_X = F_{-X}$. No entanto, $P(X = -X) = P(2X = 0) = P(X = 0) = 0$.

Exemplos de funções de distribuição. Voltemos ao exemplo do processo de Poisson, e consideremos novamente a variável aleatória X_t, onde $t > 0$. Foi visto que

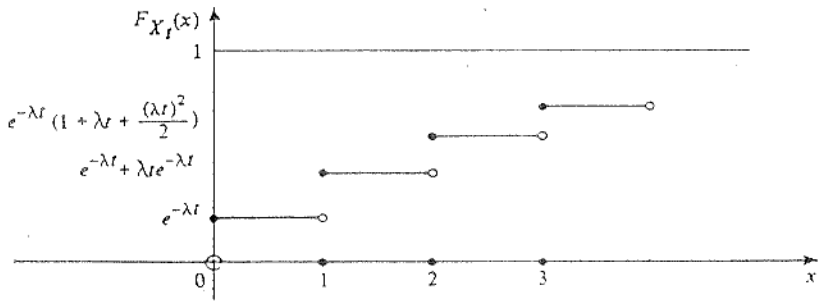
$$[X_t \le x] = \bigcup_{0 \le k \le x} [X_t = k] = \bigcup_{0 \le k \le x} A^k_{0,t}, \quad x \ge 0,$$

com $[X_t \le x] = \varnothing$ para $x < 0$ (X_t assume apenas valores não-negativos). Portanto, a função de distribuição de X_t satisfaz (veja § 1.3)

$$F_{X_t}(x) = \begin{cases} 0, & \text{se } x < 0 \\ \displaystyle\sum_{0 \le k \le x} P(X_t = k) = \sum_{k=0}^{[x]} \frac{e^{-\lambda t}(\lambda t)^k}{k!}, & \text{se } x \ge 0, \end{cases}$$

onde $[x]$ é a parte inteira de x (maior inteiro $\le x$).

Esta é a função de distribuição da *distribuição de Poisson* com parâmetro λt. O seu gráfico é:



Notemos que a função de distribuição de X_t cresce através de saltos nos valores possíveis de X_t, i.e., os números $0, 1, 2, \ldots$. O tamanho do salto em k é a probabilidade de X_t tomar o valor k, e a soma dos tamanhos de todos os saltos é igual a um. Esta propriedade é característica das variáveis aleatórias chamadas *discretas* (veja a definição a seguir), das quais as variáveis do tipo Poisson são, na prática, alguns dos principais exemplos.

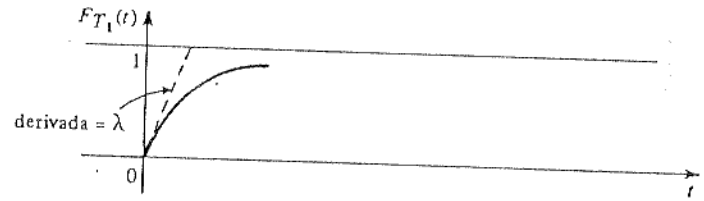
Outra variável aleatória considerada no exemplo 6 foi T_1, o tempo até a primeira chegada. Foi visto que

$$[T_1 \le t] = (A^0_{0,t})^c, \quad t \ge 0,$$

com $[T_1 \le t] = \varnothing$ para $t < 0$ (T_1 também assume apenas valores positivos). Portanto, temos

$$F_{T_1}(t) = \begin{cases} 0, & \text{se } t < 0 \\ 1 - P(A^0_{0,t}) = 1 - e^{-\lambda t}, & \text{se } t \ge 0. \end{cases}$$

Essa é a função de distribuição da *distribuição exponencial* com parâmetro λ; seu gráfico é



Notemos que a função de distribuição de T_1 é contínua, não havendo saltos, e que é derivável em todo ponto exceto em 0. Veremos adiante que disto decorre ser T_1 *absolutamente contínua.*

(*Exercício*: verifique diretamente da definição a seguir que T_1 é absolutamente contínua com densidade $f(t) = \lambda e^{-\lambda t}$, $t \ge 0$; $f(t) = 0$, $t < 0$.)

§2.2 Tipos de variáveis aleatórias

DEFINIÇÃO 2.3. (*a*) A variável aleatória X é *discreta* se toma um número finito ou enumerável de valores, i.e., se existe um conjunto finito ou enumerável $\{x_1, x_2, \ldots\} \subset \mathbb{R}$ tal que $X(\omega) \in \{x_1, x_2, \ldots\}\ \forall \omega \in \Omega$. A função $p(x_i)$ definida por $p(x_i) = P(X = x_i)$, $i = 1, 2, \ldots$, é chamada *função de probabilidade* (ou *função de freqüência*) de X.

(*b*) A variável aleatória X é (*absolutamente*) *contínua* se existe uma função $f(x) \ge 0$ tal que

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

Neste caso, dizemos que f é *função de densidade de probabilidade* de X ou simplesmente *densidade* de X.

Observações. (*a*) Se X é discreta, então $[X \le x] = \bigcup_{i:\, x_i \le x} [X = x_i]$, logo

$$F_X(x) = \sum_{i:\, x_i \le x} P(X = x_i) = \sum_{i:\, x_i \le x} p(x_i).$$

(Já verificamos que $[X = x_i]$ era evento aleatório quando vimos que o salto de F em x_i era igual a $P(X = x_i)$.)

(*b*) Se X é absolutamente contínua, então F_X, sendo uma integral indefinida de f, é contínua. Tecnicamente, a integral da Definição 2.3(b) é de Lebesgue, e X tem densidade se, e somente se, F_X é *absolutamente* contínua (i.e., F_X é a integral da sua derivada). Neste caso, $f(x) = F_X'(x)$ em todo ponto, exceto num conjunto de medida de Lebesgue nula (diz-se que $f = F_X'$ em quase toda parte). Um conjunto $B \subset \mathbb{R}$ tem medida de Lebesgue nula se tem comprimento zero, i.e., se para todo $\varepsilon > 0$, existem intervalos de comprimento total $< \varepsilon$ cuja união inclui B.

Uma função $f(x) \ge 0$ é densidade de alguma variável aleatória se, e somente se, $\int_{-\infty}^{\infty} f(x)\,dx = 1$, já que neste caso F definida por

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

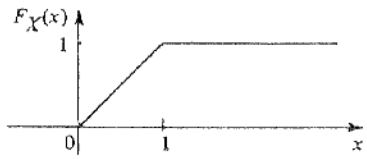
é função de distribuição, pois satisfaz F1, F2 e F3 (verifique!). Reciprocamente, se f é densidade então $\int_{-\infty}^{\infty} f(x)\,dx = 1$, pelo item (b) da definição e a propriedade F3.

Mas sem a Teoria da Medida, como vamos verificar se X tem densidade? Podemos usar o seguinte critério, válido em quase todo caso que surge na prática:

X tem densidade se F_X é (i) contínua e (ii) derivável por partes, i.e., se F_X é derivável no interior de um número finito ou enumerável de intervalos fechados cuja união é a reta $\mathbb{R}$. (Neste caso, a derivada é a densidade de X.) Em particular, X tem densidade se F_X é contínua e derivável em todo ponto exceto num número finito de pontos, ou se F_X é contínua e derivável em todo ponto a não ser nos inteiros.

Por exemplo, seja

$$F_X(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \le x \le 1 \\ 1, & x > 1. \end{cases}$$



FACULDADE DE CIÊNCIAS INSTITUTO NACIONAL MATEMÁTICA PURA E APLICADA BIBLIOTECA

Então X tem densidade, pois F_X é contínua e

$$F_X'(x) = \begin{cases} 1, & x \in (0, 1) \\ 0, & x \notin [0, 1]. \end{cases}$$

A densidade de X é dada por

$$f(x) = f_X(x) = \begin{cases} 1, & x \in (0, 1) \\ 0, & x < 0 \text{ ou } x > 1. \end{cases}$$

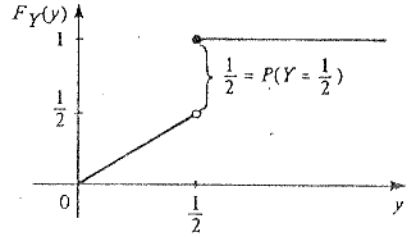
O valor de f nos pontos 0 e 1 é arbitrário, pois qualquer que seja $f(0)$ (ou $f(1)$), a integral $\int_{-\infty}^{x} f(t)\,dt$ é ainda igual a $F_X(x)$. Costuma-se definir ou $f(0) = f(1) = 1$ ou $f(0) = f(1) = 0$. (Outro exemplo de uma função de distribuição contínua e derivável por partes é a F_T, considerada anteriormente.)

Por outro lado, suponha que

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1, & x \ge 0. \end{cases}$$

Aqui X não tem densidade, pois F_X não é contínua. De fato X é uma variável aleatória discreta, e $P(X = 0) = 1$.

É fácil construir um exemplo de variável aleatória que não é discreta nem absolutamente contínua, mas sim uma mistura dos dois tipos. Por exemplo, seja tal que $X \sim U[0, 1]$ (leia-se "X tem distribuição uniforme em $[0, 1]$"), i.e., X tem a função de distribuição F_X cujo gráfico está acima. E seja $Y = \min(X, 1/2)$, i.e., Y é a variável aleatória definida por $Y(\omega) = \min(X(\omega), 1/2)$, $\omega \in \Omega$. (Y é variável aleatória, pois é função contínua da variável aleatória X). Então Y é do tipo "misto".



EXEMPLO 7. Uma função de distribuição de uma variável aleatória que não é discreta, contínua, ou mista. Nossa função será contínua, derivável em todo ponto menos num conjunto de medida de Lebesque nula, mas não será absolutamente contínua: vamos considerar a *função de Cantor*.

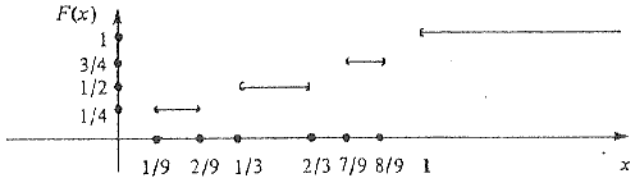


Gráfico da função de Cantor após as etapas 1 e 2.

Definamos $F(x) = 0$ para $x < 0$, $F(x) = 1$ para $x > 1$. Continuemos por etapas:

Etapa 1. Seja $F(x) = 1/2$ em $(1/3, 2/3)$. Então o valor de F nesse intervalo é a média dos valores nos dois intervalos vizinhos em que F já está definida ($(-\infty, 0)$ e $(1, \infty)$), e 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 $F(x) = 1/4$ em $(1/9, 2/9)$ e $F(x) = 3/4$ em $(7/9, 8/9)$. 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 (i.e., o conjunto onde F ainda não está definida) é o conjunto de Cantor, um conjunto de medida de Lebesgue zero (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 $1/2^n$. E 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, 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, e então F estará definida em toda a reta.

(*Exercício*: verifique que F é função de distribuição.)

Agora seja X uma variável aleatória cuja função de distribuição é F, a função de Cantor. Então X não é discreta (F é contínua), nem

do tipo misto (pela mesma razão), e nem contínua (X não tem densidade, pois $F'(x) = 0$ em C^c e $\int_{-\infty}^{x} F'(t)\,dt = 0$, i.e., F não é a integral da sua derivada, ou melhor, não é absolutamente contínua.) Dizemos que X é variável aleatória *singular*: uma variável aleatória X é chamada singular se F_X é contínua e $F_X'(x) = 0$ em quase toda parte, i.e., exceto em um conjunto de medida de Lebesgue nula.

Observemos agora que se F_X é a função de Cantor, então $P(X \in C) = 1$, onde C é o conjunto de Cantor. Com efeito,

$$C = \mathbb{R} - (-\infty, 0) - (1, \infty) - \underbrace{\left(\frac{1}{3}, \frac{2}{3}\right)}_{I_1} - \underbrace{\left(\frac{1}{9}, \frac{2}{9}\right)}_{I_2} - \underbrace{\left(\frac{7}{9}, \frac{8}{9}\right)}_{I_3} - \ldots =$$

$$\underbrace{\qquad\qquad\qquad\qquad}_{\text{1.ª etapa}} \qquad \underbrace{\qquad\qquad\qquad}_{\text{2.ª etapa}}$$

$$= (\cup I_n)^c.$$

Mas para todo n, $P(X \in I_n) = 0$, pois, por exemplo,

$$P(X \in I_3) = P\left(\frac{1}{3} < X < \frac{2}{3}\right) \leq P\left(\frac{1}{3} < X \leq \frac{2}{3}\right) =$$

$$= F\left(\frac{2}{3}\right) - F\left(\frac{1}{3}\right) = \frac{1}{2} - \frac{1}{2} = 0.$$

Então $P(X \in \cup I_n) = 0$ e portanto $P(X \in C) = 1$.

Podemos descrever o caso singular nos seguintes termos: X é singular se, e somente se, existe um conjunto B de comprimento zero tal que $P(X \in B) = 1$ *e* F_X é contínua (i.e., $P(X = x) = 0$ para todo $x \in \mathbb{R}$).

Vamos ver agora que toda variável aleatória é uma mistura dos três tipos: discreto, absolutamente contínuo e singular.

Seja X uma variável aleatória qualquer e seja F sua função de distribuição. Se $J = \{x_1, x_2, \ldots\}$ é o conjunto dos pontos de salto de F (se F for contínua, $J = \varnothing$), indiquemos com p_i o salto no ponto x_i, i.e.,

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

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.

Ocorre que uma função monótona possui derivada em quase toda parte. Seja f, então, a derivada de F, ou melhor:

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

Seja $F_{ac}(x) = \int_{-\infty}^{x} f(t)\,dt$. F_{ac} é não-decrescente, pois é integral indefinida de uma função não-negativa ($f \geq 0$ porque F é não-decrescente). A sua derivada é igual a f, pelo menos em quase toda parte, de modo que F_{ac} é absolutamente contínua (é a integral de sua derivada): 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 (F_{ac} é absolutamente contínua, logo contínua; $F - F_d$ é contínua, porque a subtração de F_d tira todos os saltos de F). 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, sendo uma função-degrau, possui derivada zero em quase toda parte. F_s é a *parte singular* de F, e

$$F = F_d + F_{ac} + F_s.$$

(*Observação.* F_s também é não-decrescente. Omitimos a prova, que depende da Teoria da Medida.)

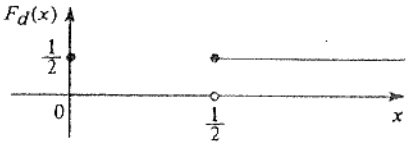
A discussão acima dá um método de *decompor* F em sua partes discreta, absolutamente contínua e singular. Consideremos um exemplo.

EXEMPLO 8. Suponha $X \sim U[0, 1]$ e $Y = \min(X, 1/2)$. Já vimos que

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

F_Y tem apenas um salto, em $x = 1/2$, e $p_1 =$ salto no ponto $1/2 = 1/2$. Logo

$$F_d(x) = \begin{cases} 0, & x < \dfrac{1}{2} \\ \dfrac{1}{2}, & x \geq \dfrac{1}{2}. \end{cases}$$



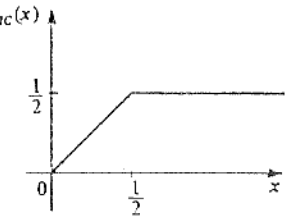
Diferenciando, temos

$$f(x) = F_Y'(x) = \begin{cases} 0, & x < 0 \text{ ou } x > \dfrac{1}{2} \\ 1, & 0 < x < \dfrac{1}{2} \end{cases}$$

$f(x) = 0$ se $x = 0$ ou $1/2$ (por definição).

Logo,

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



Como $F_d + F_{ac} = F_Y$, $F_s(x) = 0$ $\forall x$ e não há uma parte singular. Então Y é realmente uma mistura dos tipos discreto e contínuo (pode-se até dizer que metade da distribuição é discreta e concentrada no ponto $x = 1/2$ e a outra metade é absolutamente contínua e uniforme em $[0, 1/2]$). Na prática, é pouco provável que surja uma variável aleatória com uma parte singular, e quase todas as variáveis aleatórias que vamos considerar serão discretas, contínuas, ou misturas dos dois tipos.

§2.3 A distribuição de uma variável aleatória

Seja X uma variável aleatória em $(\Omega, \mathscr{A}, P)$. Por definição, $[X \leq x] \in \mathscr{A}$ $\forall x \in \mathbb{R}$. Em outras palavras, o evento $[X \in B]$ é aleatório e $P(X \in B)$ está definida se $B = (-\infty, x]$ para algum x. De fato, isso vale para todo boreliano B:

PROPOSIÇÃO 2.1. *Se X é variável aleatória em $(\Omega, \mathscr{A}, P)$, então o evento*

$$[X \in B] \stackrel{def}{=} \{\omega \in \Omega:\ X(\omega) \in B\}$$

é evento aleatório para todo boreliano B, i.e.,

$$[X \in B] \in \mathscr{A},\ \forall B \in \mathscr{B} = \sigma\text{-álgebra de Borel.}$$

Prova. Teoria da Medida. Mas podemos justificar (intuitivamente) a proposição: recorde que a σ-álgebra $\mathscr{B}$, dos borelianos, é a menor σ-álgebra contendo os intervalos. Vamos, então, verificar a conclusão da proposição para B intervalo:

(i) Se $B = (-\infty, b]$, então $[X \in B] \in \mathscr{A}$ pela Definição 2.1.

(ii) Se $B = (a, \infty)$, então $B = (-\infty, a]^c$ e $[X \in B] = [X \leq a]^c \in \mathscr{A}$, por (i). Neste caso, $P(X \in B) = P(X > a) = 1 - P(X \leq a) = 1 - F_X(a)$.

(iii) Se $B = (a, b]$, então $[X \in B] = [a < X \leq b] = [X \leq b] - [X \leq a] \in \in \mathscr{A}$, por (i). $P(X \in B) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$.

(iv) Se $B = (a, b)$, então $B = \bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n}\right]$ e $[X \in B] =$

$= \bigcup_{n=1}^{\infty} \left(a < X \leq b - \frac{1}{n}\right] \in \mathscr{A}$, por (iii). Neste caso,

$$P(X \in B) = \lim_{n \to \infty} P\left(a < X \leq b - \frac{1}{n}\right) = \lim_{n \to \infty} \left[F_X\left(b - \frac{1}{n}\right) - F_X(a)\right] =$$
$$= F_X(b-) - F_X(a).$$

De modo análogo, verifica-se que para todo intervalo B, $[X \in B] \in \in \mathscr{A}$ e $P(X \in B)$ é determinada pela função de distribuição F_X. E a mesma coisa vale se $B = \bigcup_{i=1}^{n} B_i$, onde os B_i são intervalos disjuntos, já que $[X \in B] = \bigcup_{i=1}^{n} [X \in B_i]$ e $P(X \in B) = \sum_{i=1}^{n} P(X \in B_i)$.

Então vale a conclusão da proposição na álgebra de uniões finitas de intervalos. Logo vale na σ-álgebra $\mathscr{B}$! Tecnicamente, para verificar se certa propriedade é válida para todo boreliano, basta verificar se (a) é válida para toda união finita de intervalos e (b) continua válida também para limites monótonos, i.e., se vale para B_n, para todo n, e $B_n \downarrow B$ ou $B_n \uparrow B$, então vale para B. Neste livro, não vamos nos preocupar com problemas técnicos desta natureza, limitando-nos em geral a verificar (a) e deixando (b) para o leitor interessado. Tomaremos isto como regra geral.

Salientamos outra implicação desta prova: as probabilidades $P(X \in B)$ são determinadas pela função de distribuição F_X. □

Observação. Se definimos $P_X(B) = P(X \in B)$, para B boreliano, então P_X é uma probabilidade em $\mathscr{B}$, porque os axiomas se verificam:

Axioma 1. $P_X(B) = P(X \in B) \geq 0$.

Axioma 2. $P_X(\mathbb{R}) = P(X \in \mathbb{R}) = 1$.

Axioma 3'. Se $B_1, B_2, \ldots \in \mathscr{B}$ são disjuntos, então

$$P_X(\cup B_n) = P(X \in \cup B_n) = P(\cup [X \in B_n]) =$$
$$= \sum_n P(X \in B_n) = \sum_n P_X(B_n).$$

Pelas observações feitas na prova anterior, P_X é determinada pela função de distribuição de X; por outro lado, é claro que a função de distribuição F_X é determinada por P_X, pois $F_X(x) = P(X \leq x) = = P_X((-\infty, x])$. Em outras palavras, F_X determina P_X, e vice-versa.

DEFINIÇÃO 2.4. A probabilidade P_X, definida na σ-álgebra de Borel por $P_X(B) = P(X \in B)$, é chamada *distribuição de* X.

Nos casos discreto e contínuo, podemos descrever a distribuição por meio da função de probabilidade ou densidade:

PROPOSIÇÃO 2.2. (*a*) *Se a variável aleatória X é discreta e toma valores somente no conjunto $\{x_1, x_2, \ldots\}$, então*

$$P_X(B) = \sum_{i:\, x_i \in B} P(X = x_i) = \sum_{i:\, x_i \in B} p(x_i), \quad \forall B \in B.$$

(*b*) *Se X é absolutamente contínua com densidade $f(x)$, então*

$$P_X(B) = \int_B f(x)\, dx, \quad \forall B \in \mathscr{B}.$$

Prova. (*a*) $[X \in B] = \bigcup_{i:\, x_i \in B} [X = x_i]$ e esses eventos são disjuntos. Logo

$$P_X(B) = \sum_{i:\, x_i \in B} P(X = x_i).$$

(*b*) Prova-se pela Teoria da Medida. É fácil verificar no caso de B intervalo; por exemplo, se $B = (a, b)$ então

$$P_X(B) = P(a < X < b) = \text{(pois } F_X \text{ é contínua e } P(X = b) = 0) =$$
$$= P(a < X \leq b) = F_X(b) - F_X(a) = \text{(pela Definição 2.3(b))} =$$
$$= \int_{-\infty}^{b} f(x)\, dx - \int_{-\infty}^{a} f(x)\, dx = \int_a^b f(x)\, dx = \int_B f(x)\, dx.$$

Se $B = \bigcup_{i=1}^{n} B_i$ é união finita de intervalos disjuntos, então

$$P_X(B) = \sum_{i=1}^{n} P_X(B_i) = \sum_{i=1}^{n} \int_{B_i} f(x)\, dx = \int_{\cup B_i} f(x)\, dx.$$

912723

pela aditividade da integral (por exemplo, se $a<b<c<d$, então

$$\int_{[a,b]\cup[c,d]} f(x)\,dx = \int_a^b f(x)\,dx + \int_c^d f(x)\,dx).$$

Agora é só aplicar nossa regra geral, adotada na prova da Proposição 2.1: como $\int_B f(x)\,dx = P_X(B)$ para todo B união finita de intervalos, vale também para todo B boreliano. □

Vimos, então, que a distribuição de X é determinada por qualquer das seguintes funções:

(1) A função de distribuição F_X.

(2) A densidade $f(x)$, se X é absolutamente contínua.

(3) A função de probabilidade $p(x_i)$, no caso discreto.

Veremos mais adiante (Capítulo 6) que é determinada também por:

(4) A função característica de X.

(1), (2), (3) e (4) serão chamadas *representações* da distribuição de X ou representações da *lei* de X (lei = distribuição). Para conhecer a distribuição de X, tanto faz conhecer qualquer das suas representações. Costuma-se escolher a representação mais conveniente para descrever a distribuição de uma dada variável aleatória. No caso contínuo, esta é geralmente a densidade:

EXEMPLO 9. A variável aleatória X possui distribuição normal "padrão" (notação: $X \sim N(0,1)$) se X tem densidade

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},\quad x\in\mathbb{R}.$$

Você sabe provar que f é realmente densidade? Um método de prova é o seguinte. Como f é não-negativa, é suficiente provar que $\int_{-\infty}^{\infty} f(x)\,dx = 1$. Para tanto, basta provar que o quadrado da integral é igual a 1. Mas temos

$$\left(\int_{-\infty}^{\infty} f(x)\,dx\right)^2 = \left(\int_{-\infty}^{\infty} f(x)\,dx\right)\times\left(\int_{-\infty}^{\infty} f(y)\,dy\right) =$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\,dx\,dy.$$

Exercício. Verifique que a expressão acima é igual a 1. (Sugestão: mude para coordenadas polares (θ,ρ), onde $x=\rho\cos\theta$, $y=\rho\,\text{sen}\,\theta$.)

Consideremos o seguinte problema: se $X\sim N(0,1)$ e $Y=\sigma X+\mu$, onde $\sigma>0$ e $\mu\in\mathbb{R}$, qual a distribuição de Y? Resposta: $Y\sim N(\mu,\sigma^2)$, ou seja, Y tem densidade

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(y-\mu)^2/2\sigma^2},\quad y\in\mathbb{R}.$$

É claro que Y é variável aleatória, pois $Y\le y$ se, e somente se, $X\le\dfrac{y-\mu}{\sigma}$, de modo que o evento $[Y\le y]$ é aleatório para todo y. Veremos agora, através de um resultado mais geral, como obter a densidade de Y a partir da densidade de X.

PROPOSIÇÃO 2.3. *Suponhamos que X possua densidade $f_X(x)$. Seja $Y=bx+c$, onde $b>0$ e $c\in\mathbb{R}$. Então Y tem densidade*

$$f_Y(y) = \frac{1}{b} f_X\left(\frac{y-c}{b}\right),\quad y\in\mathbb{R}.$$

(Notemos que no exemplo da normal, $c=\mu$, $b=\sigma$).

Prova. $F_Y(y) = P(Y\le y) = P(bX+c\le y) = P\left(X\le\dfrac{y-c}{b}\right) =$

$$= \int_{-\infty}^{(y-c)/b} f_X(x)\,dx = \text{(fazendo } t = bx+c) =$$

$$= \int_{-\infty}^{y} f_X\left(\frac{t-c}{b}\right)\frac{dt}{b}.$$

Pela Definição 2.3(b), $g(y)=\dfrac{1}{b}f_X\left(\dfrac{y-c}{b}\right)$ é a densidade de Y. □

(*Exercício.* Obtenha a densidade de $bX+c$ quando $b<0$. Compare com o caso $b>0$.)

Como conseqüência da proposição, vemos que quando $f(x)$ é densidade, podemos construir uma família de densidades $\{f_{b,c}\}$, definindo $f_{b,c}(x)=\dfrac{1}{b}f\left(\dfrac{x-c}{b}\right)$. Neste caso, c é chamado *parâmetro de locação* e b, *parâmetro de escala*. No Exemplo 9 (normal), μ é parâmetro de locação e σ é parâmetro de escala. (Para justificar a linguagem, notemos que multiplicação pelo fator b corresponde a uma mudança de escala, e adição da constante c resulta numa translação, i.e., mudança de locação.)

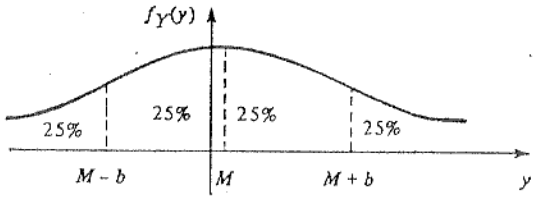
EXEMPLO 10. X tem distribuição de Cauchy (padrão) se X possui densidade

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}.$$

(Verifique se é realmente uma densidade!). Se $Y = bX + M$, onde $b > 0$, $M \in \mathbb{R}$, então Y possui densidade

$$f_Y(y) = \frac{b}{\pi(b^2 + (y-M)^2)}, \quad y \in \mathbb{R}.$$

Na família de distribuições de Cauchy, o parâmetro de locação M é a *mediana*, e o parâmetro de escala b representa a distância entre a mediana e o primeiro (ou o terceiro) *quartil*:



1º quartil 3º quartil

EXEMPLO 11. Distribuição gama. Quando $\alpha > 0$, a função $g(x) = x^{\alpha-1}e^{-x}$ é integrável no intervalo $(0, \infty)$, i.e., $\int_0^\infty x^{\alpha-1}e^{-x}\,dx < \infty$. Consideremos, então, a função gama, definida por $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$, $\alpha > 0$. Integrando por partes, vemos que $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ e, por indução, $\Gamma(n+1) = n!$ (pois $\Gamma(1) = \int_0^\infty e^{-x}\,dx = 1$).

É óbvio que $\Gamma(\alpha) > 0$; logo, f definida por

$$f(x) = \begin{cases} \dfrac{1}{\Gamma(\alpha)} x^{\alpha-1}e^{-x}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

é densidade (a da distribuição gama com parâmetros α e 1. *Notação*: $\Gamma(\alpha, 1)$). Se $Y = \dfrac{X}{\beta}$ e $X \sim \Gamma(\alpha, 1)$, então

$$f_Y(y) = \begin{cases} \dfrac{\beta}{\Gamma(\alpha)}(\beta y)^{\alpha-1}e^{-\beta y} = \dfrac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1}e^{-\beta y}, & y > 0 \\ 0, & y \le 0. \end{cases}$$

Esta é a densidade da distribuição gama com parâmetros α e β, indicada com $\Gamma(\alpha, \beta)$.

Neste caso, $\dfrac{1}{\beta}$ é parâmetro de escala (às vezes, escreve-se a densidade substituindo β por $\dfrac{1}{\beta}$, fazendo com que β seja parâmetro de escala). O parâmetro α é parâmetro de *configuração* (faça o gráfico da densidade para entender porque tem esse nome).

Observação. Quando $\alpha = 1$, a distribuição é a exponencial com parâmetro β, que tem densidade $f(y) = \beta e^{-\beta y}$, $y > 0$. Quando $\alpha = \dfrac{n}{2}$ e $\beta = \dfrac{1}{2}$, temos a distribuição qui-quadrado com n graus de liberdade (veja o §2.8).

No caso discreto, a representação mais conveniente da distribuição de X é, geralmente, a função de probabilidade. Já tratamos de um exemplo de variável aleatória discreta no §2.1, quando vimos a função de distribuição de uma variável tendo distribuição de Poisson com parâmetro λt. Com efeito, fazendo $t = 1$, vemos que se $X \sim$ Poisson (λ), então

$$F(x) = \begin{cases} 0, & \text{se } x < 0 \\ \displaystyle\sum_{0 \le k \le [x]} \frac{e^{-\lambda}\lambda^k}{k!}, & \text{se } x \ge 0. \end{cases}$$

Como o valor $p(k)$ da função de probabilidade é igual ao salto de F no ponto k, temos

$$p(k) = P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots.$$

EXEMPLO 12. Dizemos que X tem *distribuição binomial* com parâmetros n e p, onde n é um inteiro positivo e $0 < p < 1$, se

$$p(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Notação: $X \sim b(n, p)$. Esta é a distribuição que atribuímos, por exemplo ao número de caras obtido em n lançamentos de uma moeda tendo probabilidade p de dar cara.

Mais geralmente, consideremos um experimento básico em que estamos interessados em saber da ocorrência ou não de certo evento de probabilidade p, tal como a obtenção de cara na jogada de uma

moeda honesta $\left(p=\frac{1}{2}\right)$ ou a obtenção de 7 ou 11 no lançamento de dois dados equilibrados $\left(p=\frac{2}{9}\right)$. Se repetirmos o experimento básico n vezes, independentemente, e contar o número de ocorrências do evento de interesse, então este número terá distribuição $b(n, p)$.

Em tais casos, chamamos as repetições independentes do experimento básico de "ensaios" e, já que podemos interpretar cada ensaio como tendo apenas dois resultados possíveis (ocorrência ou não do evento de interesse), dizemos que se trata de um caso de "ensaios binomiais". Chamando o k-ésimo ensaio "sucesso" se ocorre o evento de interesse e "fracasso" se não ocorre, concluímos que a distribuição do *número de sucessos em n ensaios binomiais*, com probabilidade p de sucesso em cada ensaio, é $b(n, p)$. Por exemplo, a distribuição do número de caras (= sucessos) em n jogadas de uma moeda honesta é $b\left(n, \frac{1}{2}\right)$, e a distribuição do número de sucessos obtidos em n lançamentos de um par de dados equilibrados, onde a obtenção de uma soma de 7 ou 11 é considerada sucesso, é $b\left(n, \frac{2}{9}\right)$.

Observação. Na definição geral de ensaios binomiais, admitimos a possibilidade da probabilidade de sucesso variar com os ensaios. Um exemplo disto será visto no § 5.3 (exemplo 5). Quando a probabilidade de sucesso é a mesma p para todo ensaio, os ensaios binomiais são freqüentemente chamados *ensaios de Bernoulli.*

EXEMPLO 13. Ir lançando uma moeda, não necessariamente honesta, independentemente. Contar o número de lançamentos até o da primeira saída de cara, inclusive. Seja X esse número. Se p é a probabilidade de cara em um dado lançamento, então X tem função de probabilidade

$$p(k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots .$$

Dizemos que X tem *distribuição geométrica* com parâmetro p. A geométrica é a distribuição do tempo de espera até o primeiro sucesso em uma seqüência de ensaios de Bernoulli com probabilidade p de sucesso.

§2.4 Vetores aleatórios

Em muitos experimentos, o interesse do investigador recai sobre vários característicos numéricos do resultado do experimento. Um exemplo simples disto foi visto no exemplo 5, em que o resultado do experimento "escolher, ao acaso, um ponto do círculo unitário" era considerado como o valor de um par de variáveis aleatórias X e Y, as coordenadas (cartesianas) do ponto escolhido. Formalmente, tínhamos

$$\omega = (x, y) = (X(\omega), Y(\omega)), \ \omega \in \Omega = \{(x, y): \sqrt{x^2 + y^2} \leq 1\}.$$

No processo de Poisson, que representa o modelo probabilístico de um experimento bem mais complicado, é comum o experimentador se interessar pela análise simultânea de várias variáveis aleatórias. Por exemplo, ele poderia querer comparar o número acumulado de chamadas até a hora n, para $n = 1, 2, \ldots, 24$; i.e., analisar o fluxo horário de telefonemas durante o primeiro dia. Neste caso, com t expresso em horas, ele trabalharia com as variáveis aleatórias $X_1, X_2, \ldots, X_{24}$, onde $X_k(\omega) = \omega(k)$, $\omega \in \Omega =$ a classe de funções-escada do § 1.3.

Nestes exemplos, o interesse está em um vetor de variáveis aleatórias, todas definidas no mesmo espaço de probabilidade. Os vetores (X, Y), no caso do círculo, e $(X_1, \ldots, X_{24})$, no caso do processo de Poisson, são exemplos de *vetores aleatórios.*

DEFINIÇÃO 2.5. (*a*) Um vetor $\underset{\sim}{X} = (X_1, \ldots, X_n)$, cujos componentes são variáveis aleatórias definidas no mesmo espaço de probabilidade $(\Omega, \mathscr{A}, P)$, é chamado *vetor aleatório* (ou variável aleatória n-dimensional).

(*b*) A *função de distribuição* $F = F_{\underset{\sim}{X}} = F_{X_1, \ldots, X_n}$ de um vetor aleatório $\underset{\sim}{X} = (X_1, \ldots, X_n)$ é assim definida:

$$F(\underset{\sim}{x}) = F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n), \ \forall (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

F é também chamada *função de distribuição conjunta* das variáveis aleatórias $X_1, \ldots, X_n$.

Observação. O evento $[X_1 \leq x_1, \ldots, X_n \leq x_n] \overset{def}{=} \bigcap_{i=1}^{n} [X_i \leq x_i]$ é aleatório, já que as X_i são variáveis aleatórias e portanto $[X_i \leq x_i] \in \mathscr{A}\ \forall i$. Logo $F_{X_1, \ldots, X_n}$ está bem definida. Note que o vetor aleatório $\underset{\sim}{X}$ é uma função definida no espaço amostral Ω assumindo valores no $\mathbb{R}^n$, i.e., $\underset{\sim}{X}: \Omega \to \mathbb{R}^n$.

Propriedades da função de distribuição F de um vetor aleatório $(X_1, \ldots, X_n)$.

F1. $F(x_1, \ldots, x_n)$ *é não-decrescente em cada uma das variáveis.* Por exemplo, é não-decrescente em x_1: se $x < y$, então

$$F(x, x_2, \ldots, x_n) \leq F(y, x_2, \ldots, x_n).$$

Analogamente, é não-decrescente em x_2, em x_3, etc.

F2. $F(x_1, \ldots, x_n)$ *é contínua à direita em cada uma das variáveis.* Por exemplo, se $y_m \downarrow x_1$ quando $m \to \infty$, então

$$F(y_m, x_2, \ldots, x_n) \downarrow F(x_1, x_2, \ldots, x_n) \text{ quando } m \to \infty,$$

valendo resultados análogos quando $y_m \downarrow x_2$, $y_m \downarrow x_3$, etc.

F3. *Para todo* i,

$$\lim_{x_i \to -\infty} F(x_1, \ldots, x_n) = 0.$$

Também,

$$\lim_{\forall i,\, x_i \to +\infty} F(x_1, \ldots, x_n) = 1.$$

(Este é o limite quando todas as coordenadas convergem simultaneamente para $+\infty$.)

Prova. Como no caso unidimensional. Somente a propriedade F3 é um pouco diferente. É importante notar que se i é fixo, então $[X_1 \leq x_1, \ldots, X_{i-1} \leq x_{i-1},\ X_i \leq -m,\ X_{i+1} \leq x_{i+1}, \ldots, X_n \leq x_n] \downarrow \varnothing$, quando $m \to \infty$, para todo $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$. Mas

$$[X_1 \leq x_1, \ldots, X_{i+1} \leq x_{i-1}, X_i \leq m, X_{i+1} \leq x_{i+1}, \ldots, X_n \leq x_n] \uparrow$$
$$\uparrow [X_1 \leq x_1, \ldots, X_{i-1} \leq x_{i-1}, X_{i+1} \leq x_{i+1}, \ldots, X_n \leq x_n]$$

quando $m \to +\infty$. Em outras palavras, quando $x_i \to +\infty$ $F_{X_1, \ldots, X_n}$ converge para a função de distribuição conjunta das $n-1$ variáveis aleatórias $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$. Finalmente, quando todos os x_i convergem simultaneamente para $+\infty$ (i.e., $x_i \to +\infty\ \forall i$), então o evento $\bigcap_{i=1}^{n} [X_i \leq x_i]$ converge para o evento certo Ω, e $F(x_1, \ldots, x_n)$ converge para 1. □

Para $n \geq 2$, as propriedades F1, F2 e F3 *não* são suficientes para que F seja uma função de distribuição:

EXEMPLO 14. Uma função $F_0: \mathbb{R}^2 \to \mathbb{R}$ que satisfaz F1, F2 e F3 mas não é a função de distribuição de um vetor aleatório (X, Y). Seja F_0 a seguinte função definida no plano:

$$F_0(x, y) = \begin{cases} 1 \text{ se } x \geq 0 \text{ e } y \geq 0 \text{ e } x + y \geq 1 \\ 0 \text{ caso contrário.} \end{cases}$$

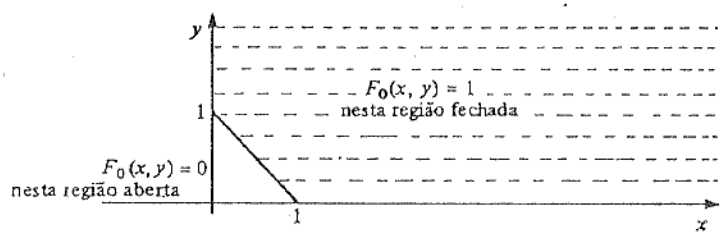


Gráfico de F_0.

É claro que as propriedades F1, F2 e F3 estão satisfeitas. Mas F_0 não é função de distribuição de um vetor aleatório (X, Y). Se fosse, então teríamos a contradição

$$0 \leq P(0 < X \leq 1, 0 < Y \leq 1) \overset{(*)}{=} F_0(1, 1) - F_0(1, 0) - F_0(0, 1) + F_0(0, 0) = 1 - 1 - 1 + 0 = -1.$$

Para verificar a equação $(*)$ acima, basta notar que quando F é a função de distribuição de um vetor aleatório (X, Y), temos

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

e, finalmente,

$$F(1, 1) - F(1, 0) - F(0, 1) + F(0, 0) = P(X \leq 1, 0 < Y \leq 1) - P(X \leq 0, 0 < Y \leq 1) = P(0 < X \leq 1, 0 < Y \leq 1).$$

De fato, se $a_1 < b_1$ e $a_2 < b_2$ e F é função de distribuição de (X, Y), então temos

$$0 \leq P(a_1 < X \leq b_1, a_2 < Y \leq b_2) = F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2).$$

Podemos descrever esta propriedade por meio de operadores de diferença. Com efeito, para $I=(a, b]$ e $g: \mathbb{R}^k \to \mathbb{R}$ definamos

$$\Delta_I g(x_1, \ldots, x_k) = g(x_1, \ldots, x_{k-1}, b) - g(x_1, \ldots, x_{k-1}, a).$$

A propriedade, então, é a seguinte, quando F é função de distribuição do vetor aleatório (X, Y): se $I_1=(a_1, b_1]$ e $I_2=(a_2, b_2]$, então

$$\Delta_{I_1}\Delta_{I_2} F(x, y) = \Delta_{I_1}[F(x, b_2) - F(x, a_2)] =$$
$$= F(b_1, b_2) - F(b_1, a_2) - [F(a_1, b_2) - F(a_1, a_2)] \geq 0.$$

Para n geral, a nova propriedade é:

F4. $\Delta_{I_1} \ldots \Delta_{I_n} F(x_1, \ldots, x_n) \geq 0, \forall I_k = (a_k, b_k], a_k < b_k, k = 1, \ldots, n.$

Essa propriedade nada mais é, portanto, que a formulação, por meio da função de distribuição F, da propriedade $P(a_1 < X_1 \leq b_1, \ldots, a_n < X_n \leq b_n) \geq 0$. Acontece que uma função satisfazendo F1, F2, F3 e F4 é realmente a função de distribuição de um vetor aleatório, i.e., as quatro propriedades são suficientes para caracterizar funções de distribuição (referência: Breiman [5], § 2.5).

DEFINIÇÃO 2.6. Uma função $F: \mathbb{R}^n \to \mathbb{R}$ que satisfaz as propriedades F1, F2, F3 e F4 é chamada *função de distribuição n-dimensional* (ou n-variada).

Os tipos discreto e absolutamente contínuo têm os seguintes análogos no caso multivariado:

DEFINIÇÃO 2.7. (*a*) Se o vetor aleatório $(X_1, \ldots, X_n)$ toma somente um número finito ou enumerável de valores, é chamado *discreto*.

(*b*) Seja $(X_1, \ldots, X_n)$ um vetor aleatório e F sua função de distribuição. Se existe uma função $f(x_1, \ldots, x_n) \geq 0$ tal que

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

então f é chamada *densidade* do vetor aleatório $(X_1, \ldots, X_n)$ ou *densidade conjunta* das variáveis aleatórias $X_1, \ldots, X_n$, e neste caso, dizemos que $(X_1, \ldots, X_n)$ é (absolutamente) *contínuo*.

São válidas as extensões para o caso n-dimensional das Proposições 2.1 e 2.2 e da Definição 2.4. Isto é:

Seja $\underset{\sim}{X} = (X_1, \ldots, X_n)$ um vetor aleatório no espaço de probabilidade $(\Omega, \mathscr{A}, P)$.

PROPOSIÇÃO 2.1'. $[\underset{\sim}{X} \in B] \in \mathscr{A}\ \forall B \in \mathscr{B}^n$, *onde* $\mathscr{B}^n$ *é a* σ*-álgebra de Borel no* $\mathbb{R}^n$.

(*Observação.* A σ-álgebra de Borel no $\mathbb{R}^n$ é a menor σ-álgebra contendo todo retângulo n-dimensional, ou seja, a σ-álgebra gerada pelos retângulos. Pelo método de prova da Proposição 2.1, pode-se ver que $[\underset{\sim}{X} \in B]$ é evento aleatório se B é retângulo ou união finita de retângulos. Por nossa regra geral, vale então para todo B boreliano. Notemos, por exemplo, que qualquer região aberta A no plano é união enumerável de retângulos; portanto, tem sentido falar na probabilidade de (X, Y) pertencer a A, se (X, Y) é vetor aleatório.)

DEFINIÇÃO 2.4'. A probabilidade definida em $\mathscr{B}^n$ por $P(\underset{\sim}{X} \in B)$ é chamada *distribuição* de $\underset{\sim}{X}$ ou *distribuição conjunta* de $X_1, \ldots, X_n$.

Notação. $P_{\underset{\sim}{X}}(B) = P(\underset{\sim}{X} \in B)$, $P_{\underset{\sim}{X}}$ é a distribuição de $\underset{\sim}{X}$.

PROPOSIÇÃO 2.2. (*a*) *Se o vetor aleatório* $\underset{\sim}{X}$ *é discreto, então*

$$P_{\underset{\sim}{X}}(B) = \sum_{i:\, \underset{\sim}{x}_i \in B} P(\underset{\sim}{X} = \underset{\sim}{x}_i), \forall B \in \mathscr{B}^n.$$

(*b*) *Se* $\underset{\sim}{X}$ *é contínuo com densidade* $f(x_1, \ldots, x_n)$, *então*

$$P_{\underset{\sim}{X}}(B) = P(\underset{\sim}{X} \in B) = \int \cdots \int_B f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n.$$

(As provas das proposições 2.1' e 2.2' são análogas às anteriores.)

§2.5 Independência

Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias definidas no *mesmo* espaço de probabilidade $(\Omega, \mathscr{A}, P)$, de modo que $\underset{\sim}{X} = (X_1, \ldots, X_n)$ é um vetor aleatório em $(\Omega, \mathscr{A}, P)$. Informalmente, as X_i são independentes se, e somente se, quaisquer eventos determinados por qualquer grupo de variáveis aleatórias distintas são independentes. Por exemplo, $[X_1 < 5]$ e $[X_2 > 9]$ são independentes; $[X_1 < 5]$, $[X_2 > 9]$ e $[0 < X_5 \leq 3]$ são independentes (se $n \geq 5$); etc.

DEFINIÇÃO 2.8. As variáveis aleatórias $X_1, \ldots, X_n$ são (coletivamente) *independentes* se

$$P(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) = \prod_{i=1}^{n} P(X_i \in B_i), \forall B_i \in \mathscr{B}, i = 1, \ldots, n.$$

Observações (1) Essa definição equivale à definição informal, pois, por exemplo, $P(X_1 \in B_1, X_2 \in B_2) = P(X_1 \in B_1, X_2 \in B_2, X_3 \in \mathbb{R}, \ldots, X_n \in \mathbb{R}) = P(X_1 \in B_1)\ P(X_2 \in B_2) \cdot 1 \cdots 1 = P(X_1 \in B_1)\ P(X_2 \in B_2)$.

(2) Usando o mesmo raciocínio do item (1) (ou a definição informal), vemos que para toda família de variáveis aleatórias independentes, qualquer subfamília é também formada por variáveis independentes. Por exemplo, se X, Y e Z são independentes, então X e Y também o são. (Essa é uma propriedade "hereditária" de variáveis independentes, segundo Breiman [6].)

(3) Ocorre que as variáveis aleatórias $X_1, \ldots, X_n$ são independentes se sua função de distribuição conjunta fatora e é o produto das funções de distribuição individuais. De fato, temos a seguinte

PROPOSIÇÃO 2.4. (*Critério para independência*).
(*a*) *Se* $X_1, \ldots, X_n$ *são independentes, então*

$$F_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i), \forall (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

(*b*) *Reciprocamente, se existem funções* $F_1, \ldots, F_n$ *tais que*

$$\lim_{x \to -\infty} F_i(x) = 1 \text{ para todo } i \text{ e}$$

$$F_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_i(x_i), \forall (x_1, \ldots, x_n) \in \mathbb{R}^n,$$

então $X_1, \ldots, X_n$ *são independentes e* $F_i = F_{X_i}, \forall\, i = 1, \ldots, n$.

(Em outras palavras, $X_1, \ldots, X_n$ são independentes se, e só se, sua função de distribuição conjunta fatora e cada fator converge para 1 em $+\infty$. No item (b), notemos que *não* é preciso verificar se F_i é função de distribuição; basta verificar se $F_i(x) \to 1$ quando $x \to +\infty$, $\forall i$.)

Prova. (a) Suponhamos $X_1, \ldots, X_n$ independentes. Então

$$F_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = P(X_1 \le x_1, \ldots, X_n \le x_n) =$$
$$= P(X_1 \in (-\infty, x_1], \ldots, X_n \in (-\infty, x_n]) =$$

$$= (\text{por hipótese}) = \prod_{i=1}^{n} P(X_i \in (-\infty, x_i]) =$$
$$= \prod_{i=1}^{n} P(X_i \le x_i) = \prod_{i=1}^{n} F_{X_i}(x_i), \forall\, (x_1, \ldots, x_n).$$

(b) $F_{X_i}(x_i) = P(X_i \le x_i) = \lim_{m \to +\infty} P(X_1 \le m, \ldots, X_{i-1} \le m, X_i \le x_i, X_{i+1} \le m, \ldots, X_n \le m) = \lim_{m \to +\infty} F_{X_1, \ldots, X_n}(\underbrace{m, \ldots, m}_{i-1 \text{ vezes}}, x_i, \underbrace{m, \ldots, m}_{n-i \text{ vezes}})$.

Pela hipótese do item (b), temos então

$$F_{X_i}(x_i) = \lim_{m \to +\infty} \left(\prod_{j=1}^{i-1} F_j(m)\, F_i(x_i) \prod_{j=i+1}^{n} F_j(m) \right) = F_i(x_i),$$

já que $\lim_{m \to \infty} F_j(m) = 1$. Logo F_i é a função de distribuição de $X_i (F_i = F_{X_i})$. Terminamos a prova com a Teoria da Medida. (A idéia da prova: queremos ver que

$$P(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{i=1}^{n} P(X_i \in B_i), \forall\, B_i \in \mathscr{B}.$$

Acabamos de provar que

$$F_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i),$$

i.e., vale o resultado se os B_i são do tipo $(-\infty, x_i]$. Vamos fazer uma verificação para $B_i = (a_i, b_i]$:

$$P(X_1 \in B_1, \ldots, X_n \in B_n) = P(a_1 < X_1 \le b_1, \ldots, a_n < X_n \le b_n) =$$
$$= \Delta_{I_1}^{1} \ldots \Delta_{I_n}^{n} F(x_1, \ldots, x_n) = \Delta_{I_1}^{1} \ldots \Delta_{I_n}^{n} (F_{X_1}(x_1) \ldots F_{X_n}(x_n)) =$$
$$= [F_{X_1}(b_1) - F_{X_1}(a_1)] \times \ldots \times [F_{X_n}(b_n) - F_{X_n}(a_n)] =$$
$$= \prod_{i=1}^{n} P(a_i < X_i \le b_i) = \prod_{i=1}^{n} P(X_i \in B_i).$$

Já que qualquer intervalo é limite de intervalos do tipo $(a, b]$, o resultado é válido se os B_i são intervalos quaisquer. Por aditividade, vale se os B_i são uniões finitas de intervalos. Para verificar que vale para todos os borelianos B_i, utilize o argumento sugerido na prova da Proposição 2.1.) □

No caso contínuo, o critério pode ser escrito da seguinte maneira:

PROPOSIÇÃO 2.5. (*Critério para independência no caso contínuo*). (*a*) *Se* $X_1, \ldots, X_n$ *são independentes e possuem densidades* $f_{X_1}, \ldots, f_{X_n}$, *então a função*

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i), \quad (x_1, \ldots, x_n) \in \mathbb{R}^n,$$

é densidade conjunta das variáveis aleatórias $X_1, \ldots, X_n$, *i.e.*, $f = f_{X_1, \ldots, X_n}$.

(*b*) *Reciprocamente, se* $X_1, \ldots, X_n$ *têm densidade conjunta* f *satisfazendo*

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_i(x_i), \ \forall (x_i, \ldots, x_n) \in \mathbb{R}^n,$$

onde $f_i(x) \geq 0$ *e* $\int_{-\infty}^{\infty} f_i(x)\,dx = 1$, $\forall i$, *então* $X_1, \ldots, X_n$ *são independentes e* f_i *é a densidade de* X_i, *para* $i = 1, \ldots, n$.

Prova. (a) Se $X_1, \ldots, X_n$ são independentes, então

$$F_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \text{(pela Proposição 2.4(a))} = \prod_{i=1}^{n} F_{X_i}(x_i) =$$

$$= \text{(pela definição de densidade)} = \prod_{i=1}^{n} \int_{-\infty}^{x_i} f_{X_i}(t_i)\,dt_i =$$

$$= \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_{X_1}(t_1) \ldots f_{X_n}(t_n)\,dt_1 \ldots dt_n.$$

Logo $\prod_{i=1}^{n} f_{X_i}$ é densidade conjunta de $X_1, \ldots, X_n$, pela Definição 2.7(b).

(b) $F_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(t_1, \ldots, t_n)\,dt_1 \ldots dt_n =$

$$= \text{(por hipótese)} = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_1(t_1) \ldots f_n(t_n)\,dt_1 \ldots dt_n =$$

$$= \prod_{i=1}^{n} \int_{-\infty}^{x_i} f_i(t_i)\,dt_i.$$

Definindo $F_i(x_i) = \int_{-\infty}^{x_i} f_i(t_i)\,dt_i$, temos, por hipótese, $\lim_{x_i \to +\infty} F_i(x_i) = 1$. A Proposição 2.4(b) implica que $X_1, \ldots, X_n$ são independentes e $F_i = F_{X_i}$, logo f_i é densidade de X_i. □

(Veja o exercício 22 para um critério para independência no caso discreto.)

EXEMPLO 15. 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 $\rho = 0$, então a densidade fatora:

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

Portanto, pela Proposição 2.5(b), se $\rho = 0$ então X e Y são independentes e $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$.

Se $\rho \neq 0$, então X e Y *não* são independentes, pois sua densidade conjunta não é produto das densidades marginais (i.e., $f \neq f_X f_Y$). Com efeito, vamos calcular as densidades de X e Y, usando a seguinte

PROPOSIÇÃO 2.6. (*a*) *Se* $F(x, y)$ *é a função de distribuição conjunta de* X *e* Y, *então a função de distribuição de* X *é*

$$F_X(x) = \lim_{y \to +\infty} F(x, y) \overset{def}{=} F(x, +\infty).$$

F_X assim obtida chama-se *função de distribuição marginal* de X.

(*b*) *Se* $f(x, y)$ *é densidade conjunta de* X *e* Y, *então* X *tem densidade dada por*

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

f_X assim obtida chama-se *densidade marginal* de X.

Prova. O item (a) já foi verificado no § 2.4 durante a prova da propriedade F3. A verificação do item (b) é deixada para o leitor (verifique se $\int_{-\infty}^{\infty} f(x, y)\,dy$ satisfaz a definição de densidade de X.) □

Observação. A Proposição 2.6 possui uma extensão natural para distribuições de vetores aleatórios de dimensão maior que 2.

Por exemplo, se $F(x, y, z)$ é a função de distribuição conjunta de X, Y e Z, então a função de distribuição marginal de (X, Y) é $F(x, y, +\infty)$, e a de X é $F(x, +\infty, +\infty)$; se $f(x, y, z)$ é a densidade conjunta de X, Y, Z, então a densidade marginal de (X, Y) é $\int_{-\infty}^{\infty} f(x, y, z)\, dz$ e a de X é $f_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y, z)\, dy\, dz$. No caso geral, obtemos a função de distribuição (densidade) marginal de uma subfamília das n variáveis aleatórias, fazendo todas as outras variáveis convergirem para $+\infty$ na função de distribuição conjunta (integrando a densidade conjunta de $-\infty$ até $+\infty$ em todas as outras variáveis).

Para um resultado análogo no caso discreto, veja o exercício 2.

Voltando ao exemplo 15, vamos calcular a densidade marginal de X, i.e., $f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$. Colocando em evidência os fatores que não dependem de y, e completando o quadrado do expoente do restante, temos

$$f_X(x) = \frac{1}{2\pi\,\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\} \times$$
$$\times \left[\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_2}{\sigma_2}\right) - \rho\left(\frac{x-\mu_1}{\sigma_1}\right)\right]^2\right\} dy.\right.$$

Falta o fator $\dfrac{1}{\sigma_2\sqrt{2\pi(1-\rho^2)}}$ para fazer do integrando a densidade da distribuição $N\left(\mu_2 + \rho\,\dfrac{\sigma_2}{\sigma_1}(x-\mu_1), (1-\rho^2)\,\sigma_2^2\right)$. Portanto, esta última integral tem o valor $\sqrt{2\pi(1-\rho^2)}\ \sigma_2$ e temos

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma_1} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\},\quad x \in \mathbb{R},$$

i.e., $X \sim N(\mu_1, \sigma_1^2)$. Analogamente, temos $Y \sim N(\mu_2, \sigma_2^2)$.

Segue-se que $\rho \neq 0$ implica que X e Y realmente não são independentes, pela Proposição 2.5(a): se o fossem, então o produto $f_X \cdot f_Y$ seria a densidade conjunta, mas é óbvio que $f \neq f_X \cdot f_Y$. (O parâmetro ρ é o *coeficiente de correlação* entre X e Y: veja o § 3.6).

Observação. Um ponto técnico: a densidade não é univocamente definida, pois podemos mudar o valor da densidade em um conjunto de medida de Lebesgue nula sem afetar o valor da integral na definição de densidade. Isto é, se f é densidade do vetor aleatório $\underset{\sim}{X}$ e $g = f$ em quase toda parte, então g também é densidade de $\underset{\sim}{X}$. Já vimos um exemplo disso, no caso de variáveis aleatórias (logo após a Definição 2.3), quando consideramos a densidade da distribuição uniforme em $[0, 1]$, aceitando as duas "versões" usuais da densidade (uma com $f(0) = f(1) = 0$, a outra com $f(0) = f(1) = 1$).

Na prática, podemos ignorar este problema técnico e tratar todas as versões da densidade como equivalentes, porque qualquer versão serve para obtermos a distribuição. Notemos, contudo, que para concluir que $f \neq f_X \cdot f_Y$ no exemplo acima, não é suficiente achar somente um ponto (x, y) em que temos a desigualdade, mas precisamos provar que $f(x, y) \neq f_X(x) f_Y(y)$ em um conjunto de área estritamente positiva (i.e., medida de Lebesgue > 0).

EXEMPLO 16. Seja $G \subset \mathbb{R}^n$ uma região tal que Vol $G > 0$, onde Vol G é o volume n-dimensional de G, de modo que Vol $G = \int_G \ldots \int 1\, dx_1 \ldots dx_n$. (Quando $n = 2$, por exemplo, Vol G = área G.) Dizemos que $\underset{\sim}{X} = (X_1, \ldots, X_n)$ é *uniformemente distribuído* em G se $\underset{\sim}{X}$ possui densidade

$$f(x_1, \ldots, x_n) = \begin{cases} \dfrac{1}{\text{Vol } G}, & (x_1, \ldots, x_n) \in G \\ 0, & (x_1, \ldots, x_n) \notin G \end{cases}$$

i.e., $f = \dfrac{I_G}{\text{Vol } G}$, onde I_G é o indicador de G.

(*Observação.* O *indicador*, ou função indicadora, de um conjunto G é a função que toma o valor 1 em G e toma o valor 0 fora de G, i.e., $I_G(\underset{\sim}{x}) = 1$ se $\underset{\sim}{x} \in G$, $I_G(\underset{\sim}{x}) = 0$ se $\underset{\sim}{x} \in G^c$.)

Neste caso, $\underset{\sim}{X}$ tem a *distribuição uniforme* em G, dada por

$$P(\underset{\sim}{X} \in B) = \int_B \ldots \int f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = \int_B \ldots \int \frac{I_G}{\text{Vol } G} =$$
$$= \frac{\text{Vol}(B \cap G)}{\text{Vol } G},\quad B \in \mathscr{B}^n.$$

Notação: $\underset{\sim}{X} \sim U(G)$. Quando a distribuição de $\underset{\sim}{X}$ é uniforme em G, todos os pontos de G são, de certa maneira, "equiprováveis", pois têm o mesmo peso relativo, representado pela densidade.

Esta definição pode ser usada também no caso de $n = 1$, com Vol G = comprimento G. Por exemplo, X tem distribuição uniforme em $[0, 1]$ se possui densidade $f = \dfrac{I_{[0,1]}}{\text{comprimento } [0, 1]} = I_{[0,1]}$.

Se G é *retângulo*, então $X_1, \ldots, X_n$ são independentes e cada uma é uniformemente distribuída. Por exemplo, seja $G = \prod_{i=1}^{n} [a_i, b_i]$. Então

$$f(x_1, \ldots, x_n) = \frac{1}{\prod_{i=1}^{n} (b_i - a_i)} I_G(x_1, \ldots, x_n) = \prod_{i=1}^{n} \left\{ \frac{1}{b_i - a_i} I_{[a_i, b_i]}(x_i) \right\},$$

onde a última igualdade se justifica por:

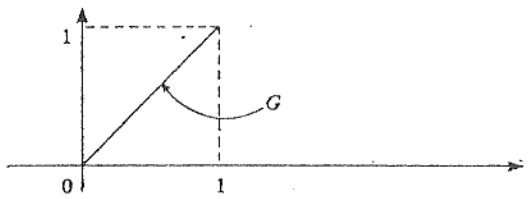
$$(x_1, \ldots, x_n) \in G \Leftrightarrow x_i \in [a_i, b_i] \ \forall i,$$

ou seja,

$$I_G(x_1, \ldots, x_n) = 1 \Leftrightarrow I_{[a_i, b_i]}(x_i) = 1, \text{ para todo } i = 1, \ldots, n.$$

Como $\frac{1}{b_i - a_i} I_{[a_i, b_i]}$ é densidade da distribuição $U[a_i, b_i]$, a Proposição 2.5(b) diz que $X_1, \ldots, X_n$ são independentes e $X_i \sim U[a_i, b_i]$.

Se Vol $G = 0$, não se pode usar a definição dada acima para definir a distribuição uniforme em G. Mas em certos casos, tal distribuição pode ser definida de maneira bem intuitiva. Para ilustrar este conceito, seja G a diagonal do quadrado unitário no plano:



Como você interpretaria "o vetor (X, Y) é uniformemente distribuído em G"? É claro que com isto queremos dizer que para todo boreliano B no plano,

$$P_{X,Y}(B) = P((X, Y) \in B) = \frac{\text{comprimento } (G \cap B)}{\sqrt{2}}.$$

Notemos que esta distribuição é *singular*; não existe uma densidade conjunta. (Suponhamos que exista uma densidade conjunta, $f(x, y)$. Então $P_{X,Y}(G) = 1 \Rightarrow \int_G \int f(x, y)\, dx\, dy = 1$. Mas área $G = 0$, logo $\int_G \int f(x, y)\, dx\, dy = 0$. Absurdo.)

É fácil verificar que $X \sim U[0, 1]$ e $Y \sim U[0, 1]$. Logo, fica provado que *se X tem densidade e Y também, não é necessariamente verdadeiro que X e Y possuam uma densidade conjunta.* Já sabemos, contudo, que vale a recíproca (Proposição 2.6(b)): se X e Y têm densidade conjunta, então existem as densidades marginais, com $f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$ e $f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$.

Cabe notar aqui que X e Y são discretas se, e somente se, (X, Y) é discreto. (Verifique!) Logo, temos o seguinte esquema, onde $X_1, \ldots, X_n$ são variáveis aleatórias em $(\Omega, \mathscr{A}, P)$:

$$X_1, \ldots, X_n \text{ discretas} \Leftrightarrow (X_1, \ldots, X_n) \text{ discreto}$$

$$X_1, \ldots, X_n \text{ absolutamente contínuas} \not\Leftarrow (X_1, \ldots, X_n) \text{ absolutamente contínuo.}$$

Observamos que sob a hipótese adicional de independência, temos equivalência nos dois casos, pois $X_1, \ldots, X_n$ independentes *e* absolutamente contínuas $\Rightarrow (X_1, \ldots, X_n)$ absolutamente contínuo, pela Proposição 2.5(a).

§ 2.6 Distribuições de funções de variáveis e vetores aleatórios

Seja $\underset{\sim}{X} = (X_1, \ldots, X_n)$ um vetor aleatório em $(\Omega, \mathscr{A}, P)$, e consideremos o problema de determinar a distribuição de $Y = g(X_1, \ldots, X_n)$. Este problema inclui o problema de determinar a distribuição da função de uma variável aleatória, ou seja, de $Y = g(X)$, pois uma variável aleatória é vetor aleatório unidimensional (i.e., $n = 1$).

Observação. Para que Y seja uma variável aleatória, vamos supor que g seja mensurável a Borel, i.e.,

$$g^{-1}(B) = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : g(x_1, \ldots, x_n) \in B\} \in \mathscr{B}^n, \forall B \in \mathscr{B}.$$

Toda função que se pode visualizar é mensurável a Borel — em particular, toda função contínua o é — e não vamos nos preocupar com esta questão.

Formalmente, o problema é de fácil solução, pois a função de distribuição de Y é

$$F_Y(y) = P(Y \leq y) = P(g(X_1, \ldots, X_n) \leq y),$$

e esta última probabilidade pode ser calculada por meio da distribuição conjunta de $X_1, \ldots, X_n$: se definirmos

$$B_y = \{(x_1, \ldots, x_n) : g(x_1, \ldots, x_n) \leq y\},$$

então $g(X_1, \ldots X_n) \le y$ se, e somente se, $(X_1, \ldots, X_n) \in B_y$, de modo que

$$F_Y(y) = P((X_1, \ldots, X_n) \in B_y) = P_{X_1, \ldots, X_n}(B_y).$$

Em outras palavras, conhecendo a distribuição conjunta de $X_1, \ldots, X_n$, podemos obter a distribuição de qualquer função (mensurável) das X_i, pelo menos teoricamente.

Quando $\underset{\sim}{X}$ é discreto, o problema é realmente de fácil solução, pois neste caso Y também é discreta (por quê?) e para obtermos sua função de probabilidade precisamos apenas somar os valores relevantes da função de probabilidade de $\underset{\sim}{X}$. Especificamente, se a função de probabilidade de $\underset{\sim}{X}$ é $p_{\underset{\sim}{X}}(\underset{\sim}{x}_i)$, $i = 1, 2, \ldots$, e se y_j é um valor possível de Y (i.e., um dos $g(x_i)$), então

$$p_Y(y_j) = \sum_{i:\, g(\underset{\sim}{x}_i) = y_j} p_{\underset{\sim}{X}}(\underset{\sim}{x}_i).$$

Vamos ver alguns exemplos dos cálculos envolvidos no caso de $\underset{\sim}{X}$ contínuo.

EXEMPLO 17. Se $X \sim U[0, 1]$, qual a distribuição de $Y = -\log(X)$?

Neste exemplo, X é um vetor unidimensional, i.e., uma variável aleatória. Como

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

e $P(0 < X < 1) = 1$, temos $F_Y(y) = 0$, $y \le 0$. Se $y > 0$, então $P(Y \le y) = P(-\log(X) \le y) = P(X \ge e^{-y}) = 1 - e^{-y}$.

Solução: $Y \sim \exp(1)$, i.e., Y tem distribuição exponencial com parâmetro 1.

EXEMPLO 18. Se X e Y são independentes, cada uma com distribuição uniforme no intervalo $[0, 1]$, qual a distribuição de $Z = X/Y$?

Como $0 < Z < \infty$ se $X > 0$ e $Y > 0$, temos

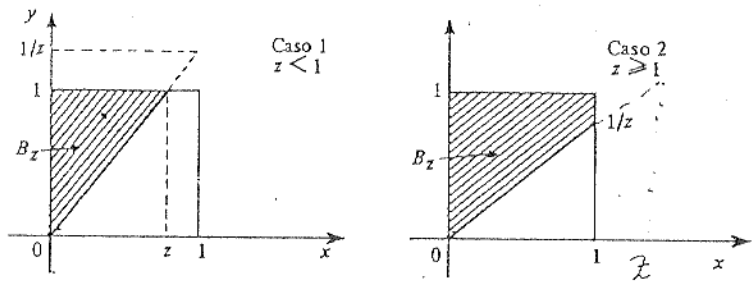
$$P(0 < Z < \infty) \ge P(0 < X \le 1, 0 < Y \le 1) = P(0 < X \le 1)\, P(0 < Y \le 1) = 1.$$

Logo, $F_Z(z) = 0$ para $z \le 0$. Notemos que segundo a nossa definição formal, pode ser que Z não seja uma variável aleatória, pois pode tomar o valor $+\infty$ (quando $Y = 0$ e $X > 0$) ou ainda ficar sem definição (quando $X = 0$ e $Y = 0$). Mas esses dois eventos excêntricos têm probabilidade zero, e podemos afirmar que, com probabilidade 1, Z está bem definida e toma valores finitos, e, em todo caso, eventos de probabilidade zero são desprezíveis para nossos propósitos. Se você quiser, poderá substituir Z pela variável aleatória Z' definida por

$$Z' = \begin{cases} X/Y, & \text{se } X > 0 \text{ e } Y > 0 \\ 0, & \text{caso contrário.} \end{cases}$$

Voltando aos cálculos, temos para $z > 0$,

$$F_Z(z) = P(X/Y \le z) = P((X, Y) \in B_z), \text{ onde:}$$



Podemos restringir nossa atenção ao quadrado unitário, já que $P(0 \le X \le 1, 0 \le Y \le 1) = 1$, i.e., $P((X, Y) \in \text{quadrado unitário}) = 1$. Como X e Y têm densidade conjunta igual a 1 no quadrado, pela Proposição 2.5(a), temos

$$P((X, Y) \in B_z) = \int_{B_z} \int 1\, dx\, dy = \text{área}\,(B_z).$$

Logo,

$$F_Z(z) = \begin{cases} 0, & z \le 0 \\ z/2, & 0 < z < 1 \\ 1 - \dfrac{1}{2z}, & z \ge 1. \end{cases}$$

Como F_Z é contínua e derivável por partes, Z possui densidade, a saber:

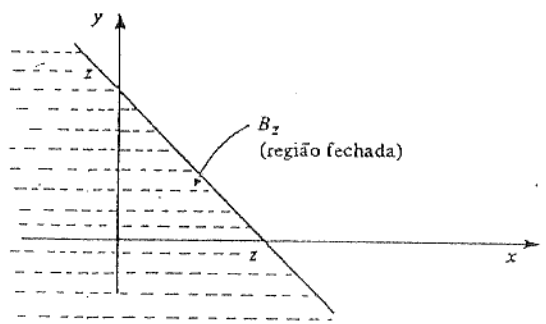
$$f_Z(z) = F_Z'(z) = \begin{cases} 0, & z < 0 \\ 1/2, & 0 < z < 1 \\ \dfrac{1}{2z^2}, & z > 1. \end{cases}$$

(Os valores de f_Z em 0 e 1 são arbitrários.)

Para certos casos "padrão", existem fórmulas que podem ser aplicadas para obter a distribuição de $g(X)$. A soma de duas variáveis aleatórias é o caso mais típico disto. Então, sejam X e Y variáveis aleatórias em $(\Omega, \mathcal{A}, P)$, com $Z = X + Y$. Calculemos a distribuição de Z. A solução geral é

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

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



Vamos supor agora que (X, Y) tenha densidade $f(x, y)$, i.e., vamos restringir nossa atenção para o caso contínuo. Neste caso,

$$F_Z(z) = \text{(pela Proposição 2.2'(b))} = \int_{B_z}\int f(x, y)\,dx\,dy =$$

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

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

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

$$= \int_{-\infty}^{z} g(s)\,ds,$$

onde $g(s) = \int_{-\infty}^{\infty} f(s-t, t)\,dt.$

Logo, g é a densidade da soma $Z = X + Y$, i.e.,

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f(z-t, t)\,dt = \text{(fazendo } s = z-t) = \int_{-\infty}^{\infty} f(s, z-s)\,ds.$$

Por isso, já está provada a seguinte proposição.

PROPOSIÇÃO 2.7. *(a) Se X e Y têm densidade conjunta $f(x, y)$, então*

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f(z-t, t)\,dt = \int_{-\infty}^{\infty} f(t, z-t)\,dt.$$

(b) Se X e Y são independentes com densidades f_X e f_Y, então (por (a) e a Proposição 2.5(a)) *$X + Y$ tem densidade*

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z-t) f_Y(t)\,dt = \int_{-\infty}^{\infty} f_X(t) f_Y(z-t)\,dt.$$

Observação. Se f_1 e f_2 são densidades de variáveis aleatórias, sua *convolução $f_1 * f_2$ é definida por*

$$f_1 * f_2(x) = \int_{-\infty}^{\infty} f_1(x-t) f_2(t)\,dt.$$

Portanto, pela proposição, se X e Y são independentes e absolutamente contínuas, então $f_X * f_Y$ é densidade da soma $X + Y$.

Voltando ao exemplo da distribuição normal bivariada, pode-se mostrar que, se (X, Y) tem distribuição normal bivariada, então $X + Y \sim$ $\sim N(\mu_1 + \mu_2, \sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2)$. (Exercício: verifique os cálculos.) Em particular, se X e Y são independentes (i.e., se $\rho = 0$), então $X + Y \sim$ $\sim N(\mu_1 + \mu_2,\ \sigma_1^2 + \sigma_2^2)$.

Podemos generalizar esse resultado para a soma de n variáveis aleatórias normais independentes. Com efeito, sejam $X_1, X_2, \ldots, X_n$ independentes, com $X_i \sim N(\mu_i, \sigma_i^2)$, $1 \leq i \leq n$. Então, por indução,

$$X_1 + \ldots + X_n \sim N(\mu_1 + \mu_2 + \ldots + \mu_n,\ \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2).$$

Fazemos a indução da seguinte maneira: $X_1 + X_2 \sim N(\mu_1 + \mu_2,$ $\sigma_1^2 + \sigma_2^2)$ e $X_1 + X_2$ é independente de X_3, logo $X_1 + X_2 + X_3 \sim$ $\sim N(\mu_1 + \mu_2 + \mu_3,\ \sigma_1^2 + \sigma_2^2 + \sigma_3^2)$, etc.. O problema que surge nesse

argumento é o seguinte: como garantir a independência entre $X_1 + X_2$ e X_3? Para resolver este problema, basta utilizar uma outra propriedade "hereditária" de famílias de variáveis aleatórias independentes:

PROPOSIÇÃO 2.8. *Se $X_1, \ldots, X_n$ são variáveis aleatórias* independentes, *então funções de famílias disjuntas das X_i também são independentes.* (As funções precisam ser mensuráveis. Exemplo: $(X_1 + X_2)^2$, e^{-X_3}, $\max(X_4, X_5, X_6)$ são independentes. No exemplo acima, $X_1 + X_2$ e X_3 são independentes, $X_1 + X_2 + X_3$ e X_4 são independentes, etc..)

Prova. Teoria da Medida. Para ilustrar o método, provemos o seguinte *caso especial*: *Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes. Se $g_1, \ldots, g_n$ são funções reais mensuráveis, então $g_1(X_1), \ldots, g_n(X_n)$ são variáveis aleatórias independentes.* (Por exemplo, $X_1^2, X_2^2, \ldots, X_n^2$ são independentes, pois $g_i(x) = x^2$ é função contínua e, portanto, mensurável. E $X_1, X_2^2, \ldots, X_n^n$ são igualmente independentes.)

Prova do caso especial. Seja $Y_i = g_i(X_i)$. Pela Proposição 2.4, basta provar que a função de distribuição conjunta das Y_i fatora:

$$F_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = P(g_1(X_1) \le y_1, \ldots, g_n(X_n) \le y_n) =$$

$$= P(X_1 \in g_1^{-1}((-\infty, y_1]), \ldots, X_n \in g_n^{-1}((-\infty, y_n])) = \text{(Definição 2.8)} =$$

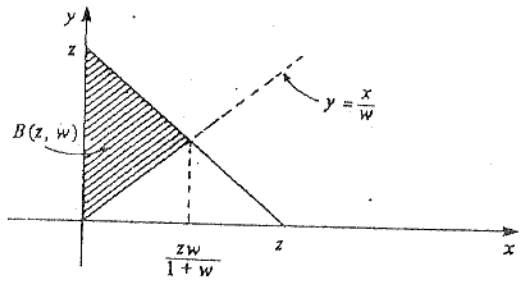
$$= \prod_{i=1}^{n} P(X_i \in g_i^{-1}((-\infty, y_i])) = \prod_{i=1}^{n} P(g_i(X_i) \le y_i) = \prod_{i=1}^{n} F_{Y_i}(y_i). \ \square$$

Passaremos a considerar o caso de funções vetoriais de vetores aleatórios, ou seja, o caso $\underset{\sim}{Y} = (g_1(\underset{\sim}{X}), \ldots, g_k(\underset{\sim}{X})) = g(\underset{\sim}{X})$, onde $g: \mathbb{R}^n \to \mathbb{R}^k$. Um exemplo simples, em que obtemos a distribuição conjunta de duas funções de um vetor aleatório bivariado, é o seguinte.

EXEMPLO 19. Sejam X e Y variáveis aleatórias independentes, cada uma com distribuição exponencial de parâmetro 1. Provar que $X + Y$ e X/Y também são independentes e achar as suas distribuições.

Solução 1. (Método anterior). Sejam $Z = X + Y$, $W = X/Y$. Aplicaremos o critério para independência, mostrando que a função de distribuição $F_{Z,W}$ é fatorável.

Para $z > 0$, $w > 0$, $F_{Z,W}(z, w) = P(Z \le z, W \le w) = P(X + Y \le z, X/Y \le w) = P((X, Y) \in B(z, w))$, onde $B(z, w)$ é:



(Podemos restringir nossa atenção ao primeiro quadrante, já que $P(X > 0, Y > 0) = 1$).

Como X e Y têm densidade conjunta

$$f(x, y) = \begin{cases} e^{-(x+y)}, & x > 0, \ y > 0 \\ 0, & \text{caso contrário} \end{cases}$$

(por quê?), temos

$$P((X, Y) \in B(z, w)) = \int_{B(z,w)} \int f(x, y)\, dx dy = \int_0^{zw/(1+w)} \int_{x/w}^{z-x} e^{-(x+y)} dy dx =$$

$$= \int_0^{zw/(1+w)} e^{-x}(-e^{-y})\Big|_{x/w}^{z-x} dx = \int_0^{zw/(1+w)} (e^{-x(1+w)/w} - e^{-z})\, dx =$$

$$= -\frac{w}{1+w} e^{-x(1+w)/w}\Big|_0^{zw/(1+w)} - \frac{zw}{1+w} e^{-z} = \left(\frac{w}{1+w}\right)(1 - e^{-z} - ze^{-z}).$$

Como $F_{Z,W}(z, w) = 0$ quando $z \le 0$ ou $w \le 0$, e além disso, os dois fatores convergem para 1 quando $z \to \infty$ e $w \to \infty$, o resultado decorre da Proposição 2.4(b): Z e W são independentes e

$$F_W(w) = \begin{cases} \dfrac{w}{1+w}, & w > 0 \\ 0, & w \le 0 \end{cases}$$

$$F_Z(z) = \begin{cases} 1 - e^{-z} - ze^{-z}, & z > 0 \\ 0, & z \le 0. \end{cases}$$

Solução 2. (Método do jacobiano. Ao leitor, sugere-se uma leitura rápida antes de passar ao próximo parágrafo, com uma relida mais profunda depois de ler o Teorema 2.1.) A transformação

$z = g_1(x, y) = x + y$, $w = g_2(x, y) = \dfrac{x}{y}$ é uma bijeção (correspondência biunívoca sobre) no primeiro quadrante (i.e., $x > 0$, $y > 0$), e $P((X, Y) \in$ $\in$ primeiro quadrante) = 1. Como

$$x = h_1(z, w) = \frac{wz}{w+1}, \quad y = h_2(z, w) = \frac{z}{w+1},$$

o jacobiano é

$$J = \begin{vmatrix} \dfrac{w}{w+1} & \dfrac{z}{(w+1)^2} \\ \dfrac{1}{w+1} & -\dfrac{z}{(w+1)^2} \end{vmatrix} = -\frac{z}{(w+1)^2}.$$

Logo

$$f_{Z,W}(z, w) = \frac{z}{(w+1)^2} f(x, y) = \frac{z}{(w+1)^2} e^{-(x+y)} = \frac{z}{(w+1)^2} e^{-z} =$$
$$= \frac{1}{(w+1)^2} \cdot ze^{-z},$$

para $z > 0$, $w > 0$ (e é igual a zero para $z \leq 0$ ou $w \leq 0$).

É fácil verificar que $\dfrac{1}{(w+1)^2}$, $w > 0$, e ze^{-z}, $z > 0$, são densidades; de fato, são as derivadas das funções de distribuição F_W e F_Z obtidas na solução 1. Decorre, então, da Proposição 2.5(b) que Z e W são independentes, e

$$f_Z(z) = \begin{cases} ze^{-z}, & z > 0 \\ 0, & z \leq 0 \end{cases}$$

$$f_W(w) = \begin{cases} \dfrac{1}{(w+1)^2}, & \omega > 0 \\ 0, & \omega \leq 0. \end{cases}$$

§2.7 O método do jacobiano

Suponha que $G_0 \subset \mathbb{R}^n$ e $G \subset \mathbb{R}^n$ sejam regiões abertas, e que $g\colon G_0 \to G$ seja uma bijeção entre G_0 e G, onde

$$g(x_1, \ldots, x_n) = (g_1(x_1, \ldots, x_n), \ldots, g_n(x_1, \ldots, x_n)) = (y_1, \ldots, y_n).$$

Então existe a função inversa $h = g^{-1}$ em G, onde

$$x_1 = h_1(y_1, \ldots, y_n), \ldots, x_n = h_n(y_1, \ldots, y_n).$$

Suponha também que existam as derivadas parciais

$$\frac{\partial x_i}{\partial y_j} = \frac{\partial h_i(y_1, \ldots, y_n)}{\partial y_j}, \quad 1 \leq i, j \leq n,$$

e que elas sejam contínuas em G. Definimos o jacobiano $J(\underset{\sim}{x}, \underset{\sim}{y})$ pelo determinante:

$$J(\underset{\sim}{x}, \underset{\sim}{y}) = \left|\left(\frac{\partial x_i}{\partial y_j}\right)\right| = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \cdots & \dfrac{\partial x_1}{\partial y_n} \\ \vdots & & \vdots \\ \dfrac{\partial x_n}{\partial y_1} & \cdots & \dfrac{\partial x_n}{\partial y_n} \end{vmatrix}.$$

Segundo um teorema do cálculo, se o jacobiano for não-nulo para todo $y \in G$, então

$$\int_A \cdots \int f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = \int_{g(A)} \cdots \int f(h_1(y_1, \ldots, y_n), \ldots,$$
$$h_n(y_1, \ldots, y_n)) \left|J(\underset{\sim}{x}, \underset{\sim}{y})\right| dy_1 \ldots dy_n,$$

para qualquer f integrável em A, onde $A \subset G_0$.

Vamos traduzir para a linguagem de variáveis aleatórias e densidades: seja f a densidade conjunta das variáveis aleatórias $X_1, \ldots, X_n$, onde $P((X_1, \ldots, X_n) \in G_0) = 1$. Sejam $Y_1, \ldots, Y_n$ as variáveis transformadas, i.e., $Y_i = g_i(X_1, \ldots, X_n)$, $i = 1, \ldots, n$. Então para $B \subset G$, B boreliano, temos

$P((Y_1, \ldots, Y_n) \in B) = P((X_1, \ldots, X_n) \in h(B)) =$ (pela Proposição 2.2'(b)) =

$$= \int_{h(B)} \cdots \int f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n =$$

$$= \int_B \cdots \int f(h_1(y_1, \ldots, y_n), \ldots, h_n(y_1, \ldots, y_n)) \cdot \left|J(\underset{\sim}{x}, \underset{\sim}{y})\right| dy_1 \ldots dy_n.$$

Como

$$P(\underset{\sim}{Y} \in G) = P(\underset{\sim}{X} \in h(G)) = P(\underset{\sim}{X} \in G_0) = 1,$$

temos, para todo boreliano B no $\mathbb{R}^n$,

$$P(\underset{\sim}{Y} \in B) = P(\underset{\sim}{Y} \in B \cap G) = \int_{B \cap G} \cdots \int f(h(\underset{\sim}{y}))\,|J(\underset{\sim}{x}, \underset{\sim}{y})|\,dy_1 \ldots dy_n.$$

Esta última integral é a integral sobre o conjunto B da função que toma o valor $f(h(\underset{\sim}{y}))\,|J(\underset{\sim}{x}, \underset{\sim}{y})|$ para $\underset{\sim}{y} \in G$ e o valor zero fora de G. Decorre, então, da definição de densidade que esta função é a densidade de $\underset{\sim}{Y}$, e o seguinte teorema já está provado:

TEOREMA 2.1. *Sob as condições dadas acima, a densidade conjunta de* $Y_1, \ldots, Y_n$ *é:*

$$f_{\underset{\sim}{Y}}(y_1, \ldots, y_n) = \begin{cases} f(h_1(y_1, \ldots, y_n), \ldots, h_n(y_1, \ldots, y_n))\,|J(\underset{\sim}{x}, \underset{\sim}{y})|, & \underset{\sim}{y} \in G \\ 0, & \underset{\sim}{y} \notin G. \end{cases}$$

Observações. (1) O teorema diz que, sob as condições dadas, para obter a densidade de $\underset{\sim}{Y}$ basta (i) substituir o valor de $\underset{\sim}{x}$ em $f(\underset{\sim}{x})$ por seu valor em função de $\underset{\sim}{y}$, i.e., substituir $\underset{\sim}{x}$ por $h(\underset{\sim}{y}) = g^{-1}(\underset{\sim}{y})$, e (ii) multiplicar pelo módulo do jacobiano de $\underset{\sim}{x}$ em relação a $\underset{\sim}{y}$, que já é função de $\underset{\sim}{y}$.

Como é freqüentemente mais fácil obter o jacobiano de $\underset{\sim}{y}$ em relação a $\underset{\sim}{x}$, pois $\underset{\sim}{Y}$ é dado em função de $\underset{\sim}{X}$, é bom lembrar que os dois jacobianos são recíprocos e pode-se obter $J(\underset{\sim}{x}, \underset{\sim}{y})$ a partir de $J(\underset{\sim}{y}, \underset{\sim}{x})$, invertendo este último e substituindo $\underset{\sim}{x}$ por $h(\underset{\sim}{y}) = g^{-1}(\underset{\sim}{y})$. (Esta regra é análoga à regra para a derivada da função inversa no caso unidimensional:

$$\frac{dg^{-1}(y)}{dy} = \frac{1}{g'(x)}\bigg|_{x = g^{-1}(y)} = \frac{1}{g'(g^{-1}(y))}\bigg).$$

Por exemplo, na solução 2 do exemplo 19 obtivemos o jacobiano

$$J((x, y), (z, w)) = -\frac{z}{(w+1)^2},$$

derivando as funções $x = \dfrac{wz}{w+1}$ e $y = \dfrac{z}{w+1}$. Mas poderíamos ter derivado as funções originais, $z = x + y$ e $w = \dfrac{x}{y}$, para obter

$$J((z, w), (x, y)) = \begin{vmatrix} 1 & 1 \\ \dfrac{1}{y} & -\dfrac{x}{y^2} \end{vmatrix} = -\frac{(x+y)}{y^2},$$

conduzindo assim a:

$$J((x, y), (z, w)) = -\frac{y^2}{(x+y)}\bigg|_{(x,y) = (wz/(w+1),\, z/(w+1))} = -\frac{z}{(w+1)^2}.$$

(2) No caso unidimensional, em que X e Y são variáveis aleatórias, o jacobiano é a derivada.

Portanto, o teorema indica uma solução alternativa para o problema do exemplo 17: se $X \sim U[0, 1]$, qual a distribuição de $Y = -\log X$? Fazendo $G_0 = (0, 1)$, $g(x) = -\log x$ e $G = (0, \infty)$, vemos que as condições do teorema estão satisfeitas; portanto, temos

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))\,|(g^{-1})'(y)|, & y \in G \\ 0, & y \notin G. \end{cases}$$

Como $g^{-1}(y) = e^{-y}$ e $(g^{-1})'(y) = -e^{-y}$, a densidade de Y é

$$f_Y(y) = \begin{cases} e^{-y}, & y > 0 \\ 0, & y \leq 0, \end{cases}$$

i.e., $Y \sim \exp(1)$.

(3) Para obter a distribuição de $\underset{\sim}{Y} = (Y_1, \ldots, Y_k) = g(\underset{\sim}{X})$, quando a dimensão de $\underset{\sim}{Y}$ é menor que a dimensão de $\underset{\sim}{X}$ (i.e., $k < n$), muitas vezes é possível completar a transformação g através da definição conveniente de outras variáveis $Y_{k+1} = g_{k+1}(\underset{\sim}{X}), \ldots, Y_n = g_n(\underset{\sim}{X})$, determinar a densidade conjunta de $Y_1, \ldots, Y_n$ utilizando o método do jacobiano e, finalmente, obter a densidade conjunta marginal de $Y_1, \ldots, Y_k$.

Por exemplo, no exemplo 18 calculamos a densidade de $Z = X/Y$ partindo da suposição de X e Y serem independentes, cada qual tendo distribuição uniforme em $[0, 1]$. O método usado foi o método básico de obter diretamente $P(Z \leq z)$.

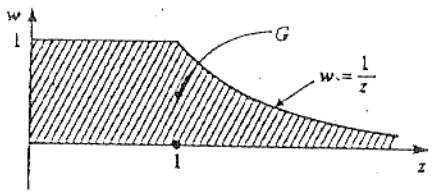
Consideremos o seguinte método alternativo: seja G_0 o quadrado aberto $(0, 1) \times (0, 1)$, de modo que

$$P((X, Y) \in G_0) = 1.$$

Definamos $g_1(x, y) = x/y$ e completemos a transformação, definindo $W = Y$, ou seja

$$g_2(x, y) = y.$$

Então a transformação g definida por $g(x, y) = (x/y, y)$ é uma bijeção entre G_0 e $G = \{(z, w) : 0 < z < 1/w,\ 0 < w < 1\}$:



Como $(x, y) = (zw, w)$, o jacobiano é

$$J((x, y), (z, w)) = \begin{vmatrix} w & z \\ 0 & 1 \end{vmatrix} = w.$$

Portanto, a densidade conjunta de Z e W é

$$f_{Z,W}(z, w) = \begin{cases} w, & (z, w) \in G \\ 0, & (z, w) \notin G. \end{cases}$$

Logo, a densidade (marginal) de Z é

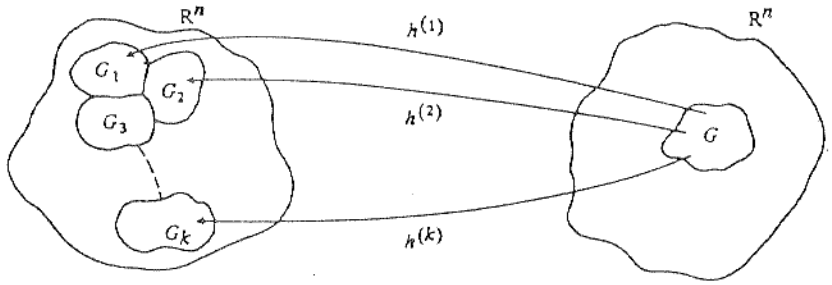
$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z,W}(z, w)\, dw =$$

$$= \begin{cases} 0, & z \le 0 \\ \displaystyle\int_0^1 w\, dw = \frac{1}{2}, & 0 < z \le 1 \\ \displaystyle\int_0^{1/z} w\, dw = \frac{1}{2z^2}, & z > 1. \end{cases}$$

Esta é a densidade obtida anteriormente.

Ocorre que para determinarmos a distribuição de $\underset{\sim}{Y} = g(\underset{\sim}{X})$, onde $\underset{\sim}{X} = (X_1, \ldots, X_n)$ e $\underset{\sim}{Y} = (Y_1, \ldots, Y_n)$, podemos utilizar o método do jacobiano em muitos casos em que a função g não é 1 a 1, bastando que g seja 1 a 1 quando restrita a cada uma de k regiões abertas cuja união contêm o valor de $\underset{\sim}{X}$ com probabilidade um. Para tanto, suponhamos que $G, G_1, \ldots, G_k$ sejam subregiões abertas do $\mathbb{R}^n$ tais que valha $P(\underset{\sim}{X} \in \bigcup_{i=1}^{k} G_i) = 1$, e tais que a função $g|_{G_\ell}$, a restrição de g a G_ℓ, seja

uma correspondência biunívoca entre G_ℓ e G, $\forall\, \ell = 1, \ldots, k$. (Neste caso podemos dizer que a função g é "k a 1".) Além disso, suponhamos que a função inversa de $g|_{G_\ell}$, denotada por $h^{(\ell)}$, satisfaça todas as condições da função h do caso anterior, e indiquemos com $J_\ell(\underset{\sim}{x}, \underset{\sim}{y})$ o jacobiano da função $h^{(\ell)}$. (Este jacobiano é função de $\underset{\sim}{y} \in G$. Notemos que $h^{(\ell)}\colon G \to G_\ell$ é uma bijeção.) Temos, então, o seguinte esquema:



G = imagem de $g|_{G_\ell}, \forall \ell$.

Desde que $P(\underset{\sim}{X} \in \bigcup_{\ell=1}^{k} G_\ell) = 1$ e $[\underset{\sim}{X} \in \bigcup_{\ell=1}^{k} G_\ell] \subset [\underset{\sim}{Y} \in G]$, temos $P(\underset{\sim}{Y} \in G) = 1$, i.e., $\underset{\sim}{Y}$ toma valores só em G (pelo menos com probabilidade 1).

TEOREMA 2.1'. *Sob as condições dadas acima, se $\underset{\sim}{X}$ tem densidade $f(x_1, \ldots, x_n)$, então $\underset{\sim}{Y}$ tem densidade*

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \begin{cases} \displaystyle\sum_{\ell=1}^{k} f(h^{(\ell)}(\underset{\sim}{y}))\cdot |J_\ell(\underset{\sim}{x}, \underset{\sim}{y})|, & \underset{\sim}{y} \in G \\ 0, & \underset{\sim}{y} \notin G. \end{cases}$$

Prova. Se $B \subset G$,

$$P(\underset{\sim}{Y} \in B) = P(g(\underset{\sim}{X}) \in B) = \sum_{\ell=1}^{k} P(g(\underset{\sim}{X}) \in B,\ \underset{\sim}{X} \in G_\ell) =$$

$$= \sum_{\ell=1}^{k} P(\underset{\sim}{X} \in h^{(\ell)}(B)) = \sum_{\ell=1}^{k} \int_{h^{(\ell)}(B)} \cdots \int f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n =$$

$$= (\text{mudança de variável}) =$$

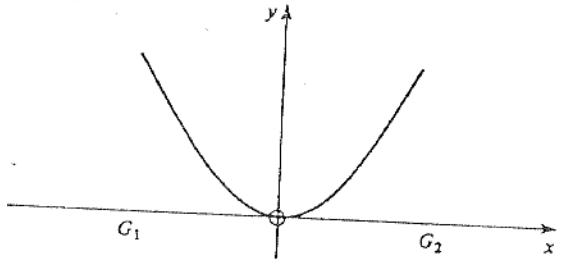
$$= \sum_{\ell=1}^{k} \int_B \cdots \int f(h^{(\ell)}(\underset{\sim}{y}))\, |J_\ell(\underset{\sim}{x}, \underset{\sim}{y})|\, dy_1 \ldots dy_n =$$

$$= \int_B \cdots \int \sum_{\ell=1}^{k} f(h^{(\ell)}(\underset{\sim}{y}))\, |J_\ell(\underset{\sim}{x}, \underset{\sim}{y})|\, dy_1 \ldots dy_n.$$

Como $P(Y \in G) = 1$, o integrando é a densidade de Y (em G). □

EXEMPLO 20. Seja X uma variável aleatória com distribuição $N(0, 1)$. Qual a densidade de $Y = X^2$?

Solução. $Y = g(X) = X^2$. A função g induz duas correspondências biunívocas quando restrita a $(-\infty, 0)$ e $(0, \infty)$:



Aqui, $G = (0, \infty)$, $h^{(1)}(y) = -\sqrt{y}$, $h^{(2)}(y) = \sqrt{y}$. Notemos que $P(X \in G_1 \cup G_2) = 1$.

Os jacobianos (neste caso as derivadas) das $h^{(\ell)}$ em relação a y são

$$J_1(x, y) = \frac{dh^{(1)}(y)}{dy} = -\frac{1}{2\sqrt{y}}, \quad y \in G \text{ (i.e., } y > 0)$$

$$J_2(x, y) = \frac{dh^{(2)}(y)}{dy} = \frac{1}{2\sqrt{y}}, \quad y \in G.$$

Como a densidade de X é $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, a densidade de Y é

$$f_Y(y) = f(h^{(1)}(y)) \cdot \frac{1}{2\sqrt{y}} + f(h^{(2)}(y)) \frac{1}{2\sqrt{y}} =$$

$$= \frac{1}{\sqrt{2\pi}} \left(e^{-y/2} \cdot \frac{1}{2\sqrt{y}} + e^{-y/2} \cdot \frac{1}{2\sqrt{y}} \right) = \frac{1}{\sqrt{2\pi}\, y} e^{-y/2}, \quad y \in G,$$

com $f_Y(y) = 0$ se $y \leq 0$.

A distribuição de Y é, por definição, qui-quadrado com 1 grau de liberdade, $Y \sim \chi^2(1)$. Como sua densidade é proporcional à densidade da distribuição $\Gamma(1/2, 1/2)$ (veja o exemplo 11) e as duas densidades têm que ter a mesma integral (=1), concluímos que as duas densidades são iguais ($\Gamma(1/2, 1/2) = \chi^2(1)$) e, em particular,

$$\frac{1}{\sqrt{2\pi}} = \frac{(1/2)^{1/2}}{\Gamma(1/2)},$$

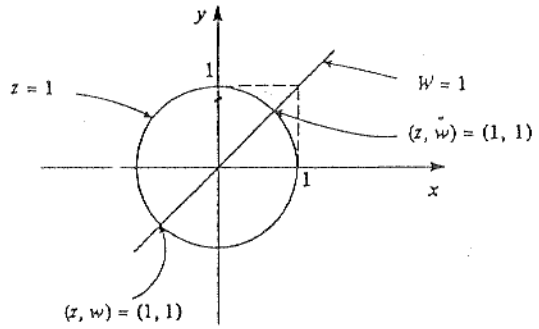
i.e.,

$$\Gamma(1/2) = \sqrt{\pi}.$$

(*Exercício.* Calcule a densidade de X^2 neste exemplo pelo método básico, primeiro obtendo a função de distribuição e depois derivando-a para obter a densidade.)

EXEMPLO 21. Sejam X e Y independentes com distribuição comum $N(0, 1)$. Provar que $Z = X^2 + Y^2$ e $W = X/Y$ são independentes e achar as suas distribuições. (Nota: por definição $Z \sim \chi^2(2)$, qui-quadrado com dois graus de liberdade. Você já sabe, talvez, que W tem distribuição de Cauchy.)

Solução. A função $g: \mathbb{R}^2 \to \mathbb{R}^2$, definida por $g(x, y) = (z, w) = (x^2 + y^2, x/y)$, é 2 a 1:



Sejam $G = \{(x, y): x > 0\}$, $G_1 = \{(x, y): y > 0\}$, $G_2 = \{(x, y): y < 0\}$.

Então $g|_{G_1}$ e $g|_{G_2}$ são correspondências biunívocas entre as regiões abertas G_ℓ e G, $\ell = 1, 2$, e $P((X, Y) \in G_1 \cup G_2) = 1$ (a probabilidade de (X, Y) tomar um valor na reta $\{(x, y): y = 0\}$ é $P(Y = 0) = 0$).

Precisamos, então, obter os jacobianos das funções inversas $h^{(1)}$ e $h^{(2)}$ em G. Mas para tanto, basta obtermos os jacobianos das funções $g|_{G_1}$ e $g|_{G_2}$, que são recíprocos dos jacobianos das inversas, e substituirmos o valor (x, y) pelo valor $h^{(1)}(z, w)$ ou $h^{(2)}(z, w)$. Veremos que neste exemplo não precisamos determinar explicitamente as inversas (o exemplo é simples). De fato, temos

$$J_1((x, y), (z, w)) = \frac{1}{\begin{vmatrix} 2x & 2y \\ \frac{1}{y} & -\frac{x}{y^2} \end{vmatrix}} = -\frac{1}{2\left(\frac{x^2}{y^2} + 1\right)} = -\frac{1}{2(w^2 + 1)}$$

e

$$J_2((x, y), (z, w)) = -\frac{1}{2(w^2+1)}.$$

Portanto, a densidade de (Z, W) é

$$f_{Z,W}(z, w) = \{ f(h^{(1)}(z, w)) + f(h^{(2)}(z, w))\} \cdot \frac{1}{2(w^2+1)} \text{ em } G.$$

Como

$$f(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2} = \frac{1}{2\pi} e^{-z/2},$$

temos

$$f_{Z,W}(z, w) = 2\left(\frac{1}{2\pi} e^{-z/2}\right) \cdot \frac{1}{2(w^2+1)} = \frac{1}{2} e^{-z/2} \cdot \frac{1}{\pi(w^2+1)},$$

para $(z, w) \in G$, i.e., $z > 0$ e $w \in \mathbb{R}$ ($e = 0$, $(z, w) \notin G$).

Como a densidade conjunta é o produto de duas densidades, concluímos (Proposição 2.5(b)) que Z e W são independentes, $Z \sim \exp(1/2)$, e $W \sim$ Cauchy-padrão.

(*Observação.* Decorre disso que $\chi^2(2) = \exp(1/2) = \Gamma(1, 1/2)$.)

EXEMPLO 22. Obteremos a densidade conjunta das estatísticas de ordem de uma amostra aleatória de uma distribuição absolutamente contínua. Primeiro, as definições necessárias:

DEFINIÇÃO 2.9. Variáveis aleatórias que possuem a mesma distribuição são chamadas *identicamente distribuídas*. Se $X_1, \ldots, X_n$ são variáveis aleatórias independentes e identicamente distribuídas, com função de distribuição comum $F = F_{X_i}$, dizemos que as X_i formam uma *amostra aleatória de tamanho n* (tirada de F, ou tirada de uma população com distribuição F). As X_i ordenadas em ordem crescente são as *estatísticas de ordem* da amostra e são representadas por $X_{(1)}, \ldots, X_{(n)}$, onde para $\omega \in \Omega$, $(X_{(1)}(\omega), \ldots, X_{(n)}(\omega))$ é qualquer permutação de $(X_1(\omega), \ldots, X_n(\omega))$ que satisfaz

$$X_{(1)}(\omega) \leq X_{(2)}(\omega) \leq \ldots \leq X_{(n)}(\omega).$$

Observação. $X_{(1)} = \min(X_1, \ldots, X_n)$ é o *mínimo* da amostra, $X_{(n)} = \max(X_1, \ldots, X_n)$ é o seu *máximo*.

Suponhamos, então que $X_1, \ldots, X_n$ formem uma amostra aleatória de uma distribuição com densidade f; deste modo f é a densidade comum às X_i e, pela independência, $X_1, \ldots, X_n$ têm densidade conjunta

$$f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i), \ (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Neste caso, as estatísticas de ordem $X_{(1)}, \ldots, X_{(n)}$ possuem densidade conjunta

$$f_{X_{(1)}, \ldots, X_{(n)}}(x_1, \ldots, x_n) = \begin{cases} n! \prod_{i=1}^{n} f(x_i), \text{ se } x_1 < x_2 < \ldots < x_n \\ 0, \text{ caso contrário.} \end{cases}$$

Prova. Provaremos primeiro para $n = 2$. Definamos

$$g(x_1, x_2) = \begin{cases} (x_1, x_2) \text{ se } x_1 < x_2 \\ (x_2, x_1) \text{ se } x_2 < x_1 \\ (x_1, x_1) \text{ se } x_1 = x_2. \end{cases}$$

Então $Y = (X_{(1)}, X_{(2)}) = g(X_1, X_2)$ e g é 2 a 1. De fato, definindo $G = G_1 = \{(x_1, x_2): x_1 < x_2\}$ e $G_2 = \{(x_1, x_2): x_2 < x_1\}$, vemos que $g|_{G_1}$ e $g|_{G_2}$ são correspondências biunívocas entre as regiões abertas G_ℓ e G, $\ell = 1, 2$. Além disso, temos $P((X_1, X_2) \in G_1 \cup G_2) = 1$, pois

$$P((X_1, X_2) \notin G_1 \cup G_2) = P(X_1 = X_2) = \iint_{\{(x_1, x_2): x_1 = x_2\}} f(x_1) f(x_2)\, dx_1\, dx_2 = 0$$

(a área da diagonal é nula, portanto a integral sobre a diagonal também é nula).

Como $h^{(1)}(y_1, y_2) = (y_1, y_2)$ e $h^{(2)}(y_1, y_2) = (y_2, y_1)$, os jacobianos de $h^{(1)}$ e $h^{(2)}$ são, respectivamente, iguais a 1 e -1. Em cada caso o módulo do jacobiano é 1 em G, logo a densidade conjunta de $X_{(1)}$ e $X_{(2)}$ é

$$f_{X_{(1)}, X_{(2)}}(y_1, y_2) = f_{X_1, X_2}(y_1, y_2) + f_{X_1, X_2}(y_2, y_1) = f(y_1) f(y_2) + f(y_2) f(y_1) = 2 f(y_1) f(y_2),$$

para $(y_1, y_2) \in G$, ou seja, para $y_1 < y_2$ (e é igual a zero se $y_2 \leq y_1$), como queríamos demonstrar.

Para $n > 2$, a prova é análoga. Neste caso, a função g é $n!$ a 1 e há $n!$ regiões G_ℓ, correspondentes às $n!$ permutações de $G = \{(x_1, \ldots x_n): x_1 < x_2 < \ldots < x_n\}$. Como o jacobiano de cada permutação é 1 ou -1,

e como o produto dos n termos $f(y_i)$ não depende da ordem dos termos, segue-se o resultado. □

Consideremos o seguinte exemplo específico: se $X_1, \ldots, X_n$ são independentes e identicamente distribuídas, com $X_i \sim U[0,1]$, então $f(x) = I_{[0,1]}(x)$ e a densidade conjunta das estatísticas de ordem é

$$f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = \begin{cases} n! \text{ se } 0 \le x_1 < \ldots < x_n \le 1 \\ 0 \text{ caso contrário,} \end{cases}$$

de modo que $X_{(1)}, \ldots, X_{(n)}$ têm distribuição uniforme na pirâmide $\{(x_1, \ldots, x_n)\colon 0 \le x_1 < \ldots < x_n \le 1\}$.

§2.8. Observações adicionais — variáveis e vetores aleatórios

(a) Se $X_1, \ldots, X_n$ têm densidade conjunta $f(x_1, \ldots, x_n)$, então, como no caso unidimensional, f é a derivada de $F = F_{X_1,\ldots,X_n}$, no seguinte sentido:

$$f(x_1, \ldots, x_n) = \frac{\partial^n F(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n}$$

em quase toda parte, i.e., em todo ponto exceto num conjunto de medida de Lebesgue nula (volume zero).

(b) Seja $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função não-negativa ($f(x_1, \ldots, x_n) \ge 0$). Como no caso unidimensional (veja o §2.2), f é densidade de algum vetor aleatório se, e só se,

$$\int \ldots \int f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = 1.$$

O argumento é o mesmo: se a integral é igual a 1, então F definida por

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_n} \ldots \int_{-\infty}^{x_1} f(t_1, \ldots, t_n)\, dt_1 \ldots dt_n$$

satisfaz as quatro condições definidoras de função de distribuição n-dimensional (verifique!). Reciprocamente, se f é densidade, então a Definição 2.7(b) e a propriedade F3 implicam que a integral de f em $\mathbb{R}^n$ é igual a 1.

(c) Eis uma relação de algumas outras distribuições úteis na Estatística:

(i) Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com distribuição comum $N(0, 1)$.

Dizemos que a soma $X_1^2 + X_2^2 + \ldots + X_n^2$ tem *distribuição qui-quadrado* com n graus de liberdade. Notação:

$$X_1^2 + \ldots X_n^2 \sim \chi^2(n).$$

Para verificar que a distribuição $\chi^2(n)$ é a $\Gamma(n/2, 1/2)$, siga este caminho: verifique primeiro que $X_1^2 \sim \Gamma(1/2, 1/2)$ (veja o exemplo 20); prove a seguir que se X e Y são independentes e $X \sim \Gamma(\alpha_1, \beta)$, $Y \sim \Gamma(\alpha_2, \beta)$, então $X + Y \sim \Gamma(\alpha_1 + \alpha_2, \beta)$; e finalmente, mostre por indução e pela propriedade hereditária da independência que $X_1^2 + \ldots + X_n^2 \sim \Gamma(n/2, 1/2)$. (Para verificar que $X + Y \sim \Gamma(\alpha_1 + \alpha_2, \beta)$, use a convolução – veja a Proposição 2.6(b).)

Nota: quando $n = 2$, a distribuição é exponencial.

(ii) Se $X \sim N(0, 1)$, $Y \sim \chi^2(n)$, e X, Y são independentes, então

$$T = \frac{X}{\sqrt{Y/n}}$$

tem *distribuição t de Student* com n graus de liberdade.

Por exemplo, sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas, com $X_i \sim N(0, \sigma^2)$, onde $\sigma^2 > 0$. Definamos

$$\overline{X} = \frac{1}{n}(X_1 + \ldots + X_n) = \text{"média amostral"},$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \overline{X})^2 = \text{"variância amostral"}.$$

É fácil verificar que $\dfrac{\sqrt{n}\,\overline{X}}{\sigma}$ possui distribuição $N(0, 1)$. Acontece que $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ e $\overline{X}$ e S^2 são independentes (você verá isso em algum curso de Estatística), logo

$$T = \frac{\sqrt{n}\,\overline{X}}{S} \sim t(n-1).$$

(iii) Se $X \sim \chi^2(k)$, $Y \sim \chi^2(n)$, e X, Y são independentes, então

$$F = \frac{X/k}{Y/n}$$

tem *distribuição F* com k e n graus de liberdade, i.e., $\sim F(k, n)$. Pelo item (ii), se T possui distribuição $t(n)$, então $T^2 \sim F(1, n)$.

EXERCÍCIOS DO CAPÍTULO 2

§2.1

1. Seja X o número de caras obtidas em 4 lançamentos de uma moeda honesta. Desenhe o gráfico da função de distribuição de X.
2. Um ponto é selecionado, ao acaso, do quadrado unitário $[0,1] \times [0,1]$. Seja X a primeira coordenada do ponto selecionado. Faça o gráfico da função de distribuição de X.
3. Se adotássemos $F(x) = P(X < x)$ como definição da função de distribuição de X, qual seria a distinção entre o gráfico de F_{X_6} é o desenhado no §2.1? Haveria alguma mudança na função de distribuição de T_1 no mesmo exemplo?
4. Seja X uma variável aleatória com distribuição de Poisson, parâmetro $\lambda > 0$. Mostre que a função de distribuição de X é

$$F(x) = \begin{cases} \dfrac{1}{n!}\displaystyle\int_{\lambda}^{\infty} e^{-t}t^n\,dt & \text{se } n \le x < n+1,\ n = 0, 1, 2, \ldots \\ 0 & \text{se } x < 0. \end{cases}$$

5. Suponha que a vida útil de certo tipo de lâmpada tenha distribuição exponencial com parâmetro λ.
   (a) (*Falta de memória* da distribuição exponencial. Compare com o exercício 27(c) do Capítulo 1.) Seja T a vida de uma lâmpada desse tipo. Mostre que

$$P(T > t + s \mid T > t) = P(T > s)\ \forall s, t > 0.$$

   (b) Suponha que $\lambda = 3$ quando a vida é expressa em dias. Uma lâmpada solitária é ligada em uma sala no instante $t = 0$. Um dia depois, você entra na sala e fica ali durante 8 horas, saindo no final desse período.
      (i) Qual a probabilidade de que você entre na sala quando já está escura?
      (ii) Qual a probabilidade de você entrar na sala com a lâmpada ainda acesa e sair da sala depois da lâmpada queimar?

§2.2

6. Seja X uma variável aleatória com densidade

$$f(x) = \begin{cases} cx^2, \text{ se } -1 \le x \le 1 \\ 0, \text{ caso contrário.} \end{cases}$$

   (a) Determine o valor da constante c.
   (b) Ache o valor α tal que $F_X(\alpha) = 1/4$. (α é o primeiro *quartil* da distribuição de X.)
7. Uma variável aleatória X tem função de distribuição

$$F(x) = \begin{cases} 1 & \text{se } x > 1 \\ x^3 & \text{se } 0 \le x \le 1 \\ 0 & \text{se } x < 0. \end{cases}$$

   Qual é a densidade de X?
8. Verifique que a função de Cantor é uma função de distribuição.
9. Seja X uma variável aleatória com densidade

$$f(x) = \begin{cases} \dfrac{1}{(1+x)^2}, & \text{se } x > 0 \\ 0 & \text{, caso contrário.} \end{cases}$$

   Seja $Y = \max(X, c)$, onde c é uma constante > 0.
   (a) Ache a função de distribuição de Y.
   (b) Decomponha F_Y em suas partes discreta, absolutamente contínua e singular.
10. Se X é uma variável aleatória com distribuição exponencial de parâmetro $\lambda > 0$, qual a distribuição da variável aleatória $Y = \min(\lambda, X)$? Faça a decomposição de F_Y.
11. Suponha que certa máquina seja colocada a funcionar no instante $t = 0$. Para $t > 0$, seja $Q(t + \Delta t \mid t)$ a probabilidade condicional da máquina pifar até o instante $t + \Delta t$, dado que funcionou até o instante t. A *taxa de falha* da máquina é a função

$$h(t) = \lim_{\Delta t \to 0} \frac{Q(t + \Delta t \mid t)}{\Delta t},$$

se este limite existe. Suponha que $h(t) = \lambda\alpha t^{\alpha - 1}$, onde $\lambda > 0$, $\alpha > 0$.
   (a) Ache a equação diferencial satisfeita por $P(t) = P(T > t)$, $t \ge 0$, onde T é a vida útil da máquina. (Suponha que $P(t)$ seja contínua, com $P(0) = 1$, e que as derivadas à direita e à esquerda são iguais.)
   (b) Resolva a equação diferencial do item (a). Qual a densidade de T? (Observação: A distribuição de T é a de *Weibull* com parâmetros α e λ. Quando $\alpha = 1$, a distribuição é exponencial; quando $\alpha = 2$, é de *Rayleigh*. Estes dois casos foram considerados no exercício 27 do Capítulo 1.)

U. F. M. S. M.
911066
FACULDADE DE CIÊNCIAS MATEMÁTICAS
BIBLIOTECA

§2.3

12. Determine a densidade de $Y = (b-a)X + a$, onde $X \sim U[0,1]$. (É a densidade da distribuição uniforme em $[a,b]$, e escrevemos $Y \sim U[a,b]$.) Faça o gráfico da função de distribuição de Y.
13. Se X tem densidade $f(x) = e^{-|x|}/2$, $-\infty < x < +\infty$, qual a distribuição de $Y = |X|$?
14. Cinco pontos são escolhidos, independentemente e ao acaso, do intervalo $[0,1]$. Seja X o número de pontos que pertencem ao intervalo $[0,c]$ onde $0 < c < 1$. Qual a distribuição de X?
15. Determine a distribuição do tempo de espera até o segundo sucesso em uma seqüência de ensaios de Bernoulli com probabilidade p de sucesso.
16. Uma massa radioativa emite partículas segundo um processo de Poisson a uma taxa média de 10 partículas por segundo, Um contador é colocado ao lado da massa. Suponha que cada partícula emitida atinge o contador com probabilidade 1/10, que o contador registra todas as partículas que o atingem, e que não há interação entre as partículas (elas se movimentam independentemente).
    (a) Qual a distribuição de $X_t \stackrel{def}{=}$ número de partículas *emitidas* até o tempo t, $t > 0$?
    (b) Prove que Y_t tem distribuição de Poisson, onde Y_t é o número de partículas *registradas* (contadas) até o tempo t, $t > 0$. Qual o parâmetro?

§2.4

17. (a) Demonstre que a função

$$F(x,y) = \begin{cases} 1 - e^{-x-y}, & \text{se } x \geq 0 \text{ e } y \geq 0 \\ 0, & \text{caso contrário} \end{cases}$$

não é função de distribuição de um vetor aleatório.
    (b) Mostre que a seguinte função *é* função de distribuição de algum (X, Y):

$$F(x,y) = \begin{cases} (1 - e^{-x})(1 - e^{-y}), & x \geq 0 \text{ e } y \geq 0 \\ 0, & \text{caso contrário.} \end{cases}$$

18. Uma urna contém três bolas numeradas 1, 2 e 3. Duas bolas são tiradas sucessivamente da urna, ao acaso e *sem* reposição. Seja X o número da primeira bola tirada e Y o número da segunda.
    (a) Descreva a distribuição conjunta de X e Y.
    (b) Calcule $P(X < Y)$.
19. Dizemos que a distribuição conjunta de $X_1, \ldots, X_n$ é invariante para permutações se toda permutação das X_i tem a mesma distribuição, i.e., se $(X_{\pi_1}, X_{\pi_2}, \ldots, X_{\pi_n}) \sim (X_1, \ldots, X_n)$ para toda permutação $(\pi_1, \ldots, \pi_n)$ do vetor $(1, \ldots, n)$.
    (a) Mostre que se $(X, Y) \sim (Y, X)$ e X e Y possuem densidade conjunta $f(x,y)$, então $P(X < Y) = P(X > Y) = 1/2$, com $P(X = Y) = 0$.
    (b) Generalize o item (a), provando que se a distribuição conjunta de $X_1, \ldots, X_n$ é invariante para permutações e $X_1, \ldots, X_n$ possuem densidade conjunta $f(x_1, \ldots, x_n)$, então

$$P(X_1 < X_2 < \ldots < X_n) = P(X_{\pi_1} < X_{\pi_2} < \ldots < X_{\pi_n}) = \frac{1}{n!}$$

    e $P(X_i = X_j \text{ para algum par } (i,j) \text{ tal que } i \neq j) = 0$.
20. Seleciona-se, ao acaso, um ponto do círculo unitário $\{(x,y): x^2 + y^2 \leq 1\}$. Sejam X e Y as coordenadas do ponto selecionado.
    (a) Qual a densidade conjunta de X e Y?
    (b) Determine $P(X < Y)$, $P(X > Y)$ e $P(X = Y)$.
21. Seleciona-se, ao acaso, um ponto do quadrado unitário $\{(x,y): 0 \leq x \leq 1, 0 \leq y \leq 1\}$. Sejam X e Y as coordenadas do ponto selecionado.
    (a) Qual a densidade conjunta de X e Y?
    (b) Calcule $P(|Y/X - 1| \leq 1/2)$.
    (c) Calcule $P(Y \geq X \mid Y \geq 1/2)$.

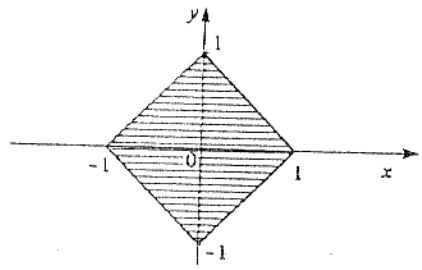
§2.5

22. (Critério para independência no caso discreto.) (a) Sejam X e Y variáveis aleatórias discretas, tomando respectivamente os valores $x_1, x_2, \ldots$ e $y_1, y_2, \ldots$. Prove que X e Y são independentes se, e somente se, $P(X = x_i, Y = y_j) = P(X = x_i)\, P(Y = y_j)\ \forall i, j$.
    (b) Mostre que se X e Y tomam somente um número finito de valores, digamos $x_1, \ldots, x_m$ e $y_1, \ldots, y_n$, então X e Y são independentes se $P(X = x_i, Y = y_j) = P(X = x_i)\, P(Y = y_j)$ para $1 \leq i \leq m-1$, $1 \leq j \leq n-1$. (Em outras palavras, para provar independência, basta verificar $(m-1)(n-1)$ equações.)

(c) Generalize o item (a) para o caso de n variáveis aleatórias. Compare com a Proposição 2.5 e explique porque é suficiente verificar se a função de probabilidade conjunta é igual ao produto de n funções de probabilidade unidimensionais.

23. Demonstre ou exiba um contra-exemplo: se X, Y e Z são independentes 2 a 2, então elas são independentes.
24. Ache a densidade conjunta e as distribuições marginais das variáveis aleatórias X e Y cuja função de distribuição conjunta está no exercício 17(b). X e Y são independentes?
25. Determine as distribuições marginais das variáveis aleatórias discretas X e Y definidas no exercício 18. X e Y são independentes?
26. Demonstre a Proposição 2.6(b).
27. Sejam X, Y e Z independentes, cada uma tendo distribuição uniforme em $[0, 1]$. Qual a probabilidade da equação quadrática $Xt^2 + Yt + Z = 0$ ter raizes reais?
28. Sejam X e Y variáveis aleatórias independentes, com $X \sim U[0, a]$ e $Y \sim U[a, a+b]$, onde $a > 0$, $b > 0$. Qual a probabilidade de que os três segmentos $[0, X]$, $[X, Y]$, $[Y, a+b]$ possam formar um triângulo?
29. Demonstre: se a variável aleatória X é independente de si mesma, então X é constante com probabilidade 1 (i.e., existe uma constante c tal que $P(X = c) = 1$).
30. Suponha que as vidas úteis T_1 e T_2 de máquinas I e II sejam variáveis aleatórias independentes tendo distribuições exponenciais com, respectivamente, parâmetros λ_1 e λ_2. Um inspetor escolhe uma das máquinas ao acaso, cada uma tendo a mesma probabilidade de ser a escolhida, e depois observa a máquina escolhida durante a vida útil dela. (Suponha que a escolha seja independente das vidas.)
    (a) Determine a densidade de T, onde T é a vida observada.
    (b) Suponha que o inspetor parou de observar a máquina escolhida depois de cem horas, com a máquina ainda funcionando. Qual a probabilidade condicional da máquina escolhida ter sido a máquina I?
    (c) Qual a distribuição de T se $\lambda_1 = \lambda_2 = \lambda$?
31. Suponhamos que os tempos que dois estudantes demoram para resolverem um problema sejam independentes e exponenciais com parâmetro $\lambda > 0$. Calcule a probabilidade do primeiro estudante demorar pelo menos duas vezes o tempo do segundo para resolver o problema.

32. Um ponto é selecionado, ao acaso (i.e., conforme a distribuição uniforme), do seguinte quadrado:



Sejam X e Y as coordenadas do ponto selecionado.
    (a) Qual a densidade conjunta de X e Y?
    (b) Obtenha a densidade marginal de X.
    (c) X e Y são independentes?

33. Suponhamos que X e Y tenham distribuição conjunta dada pela seguinte tabela:

Y \ X	1	2	3
1	0	1/5	0
2	1/5	1/5	1/5
3	0	1/5	0

(Por exemplo, $P(X = 1,\ Y = 1) = 0$ e $P(X = 2,\ Y = 1) = 1/5$.)
    (a) Determine as distribuições marginais de X e Y.
    (c) X e Y são independentes? Por quê?

§2.6

34. Sejam X e Y variáveis aleatórias independentes com distribuição uniforme em $[\theta - 1/2, \theta + 1/2]$, onde $\theta \in \mathbb{R}$. Prove que a distribuição de $X - Y$ não depende de θ, achando sua densidade.
35. Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes com densidade comum de Rayleigh com parâmetro $\theta > 0$:

$$f(x) \begin{cases} \dfrac{x}{\theta^2} \exp\left(-\dfrac{x^2}{2\theta^2}\right), & x > 0 \\ 0, & x \le 0. \end{cases}$$

(a) Determine a densidade conjunta de $Y_1, \ldots, Y_n$, onde $Y_i = X_i^2$.
(b) Qual a distribuição de $U = \min_{1 \le i \le n} X_i$? (Como se chama essa distribuição?)
(c) Calcule a distribuição de $Z = \dfrac{X_1}{X_2}$.

36. Sejam as variáveis aleatórias $X_1, \ldots, X_n$ independentes e exponenciais com, respectivamente, parâmetros $\alpha_1, \ldots, \alpha_n$.
(a) Mostre que a distribuição de $Y = \min_{1 \le i \le n} X_i$ é exponencial. Qual o parâmetro?
(b) Prove que para $k = 1, \ldots, n$.

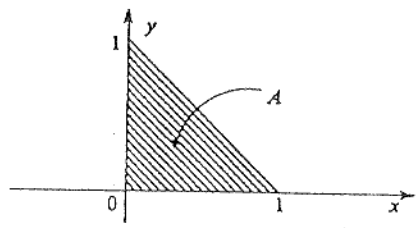
$$P(X_k = \min_{1 \le i \le n} X_i) = \frac{\alpha_k}{\alpha_1 + \ldots + \alpha_n}.$$

(*Sugestão.* X_k e $\min_{i \ne k} X_i$ são independentes e, pelo item (a), exponenciais. Considere o evento $[X_k < \min_{i \ne k} X_i]$.)

37. Seja X uma variável aleatória cuja função de distribuição F é uma função contínua na reta. Prove que a distribuição de $Y = F(X)$ é $U[0, 1]$. (*Sugestão.* Prove primeiro no caso de F estritamente crescente. Observe que não é suficiente provar no caso absolutamente contínua; vale também quando F é a função de Cantor.)
38. (a) As variáveis X, Y e Z são independentes, cada uma uniformemente distribuída no intervalo $[0, 1]$. Determine $P(X < Y < Z)$ e $P(X \le Y \le Z)$.
(b) Se X, Y e Z são independentes e identicamente distribuídas, e a função de distribuição comum F é contínua, qual é $P(X < Y < Z)$? Justifique sua resposta. (*Sugestão.* Exercício anterior.)
39. (a) Sejam X e Y independentes com distribuições de Poisson tendo, respectivamente, parâmetros λ_1 e λ_2. Mostre que $X + Y \sim$ Poisson $(\lambda_1 + \lambda_2)$.
(b) Mostre que se $X_1, \ldots, X_n$ são independentes tais que $X_i \sim$ Poisson (λ_i), $i = 1, \ldots, n$, então $X_1 + \ldots + X_n \sim$ Poisson $(\lambda_1 + \ldots + \lambda_n)$.
40. Certo supermercado tem duas entradas, A e B. Fregueses entram pela entrada A conforme um processo de Poisson com taxa média de 15 fregueses por minuto. Pela entrada B, entram fregueses conforme outro processo de Poisson, independente do primeiro, a uma taxa média de 10 por minuto.

(a) Seja X_t o número total de fregueses que entram no supermercado até o instante t (inclusive), para $t \ge 0$. Então $\{X_t : t \ge 0\}$ também é processo de Poisson (*não* é preciso provar). Qual o parâmetro deste processo? Justifique sua resposta.
(b) Seja T_1 o tempo em que o primeiro freguês entra pela entrada A, com V_1 o tempo em que o primeiro freguês entra pela entrada B. Ache a distribuição de $\min(T_1, V_1)$, o mínimo dos dois tempos.
(c) Determine a probabilidade de que o primeiro freguês a entrar no mercado entre pela entrada A.

41. Seja A o seguinte triângulo:



Suponha que X e Y tenham densidade conjunta $f(x, y) = cI_A(x, y)$.
(a) Determine o valor da constante c.
(b) Calcule a distribuição de X, a de Y e a de $Z = X + Y$.
(c) X e Y são independentes? Por quê?

42. Se X e Y são as coordenadas de um ponto selecionado, ao acaso, do círculo unitário $\{(x, y): x^2 + y^2 \le 1\}$, qual a distribuição da variável aleatória $Z = X^2 + Y^2$?
43. Sejam X e Y variáveis aleatórias independentes, tendo distribuição comum $U[0, 1]$.
(a) Qual a densidade da variável aleatória $Z = X + Y$?
(b) Ache a probabilidade da equação quadrática $Xt^2 + Yt + Z = 0$ ter raízes reais.
44. Dizemos que X tem distribuição de Weibull com parâmetros α e λ se X tem densidade

$$f_X(x) = \lambda\alpha x^{\alpha - 1} e^{-\lambda x^\alpha} I_{(0, \infty)}(x),$$

onde $\alpha > 0$ e $\lambda > 0$. Suponha que a vida útil de certo tipo de máquina, i.e., o tempo que ele funciona até pifar, possua distribuição Weibull (α, λ). Colocam-se em funcionamento, simultaneamente, n dessas máquinas. Qual a distribuição do tempo de espera até alguma máquina pifar? (Suponha independência das máquinas.)

45. Sejam X e Y variáveis aleatórias independentes, X tendo distribuição de Poisson com parâmetro $\lambda = 5$, e Y tendo distribuição uniforme em $[0, 1]$. Ache a densidade de $Z = X + Y$.
46. Lança-se um dado equilibrado duas vezes, independentemente. Sejam X e Y as variáveis aleatórias que representam os números obtidos em, respectivamente, o primeiro e o segundo lançamento.
    (a) Determine $P(X = Y)$.
    (b) Descreva a distribuição de $W = |X - Y|$.
    (c) Seja
    $$Z = \begin{cases} 1 \text{ se } X + Y \text{ é par} \\ 0 \text{ se } X + Y \text{ é ímpar.} \end{cases}$$
    Explique por que X e Z são, ou não são, independentes.
47. Escolhe-se um ponto ao acaso (i.e., conforme a distribuição uniforme) *dos lados* do quadrado de vértices $(1, 1)$, $(1, -1)$, $(-1, -1)$ e $(-1, 1)$. Sejam X e Y as coordenadas do ponto escolhido.
    (a) Determine a distribuição de $X + Y$.
    (b) Ache $P(W > 0)$, onde W é o máximo de X e Y.

§2.7

48. Sejam X e Y variáveis aleatórias independentes com distribuição comum $N(0, 1)$. Mostre que $U = \dfrac{X + Y}{\sqrt{2}}$ e $V = \dfrac{X - Y}{\sqrt{2}}$ também são independentes e $N(0, 1)$.
49. Sejam X e Y variáveis aleatórias independentes com distribuição comum $U[0, 1]$. Ache a densidade conjunta de W e Z, onde $W = X + Y$ e $Z = X - Y$. W e Z são independentes?
50. Suponha que X seja uma variável aleatória com distribuição $N(0, 1)$. Calcule a densidade de $Y = X^4$ e a de $Z = 1/X$. Y e Z possuem densidade conjunta? Por quê?
51. Seja X uma variável aleatória possuindo densidade $f(x)$.
    (a) Ache a densidade de $Y = |X|$ pelo método básico, obtendo a função de distribuição de X e derivando-a.
    (b) Ache a densidade de Y pelo método do jacobiano.
52. Suponha que X, Y e Z possuam densidade conjunta
    $$f(x, y, z) = \begin{cases} \dfrac{6}{(1 + x + y + z)^4} & \text{se } x > 0,\ y > 0 \text{ e } z > 0 \\ 0 & \text{caso contrário.} \end{cases}$$
    Obtenha a densidade da variável aleatória $W = X + Y + Z$ de duas maneiras diferentes (método básico e método do jacobiano).

53. Sejam X e Y variáveis aleatórias independentes com distribuição comum $\exp(\lambda)$. Prove que $Z = \dfrac{X}{X + Y} \sim U[0, 1]$.
54. (Extensão do método do jacobiano para o caso de k infinito.) Seja $\underset{\sim}{Y} = g(\underset{\sim}{X})$, onde $\underset{\sim}{X} = (X_1, \ldots, X_n)$ e $\underset{\sim}{Y} = (Y_1, \ldots, Y_n)$. Suponhamos que G, G_1, G_2, ... sejam subregiões abertas do $\mathbb{R}^n$ tais que $P(\underset{\sim}{X} \in \bigcup_n G_n) = 1$, e tais que $g|_{G_n}$ seja uma correspondência biunívoca entre G_n e G, $\forall\, n \geq 1$. Demonstre o seguinte teorema: se a função $h^{(n)}$, a inversa de $g|_{G_n}$, satisfaz as condições da função h do Teorema 2.1, $\forall\, n \geq 1$, então $\underset{\sim}{Y}$ tem densidade
    $$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \begin{cases} \sum_{n=1}^{\infty} f(h^{(n)}(\underset{\sim}{y}))\, |J_n(\underset{\sim}{x}, \underset{\sim}{y})|, & \underset{\sim}{y} \in G \\ 0, & \underset{\sim}{y} \notin G, \end{cases}$$
    onde $J_n(\underset{\sim}{x}, \underset{\sim}{y})$ é o jacobiano de $h^{(n)}$.
55. Se X possui densidade $f(x)$, qual a densidade de $Y = \cos X$? (*Sugestão*: Use o exercício 54.)
56. Sejam X e Y variáveis aleatórias independentes, tendo distribuição comum $U[0, 1]$, e sejam $R = \sqrt{2 \log (1/(1 - X))}$ e $\Theta = \pi(2Y - 1)$.
    (a) Mostre que $\Theta \sim U[-\pi, \pi]$ e que R tem distribuição de Rayleigh com densidade
    $$f(r) = \begin{cases} re^{-r^2/2}, & r > 0 \\ 0, & r \leq 0. \end{cases}$$
    (b) Mostre que Z e W, definidas por $Z = R \cos \Theta$ e $W = R \operatorname{sen} \Theta$, são independentes com distribuição comum $N(0, 1)$.
    (*Observação*. Este resultado é de interesse na simulação de variáveis aleatórias independentes e normais, pois indica como transformar números "pseudo-aleatórios" (simulações de variáveis aleatórias independentes e $U[0, 1]$) gerados por computador.)
57. (a) Se X e Y têm densidade conjunta $f(x, y)$, ache a densidade conjunta de W e Z, onde $W = aX + b$ e $Z = cY + d$, $a > 0$, $c > 0$, $b \in \mathbb{R}$, $d \in \mathbb{R}$.
    (b) Seja (X, Y) um vetor aleatório tendo distribuição normal bivariada com a densidade dada no exemplo 13 (§2.5). Qual a densidade de $(W, Z) = \left(\dfrac{X - \mu_1}{\sigma_1}, \dfrac{Y - \mu_2}{\sigma_2}\right)$? Que distribuição é essa?
    (c) Se (X, Y) tem distribuição uniforme no círculo unitário $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$, qual a distribuição conjunta de W e Z (como definidas no item (a))?

58. Suponha que X e Y sejam independentes, com $X \sim \Gamma(\alpha_1, 1)$ e $Y \sim \Gamma(\alpha_2, 1)$, onde $\alpha_1 > 0$ e $\alpha_2 > 0$. Mostre que $X + Y$ e X/Y são independentes e ache as suas distribuições.
59. Suponha que $X_1, \ldots, X_n$ formem uma amostra aleatória de uma distribuição com densidade $f(x)$. Mostre que $P(X_1 < \ldots < X_n) = 1/n!$ e que $P(X_i = X_j \text{ para algum par } (i, j) \text{ tal que } i \neq j) = 0$. (Veja o exercício 19.)

§2.8

60. Suponha que $X_1, \ldots, X_n$ sejam independentes e identicamente distribuídas, com densidade comum f. Mostre que a densidade conjunta de $U = \min_{1 \le i \le n} X_i$ e $V = \max_{1 \le i \le n} X_i$ é

$$f_{U,V}(u, v) = \begin{cases} n(n-1)\,[F(v) - F(u)]^{n-2} f(u) f(v), & \text{se } u < v \\ 0, & \text{se } u \ge v. \end{cases}$$

(*Sugestão.* Primeiro obtenha $P(u \le U,\ V \le v)$. Depois, calcule a derivada da função de distribuição conjunta.)

61. Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas, com distribuição uniforme em $[0, \theta]$, onde $\theta > 0$. Sejam

$$U = \min_{1 \le i \le n} X_i, \quad V = \max_{1 \le i \le n} X_i.$$

(a) Prove que a densidade conjunta de (U, V) é

$$f(u, v) = \begin{cases} n(n-1)\,(v-u)^{n-2}/\theta^n, & 0 \le u < v \le \theta \\ 0, & \text{caso contrário.} \end{cases}$$

(*Sugestão.* Exercício 60).

(b) Prove que a densidade de $V - U$ está dada por

$$f(w) = \begin{cases} \dfrac{n(n-1)\,w^{n-2}}{\theta^{n-1}}\left(1 - \dfrac{w}{\theta}\right), & 0 \le w \le \theta \\ 0, & \text{caso contrário.} \end{cases}$$

62. Se $X_1, \ldots, X_n$ são independentes com distribuição comum $U[0, 1]$, mostre que

$$-2n \log Y \sim \chi^2(2n),$$

onde Y é a média geométrica das X_i, definida por

$$Y = \left(\prod_{i=1}^{n} X_i\right)^{1/n}.$$

63. Mostre que se $X \sim t(1)$, então X tem distribuição de Cauchy.



CAPÍTULO III

ESPERANÇA MATEMÁTICA

§3.1 Preliminares: a integral de Stieltjes

Não é necessário ler esta seção primeiro para poder acompanhar as seções seguintes. O leitor que já tenha alguma familiaridade com as propriedades da integral de Riemann-Stieltjes, que são parecidas com as da integral de Stieltjes, pode omitir esta seção e consultá-la quando precisar. Ao leitor que não conheça a integral de Riemann-Stieltjes, sugere-se uma leitura rápida antes de prosseguir à seção seguinte.

Se φ é uma função contínua definida no intervalo $[a, b]$ e F é uma função de distribuição, define-se a *integral de Riemann-Stieltjes* de φ em $[a, b]$, em relação a F (ou ponderada por F), como o limite de "somas de Riemann" da forma

$$\sum_{i=1}^{n} \varphi(y_i)\,[F(x_{i+1}) - F(x_i)], \tag{3.1}$$

onde $a = x_1 < x_2 < \ldots < x_{n+1} = b$, y_i é um ponto arbitrário de $[x_i, x_{i+1}]$, e toma-se o limite quando a norma da partição tende a zero. (A partição consiste nos pontos x_i, e a sua norma é definida como a maior distância entre seus pontos vizinhos, ou seja, $\max_{1 \le i \le n} (x_{i+1} - x_i)$.) Tal limite existe e é finito sob as condições descritas, e é representado por

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

A função φ é chamada de *integrando*, F de *integrador*.

Não é preciso supor que F seja uma função de distribuição: se F é uma função monótona, ou mais geralmente, de variação limitada, o limite de (3.1) existe e é a integral de Riemann-Stieltjes. No entanto,

o caso em que o integrador é uma função de distribuição será do maior interesse para nós.

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

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

se o limite existe. Veremos adiante que para a definição de esperança de uma variável aleatória, a função $\varphi(x) = x$ assume a maior importância. Neste caso, pode-se mostrar que quando a integral imprópria de Riemann-Stieltjes existe, é um simples limite de somas da forma

$$\sum_{i=-\infty}^{\infty} y_i[F(x_{i+1}) - F(x_i)], \tag{3.2}$$

onde os pontos x_i formam uma seqüência crescente, $\lim_{i\to+\infty} x_i = +\infty$, $\lim_{i\to-\infty} x_i = -\infty$, $y_i \in [x_i, x_{i+1}]$ e toma-se o limite quando

$$\sup_{-\infty<i<\infty} (x_{i+1} - x_i) \to 0.$$

A definição da integral de Riemann-Stieltjes pode ser estendida a outras funções φ além das contínuas. Para uma função φ qualquer, define-se

$$\int_a^b \varphi(x)\mathrm{dF(x)}$$

como o limite das somas (3.1) quando a norma da partição tende a zero, se o limite existe. O problema desta definição é que até funções bem simples deixam de possuir integrais, como vemos no seguinte exemplo.

EXEMPLO 1. Inexistência da integral de Riemann-Stieltjes para um caso "simples". Seja F_0 a função de distribuição definida por

$$F_0(x) = \begin{cases} 1 & \text{se } x \geq 0, \\ 0 & \text{se } x < 0, \end{cases}$$

e consideremos a integral de F_0 em $[-1, 1]$ em relação a F_0, ou seja, F_0 é ao mesmo tempo integrador e integrando. Então, não existe o limite das somas (3.1), pois quando 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 ou 0 ou 1, dependendo do valor escolhido para y_i ser menor que 0, ou não.

Por causa desta deficiência da definição, a integral de Riemann-Stieltjes mostra-se insuficiente para nossos propósitos, e teremos que utilizar uma integral mais geral, a saber, a de *Lebesgue-Stieltjes*, que, doravante, será chamada simplesmente *integral de Stieltjes*. Não daremos a definição formal desta integral, pois depende de conceitos da Teoria da Medida, que não devem ser introduzidos a esta altura. Entretanto, faremos agora algumas observações sobre a integral de Stieltjes que deverão proporcionar ao leitor condições para poder calcular a integral em quase todos os casos de interesse. (Como caso particular, notemos que quanto ao exemplo acima, o item 6 abaixo implica $\int_{-1}^{1} F_0(x)dF_0(x) = 1$.)

No que se segue, o integrando φ é uma função real mensurável, e o integrador F é uma função de variação limitada, contínua à direita, i. e., a diferença entre duas funções monótonas crescentes, limitadas e contínuas à direita (veja Rudin [16], Teorema 6.27). Na grande parte dos casos em que usaremos a integral de Stieltjes, o integrador F será uma função de distribuição.

(1) *Notações.* $\int_a^b \varphi dF$ significa $\int_a^b \varphi(x)dF(x)$. Quando não aparecem limites de integração, a integral é sobre toda a reta:

$$\int \varphi dF = \int_{-\infty}^{\infty} \varphi dF = \lim_{\substack{a\to-\infty \\ b\to+\infty}} \int_a^b \varphi dF.$$

(2) Quando o integrando é contínuo em $[a, b]$, a integral de Stieltjes torna-se uma simples integral de Riemann-Stieltjes, e podemos utilizar as propriedades desta, tais como as descritas em Rudin [16], Capítulo 6. Com efeito, estas facilitarão nossa discussão das propriedades da esperança, pois nesse caso o integrando será a função contínua $\varphi(x) = x$ e, além disso, a integral será o limite das somas (3.2), se existir tal limite.

(3) $\int_a^b dF(x) = F(b) - F(a)$, i. e., a diferença de F sobre um intervalo é a integral da sua diferencial. Esta propriedade é análoga ao

teorema fundamental do cálculo: $\int_a^b \varphi'(x)dx = \varphi(b) - \varphi(a)$, onde $\varphi'(x) = \frac{d\varphi(x)}{dx}$, podendo-se reescrever a igualdade formalmente como

$$\int_a^b \varphi'(x)dx = \int_a^b d\varphi(x) = \varphi(b) - \varphi(a).$$

(4) A integral de Stieltjes é *linear*, tanto no integrando quanto no integrador. Em outras palavras, para $\varphi(x) = \alpha f(x) + \beta g(x)$ temos

$$\int_a^b \varphi dF = \alpha \int_a^b f dF + \beta \int_a^b g dF$$

e

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

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

$$\int_a^b \varphi dH = \alpha \int_a^b \varphi dF + \beta \int_a^b \varphi dG$$

e

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

onde valem as equações acima desde que as integrais estejam bem definidas e as somas tenham sentido.

(5) A integral de Stieltjes é *aditiva*. Por exemplo, se $a < b < c$ então

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

Isto vale também quando os intervalos são infinitos. Por exemplo,

$$\int \varphi dF = \int_{-\infty}^a \varphi dF + \int_a^\infty \varphi dF.$$

Novamente, estas equações são válidas quando as integrais estão bem definidas e as somas têm sentido. Para ter uma idéia do significado

dessa restrição, no caso da esperança, veja a observação 3, logo abaixo da Definição 3.2.

(6) Quando F é a função de distribuição de uma variável aleatória discreta X, a integral de Stieltjes reduz-se a uma série: se $P(X = x_i) = p(x_i) > 0$ e $\sum_i p(x_i) = 1$, i. e., se p é a função de probabilidade de X, então $p(x_i)$ é o salto de F em x_i e

$$\int \varphi dF = \int_{-\infty}^{\infty} \varphi(x)dF(x) = \sum_i \varphi(x_i)p(x_i).$$

Uma explicação intuitiva desta propriedade resulta da interpretação da diferencial $dF(x)$ como igual a $p(x_i)$ no ponto x_i e zero nos pontos x que não são pontos de salto de F (notemos que F cresce apenas em seus pontos de salto).

Quando a região de integração é um intervalo finito, temos

$$\int_a^b \varphi dF = \sum_{i: a < x_i \leq b} \varphi(x_i)p(x_i),$$

como é explicado nos itens 9 e 10 abaixo.

(7) Quando F é a função de distribuição de uma variável aleatória contínua tendo densidade f, então f é a derivada de F (em quase toda parte), temos $dF(x) = f(x)dx$ e, em analogia com uma propriedade da integral de Riemann-Stieltjes (Rudin [16], Teorema 6.17),

$$\int_a^b \varphi dF = \int_a^b \varphi(x)f(x)dx,$$

$$\int \varphi dF = \int_{-\infty}^{\infty} \varphi(x)f(x)dx.$$

(8) No caso de uma função de distribuição geral F, onde a decomposição de F nas partes discreta, absolutamente contínua e singular é dada por $F = F_d + F_{ac} + F_c$, temos, por linearidade,

$$\int_a^b \varphi dF = \int_a^b \varphi F_d + \int_a^b \varphi dF_{ac} + \int_a^b \varphi dF_s.$$

Em particular, quando F não possui parte singular ($F_s(x) = 0\ \forall x$), então

$$\int_a^b \varphi dF = \int_a^b \varphi dF_d + \int_a^b \varphi dF_{ac}$$

$$= \sum_{i:a<x_i\le b} \varphi(x_i)p(x_i) + \int_a^b \varphi(x)f(x)dx$$

e

$$\int \varphi dF = \int \varphi dF_d + \int \varphi dF_{ac} = \sum_i \varphi(x_i)p(x_i) + \int_{-\infty}^{\infty} \varphi(x)f(x)dx,$$

onde $p(x_i)$ é o salto de F em x_i e f é a derivada de F (f é também a derivada de F_{ac}, de modo que $f(x) = \dfrac{dF(x)}{dx}$ e $f(x) = \dfrac{dF_{ac}(x)}{dx}$, com as duas igualdades valendo em quase toda parte).

(9) Já que F é contínua à direita, a integral de Riemann-Stieltjes em $[a, b]$, se existe, ignora um eventual salto de F no ponto a, mas leva em consideração um eventual salto de F no ponto b. Por exemplo, definamos

$$F_1(x) = \begin{cases} 0 & \text{se } x < a \\ \dfrac{1}{2} & \text{se } a \le x < b \\ 1 & \text{se } b \le x. \end{cases}$$

Vemos que a função de distribuição F_1, embora possua dois pontos de salto pertencentes ao intervalo $[a, b]$, salta somente uma vez nesse intervalo:

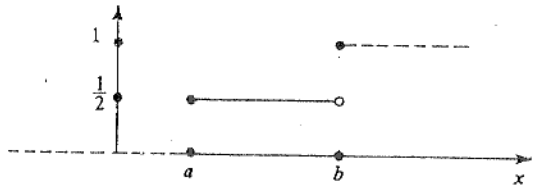


Gráfico de F_1 em [a, b]

Se φ for uma função contínua em $[a, b]$, a integral de Riemann-Stieltjes de φ em $[a, b]$, ponderada por F_1, será

$$\int_c^b \varphi dF_1 = \frac{1}{2}\varphi(b),$$

como vemos pela definição da integral como limite das somas (3.1). Portanto, a integração leva em conta o salto em b e ignora o salto em a. Como essa propriedade é também da integral de Stieltjes sobre o intervalo $(a, b]$ (veja o item seguinte), utilizaremos o símbolo $\int_a^b$ para representar esta integral:

$$\int_a^b \varphi dF \overset{def}{=} \int_{(a,b]} \varphi dF, \tag{3.3}$$

onde o termo à direita é a integral de Stieltjes em $(a, b]$.

Desta maneira, quando a integral de Riemann-Stieltjes existir, será igual à integral de Stieltjes e o termo à esquerda em (3.3) indicará as duas integrais.

(10) A integral de Stieltjes de φ em um intervalo é definida como a integral sobre a reta toda do produto de φ com o indicador do intervalo. (A integral de Riemann possui uma propriedade análoga, pois é fácil ver que a integral de Riemann de φ em $[a, b]$ satisfaz $\int_a^b \varphi(x)dx = \int_{-\infty}^{\infty} \varphi(x)I_{[a,b]}(x)dx$.)

Consideremos, por exemplo, $\int_{(a,b]} \varphi dF$. Já que $I_{(a,b]}$ toma o valor 1 no intervalo $(a, b]$ e 0 fora dele, temos

$$\varphi(x)I_{(a,b]}(x) = \begin{cases} 0 & \text{se } x \notin (a, b]. \\ \varphi(x) & \text{se } x \in (a, b]. \end{cases}$$

Tomando como exemplo a função de distribuição F_1 do item 9, que é uma função de distribuição discreta, temos (veja o item 6)

$$\int_a^b \varphi dF_1 = \int_{(a,b]} \varphi dF_1 = \int \varphi(x)I_{(a,b]}(x)dF_1(x) = \frac{1}{2}\varphi(a)I_{(a,b]}(a) +$$

$$+ \frac{1}{2}\varphi(b)I_{(a,b]}(b) = \frac{1}{2}\varphi(b).$$

Analogamente, temos

$$\int_{[a,b]} \varphi dF_1 = \int \varphi(x)I_{[a,b]}(x)dF_1(x) = \frac{1}{2}[\varphi(a) + \varphi(b)],$$

$$\int_{(a,b)} \varphi dF_1 = \int \varphi(x)I_{(a,b)}(x)dF_1(x) = \frac{1}{2}\varphi(a),$$

$$\int_{(a,b)} \varphi dF_1 = \int \varphi(x)I_{(a,b)}(x)dF_1(x) = 0.$$

Portanto, vemos que no caso discreto a integral inclui ou não a parcela $\varphi(x_i)p(x_i)$ dependendo de x_i pertencer ou não ao intervalo de integração.

Na Teoria da Medida estende-se esse conceito, definindo-se a integral sobre um boreliano B qualquer por

$$\int_B \varphi dF = \int \varphi I_B dF,$$

de maneira que no caso discreto, por exemplo,

$$\int_B \varphi dF = \sum_{i:x_i \in B} \varphi(x_i)p(x_i).$$

Notemos que a aditividade da integral (item 5) decorre agora da linearidade (item 4), pois se $a < b < c$ temos

$$I_{(a,c]}(x) = I_{(a,b]}(x) + I_{(b,c]}(x)$$

e, portanto,

$$\int_a^c \varphi dF = \int \varphi I_{(a,c]} dF = \int \varphi I_{(a,b]} dF + \int \varphi I_{(b,c]} dF =$$

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

§3.2 Esperança

Seja X uma variável aleatória discreta com função de probabilidade $p(x_i)$. O leitor certamente conhece a definição de *esperança matemática* para o caso discreto: a esperança de X é definida por

$$EX = \sum_i x_i p(x_i) = \sum_i x_i P(X = x_i). \tag{3.4}$$

Este valor está bem definido quando a soma não depende da ordem dos termos, em particular quando a série converge absolutamente (i.e., $\sum_i |x_i| p(x_i) < \infty$). A esperança de X é também chamada *média* de X, ou *valor esperado* de X. Com efeito, EX é uma *média ponderada*, onde os pesos são as probabilidades $p(x_i)$, i.e., EX é uma média dos valores possíveis de X, ponderada conforme a distribuição de X.

Uma possível explicação intuitiva desta definição reside na interpretação de probabilidade como limite de freqüências relativas: interpretando X novamente como um característico numérico do resultado de um experimento, suponhamos que vamos repetir (pelo menos conceitualmente) o experimento n vezes, independentemente, e observar os valores desse característico numérico. Nesses n experimentos, se n é grande, as observações tomarão o valor x_i com freqüência relativa de aproximadamente $p(x_i)$, para todo i, isto é, x_i aparecerá mais ou menos $np(x_i)$ vezes nas n observações. Portanto, o valor médio observado nesses n ensaios do experimento, i.e., a média aritmética dos n valores observados, será aproximadamente igual a

$$\frac{1}{n}\sum_i [x_i \cdot np(x_i)] = \sum_i x_i p(x_i).$$

Este valor será o limite quando $n \to \infty$, i.e., o valor médio obtido em n ensaios do experimento convergirá para EX quando $n \to \infty$. (Esta é uma versão da *Lei dos Grandes Números*. Veremos mais sobre essa lei no Capítulo 5.) Portanto, podemos dizer que "esperamos" obter a longo prazo um valor médio EX.

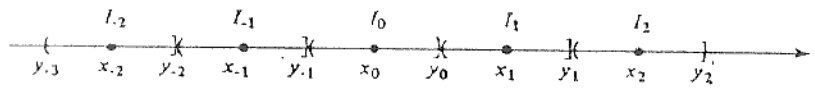
EXEMPLO 2. No exemplo do processo de Poisson (exemplo 8, §1.3), foi dito que λ era o número médio de chegadas durante um intervalo unitário de tempo. Verifiquemos isso agora, recordando que o número de chegadas em tal intervalo possui distribuição de Poisson de parâmetro λ. Em particular, temos $X_1 \sim$ Poisson (λ), portanto

$$EX_1 = \sum_{k=0}^{\infty} kP(X_1 = k) = \sum_{k=0}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} =$$

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

Observamos agora que quando X é discreta, o item 6 do §3.1 diz que $EX = \int x dF(x)$. É essa definição que adotaremos no caso geral, mantendo assim EX como uma média dos valores possíveis de X, ponderada conforme a distribuição de X. Para justificarmos o uso dessa definição, partindo de (3.4), vamos aproximar X por uma variável aleatória discreta.

Consideremos uma partição da reta em pequenos intervalos I_i centrados em pontos x_i, com extremos y_{i-1} e y_i:



Os intervalos não precisam ser de mesma comprimento; suponhamos apenas que eles sejam uniformemente pequenos, no sentido de que $\sup_i (y_i - y_{i-1}) \stackrel{def}{=} M$ seja pequeno.

Definamos uma variável aleatória discreta Y como a variável aleatória que assume o valor x_i quando X assume um valor em I_i, ou seja, em notação formal,

$$Y = \sum_i x_i I_{[X \in I_i]}.$$

Como $[Y = x_i] = [X \in I_i]$, que é um evento aleatório, Y é variável aleatória. É discreta, pois assume somente os valores x_i, e sua esperança é

$$EY = \sum_i x_i P(Y = x_i) = \sum_i x_i P(X \in I_i).$$

Ocorre que EY é finita se, e somente se $\sum_i |x_i| P(X \in I_i) < \infty$.

Mas Y é uma boa aproximação para X, pois $|Y - X| \leq \frac{M}{2}$ (para $\omega \in [X \in I_i]$, $X(\omega) \in I_i$ e $Y(\omega) = x_i$, logo $|Y(\omega) - X(\omega)| \leq (y_i - y_{i-1})/2 \leq M/2$).

Já que os valores de Y e X têm uma diferença sempre $\leq \frac{M}{2}$, é bem intuitivo requerer que nossa definição da esperança (= média) satisfaça $|EX - EY| \leq \frac{M}{2}$. Em outras palavras, queremos que EX seja o limite de EY quando $M \to 0$, se o limite existir, i.e.,

$$EX = \lim_{M \to 0} \sum_i x_i P(X \in I_i).$$

(Pode-se mostrar que existe o limite se $\sum_i |x_i| P(X \in I_i) < +\infty$ para alguma partição com $M < \infty$, ou seja, se EY é finita para alguma tal partição.)

Calculando o limite, temos

$$EX = \lim_{M \to 0} \sum_i x_i P(y_{i-1} < X \leq y_i) = \lim_{M \to 0} \sum_i x_i [F_X(y_i) - F_X(y_{i-1})] =$$

$$= \int_{-\infty}^{\infty} x dF_X(x),$$

onde a integral de Stieltjes é, de fato, de Riemann-Stieltzes (veja a discussão em torno da fórmula (3.2)).

DEFINIÇÃO 3.1. Seja X uma variável aleatória qualquer e F sua função de distribuição. A *esperança* de X é definida por

$$EX = \int_{-\infty}^{\infty} x dF(x),$$

quando a integral imprópria de Riemann-Stieltjes está bem definida.

DEFINIÇÃO 3.2. Se EX é finita, dizemos que X é *integrável*.

Observações. (1) Para muitos autores (por exemplo, Gnedenko [11]), "a esperança existe" quer dizer "X é integrável". Nós, entretanto, admitiremos a "existência" de esperanças infinitas – veja a observação 3 a seguir.

(2) Na literatura matemática, usam-se várias integrais para representar a esperança, as mais comuns sendo

$$EX = \int x dF_X(x) = \int x P_X(dx) = \int X dP.$$

As duas últimas integrais são de Lebesgue: a última é uma integral de Lebesgue no espaço de probabilidade $(\Omega, \mathcal{A}, P)$. Você não precisa entender essas integrais, mas no futuro poderá ser conveniente reconhecê-las como equivalentes à esperança.

(3) A esperança estará bem definida se $\int_0^{\infty} x dF(x)$ ou $\int_{-\infty}^{0} x dF(x)$ for finita. De fato, escrevendo

$$\int_{-\infty}^{\infty} x dF_X(x) = \int_{-\infty}^{0} x dF_X(x) + \int_0^{\infty} x dF_X(x) \stackrel{def}{=} I + II,$$

temos $I \leq 0$, $II \geq 0$ e

(i) se I e II são finitas, então X é integrável,
(ii) se I é finita e $II = +\infty$, então $EX = +\infty$,
(iii) se $I = -\infty$ e II é finita, então $EX = -\infty$,
(iv) se $I = -\infty$ e $II = +\infty$, então EX não está definida.

Assim, X é integrável se, e somente se, $\int |x| dF_X(x) < \infty$. Mais tarde veremos que $\int |x| dF_X(x) = E|X|$, de modo que X é integrável $\Leftrightarrow E|X| < \infty$.

(4) Se X tem densidade $f(x)$, então pelo item 7 do §3.1,

$$EX = \int x dF_X(x) = \int x f(x) dx.$$

Se a densidade f for integrável a Riemann (no sentido usual) então esta última integral também será de Riemann (Rudin [16], Teorema 6.17). Em outras palavras, você pode continuar a trabalhar com a integral de Riemann no caso contínuo.

No caso discreto, já vimos que $EX = \Sigma_i x_i p(x_i)$. Pelo item 6 do §3.1, esta definição concorda com a Definição 3.1

No caso geral, suponha $F_X = F_d + F_{ac} + F_s$. Então

$$EX = \int x dF_X(x) = \sum_i x_i p(x_i) + \int x f(x) dx + \int x dF_s(x),$$

pelo item 8, §3.1. Como a parte singular costuma ser nula, na prática a esperança reduz-se a uma série e/ou uma integral imprópria de (usualmente) Riemann.

EXEMPLO 3. Sejam $X \sim U[0,1]$, $Y = \min\left(X, \frac{1}{2}\right)$. Então

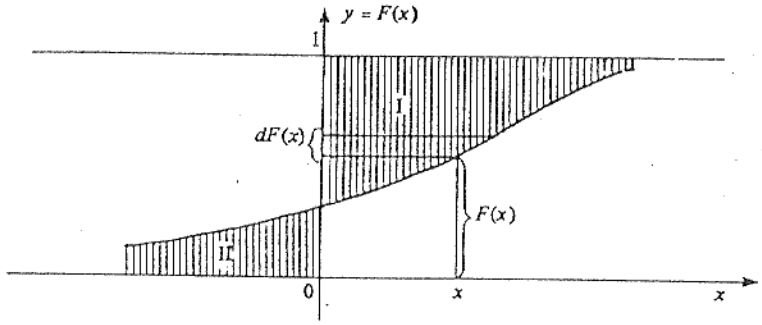
$$EY = \int y dF_Y(y) = \int y dF_d(y) + \int y dF_{ac}(y) =$$

$$= \frac{1}{2} p\left(\frac{1}{2}\right) + \int_0^{1/2} y \cdot 1 \, dy = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}.$$

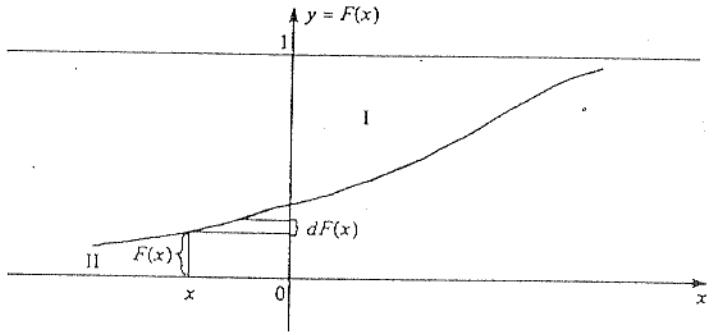
(Este exemplo é continuação do exemplo 8, §2.2.)

Ocorre que a esperança de X é sempre igual a uma integral de Riemann. Para entendermos isso, consideremos o seguinte argumento:

$EX = \int x dF(x)$ significa que a esperança é a integral da diferencial $x dF(x)$. Mas $x dF(x)$ é uma diferencial de área:



Para $x \geq 0$, $x dF(x)$ é uma diferencial de área da região I.



Para $x \leq 0$, $-x dF(x)$ é uma diferencial de área da região II.

Logo EX = área I − área II. Mas área $I = \int_0^\infty (1 - F(x))dx$ e área $II = \int_{-\infty}^0 F(x)dx$, logo deduzimos a seguinte

PROPOSIÇÃO 3.1. $EX = \int_0^\infty (1 - F(x))dx - \int_{-\infty}^0 F(x)dx$.

Prova formal. Vamos provar que $\int_0^\infty x dF(x) = \int_0^\infty (1 - F(x))dx$, deixando a outra parte ($\int_{-\infty}^0 x dF(x) = -\int_{-\infty}^0 F(x)dx$) para o leitor (a prova é análoga). Usaremos integração por partes, com a diferencial $d(xF(x)) = x dF(x) + F(x)dx$ (para uma justificação formal dessa integração, veja Rudin [16], Teorema 6.30):

$$\forall\, b > 0, \quad \int_0^b x dF(x) = bF(b) - \int_0^b F(x)dx = \int_0^b [F(b) - F(x)]dx.$$

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

$$\int_0^b x dF(x) = \int_0^b [F(b) - F(x)]dx \leq \int_0^\infty [1 - F(x)]dx, \ \forall\, b > 0,$$

de modo que

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

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

$$\int_0^b [F(b) - F(x)]dx \geq \int_0^\lambda [F(b) - F(x)]dx =$$

$$= \int_0^\lambda [F(b) - 1]dx + \int_0^\lambda [1 - F(x)]dx =$$

$$= \lambda[F(b) - 1] + \int_0^\lambda [1 - F(x)]dx,$$

e portanto,

$$\int_0^\infty x dF(x) = \lim_{b\to\infty} \int_0^b x dF(x) = \lim_{b\to\infty} \int_0^b [F(b) - F(x)]dx \geq$$

$$\geq \int_0^\lambda [1 - F(x)]dx + \lim_{b\to\infty} \lambda[F(b) - 1] = \int_0^\lambda [1 - F(x)]dx.$$

Já que isto vale para todo $\lambda > 0$, temos

$$\int_0^\infty x dF(x) \geq \lim_{\lambda\to\infty} \int_0^\lambda [1 - F(x)]dx = \int_0^\infty [1 - F(x)]dx. \ \square$$

COROLÁRIO 1. *Se X tomar somente valores não-negativos, ou seja, $X(\omega) \geq 0\ \forall\, \omega \in \Omega$, então $F_X(x) = 0$ para $x < 0$ e*

$$EX = \int_0^\infty [1 - F_X(x)]dx = \int_0^\infty P(X > x)dx.$$

Prova. Imediata. $\square$

Quando se conhece F_X ou $P(X > x)$, esta fórmula pode simplificar o cálculo de EX. Por exemplo, se $X \sim \exp(\lambda)$, i.e., X tem distribuição exponencial de parâmetro λ, então $P(X > x) = e^{-\lambda x}$, $x \geq 0$ e

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

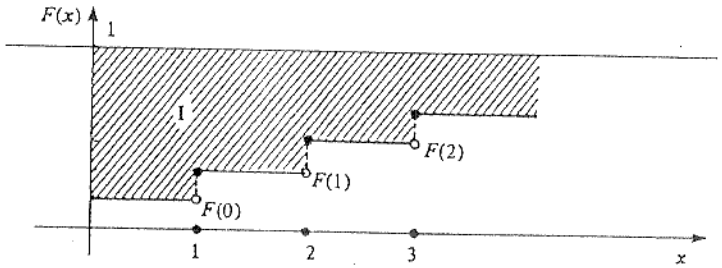
O cálculo foi simplificado, pois evitou-se uma integração por partes (o que, afinal, já foi feito na prova da proposição.)

A fórmula do Corolário 1 possui uma forma simples no caso de X ser discreta e assumir valores inteiros:

COROLÁRIO 2. *Se a variável aleatória X assumir somente valores inteiros não-negativos, então*

$$EX = \sum_{n=0}^\infty P(X > n) = \sum_{n=1}^\infty P(X \geq n).$$

Prova.



Pela proposição e pelo desenho,

$$EX = \text{área } I = 1 - F(0) + 1 - F(1) + 1 - F(2) + \ldots =$$

$$= \sum_{n=0}^\infty [1 - F(n)] = \sum_{n=0}^\infty P[X > n].$$

Como X assume apenas valores inteiros, temos

$$P(X > n) = P(X \geq n + 1),$$

logo,

$$\sum_{n=0}^\infty P(X > n) = \sum_{n=0}^\infty P(X \geq n + 1) = \sum_{n=1}^\infty P(X \geq n). \ \square$$

EXEMPLO 4. Lançar uma moeda independentemente até obter a primeira cara. Seja X o número de lançamentos necessários e p a probabilidade de obter cara em um dado lançamento. Então X toma apenas os valores $1, 2, 3, \ldots$ e $P(X \geq n)$ é a probabilidade de não obter cara nos lançamentos $1, 2, \ldots, n-1$, i.e., a probabilidade de sair coroa nos lançamentos 1 até $n-1$ $(=(1-p)^{n-1})$. Logo,

$$EX = \sum_{n=1}^{\infty} P(X \geq n) = \sum_{n=1}^{\infty} (1-p)^{n-1} = \sum_{n=0}^{\infty} (1-p)^n = \frac{1}{p}.$$

Notemos que o cálculo direto $EX = \sum_{n=1}^{\infty} nP(X=n) = \sum_{n=1}^{\infty} n(1-p)^{n-1}p$, utilizando a função de probabilidade da distribuição geométrica, é mais árduo.

Vemos, portanto, que a esperança do tempo de espera até o primeiro sucesso em uma seqüência de ensaios de Bernoulli é $1/p$, onde p é a probabilidade de sucesso em cada ensaio.

Observação. Como $|X| \geq 0$, o Corolário 1 implica que

$$E|X| = \int_0^{\infty} P(|X| > x)dx = \int_0^{\infty} [P(X > x) + P(X < -x)]dx =$$

$$= \int_0^{\infty} [1 - F_X(x) + F_X((-x)-)]dx.$$

É bom recordar que $F_X((-x)-)$ é o limite da função $F_X(y)$ quando $y \uparrow (-x)$. Logo, $F_X((-x)-) = F_X(-x)$ quando $-x$ é ponto de continuidade de F_X. Por isso, $F_X((-x)-)$ e $F_X(-x)$ são funções monótonas (decrescentes em x) iguais, exceto em um número finito ou enumerável de pontos. Daí concluímos que

$$\int_0^{\infty} F_X((-x)-)dx = \int_0^{\infty} F_X(-x)dx = \text{(mudança de variável)} =$$

$$= \int_{-\infty}^{0} F_X(x)dx.$$

Resumindo, temos

$$E|X| = \int_0^{\infty} [1 - F_X(x)]dx + \int_{-\infty}^{0} F_X(x)dx.$$

Então, pela prova da Proposição 3.1,

$$E|X| = \text{área } I + \text{área } II = \int_0^{\infty} xdF_X(x) - \int_{-\infty}^{0} xdF_X(x) =$$

$$= \int_0^{\infty} |x|dF_X(x) + \int_{-\infty}^{0} |x|dF_X(x) = \int_{-\infty}^{\infty} |x|dF_X(x).$$

Assim provamos que (veja a observação 3, abaixo da Definição 3.2) $E|X| = \int_{-\infty}^{\infty} |x|dF_X(x)$, e que X é integrável se, e somente se, $E|X| < \infty$.

(*Nota*: daqui a pouco veremos que a fórmula

$$E\varphi(x) = \int_{-\infty}^{\infty} \varphi(x)dF_X(x)$$

vale para φ geral.)



§3.3 Propriedades da esperança

E1. *Se $X = c$ (i.e., $X(\omega) = c \ \forall \omega \in \Omega$), então $EX = c$.*

Prova. X é uma variável aleatória discreta e $EX = cP(X = c) = c \cdot 1 = c$. □

E2. *Se $X \leq Y$ então $EX \leq EY$, se as esperanças estão bem definidas.* (Basta uma das esperanças ser finita, ou $EY = -\infty$, ou $EX = +\infty$).

Prova. $Y \leq z \Rightarrow X \leq z$, logo $[Y \leq z] \subset [X \leq z]$. Portanto, $F_Y(z) \leq F_X(z)$ e $1 - F_Y(z) \geq 1 - F_X(z)$. Pela Proposição 3.1,

$$EY = \int_0^{\infty} (1 - F_Y(z))dz - \int_{-\infty}^{0} F_Y(z)dz \geq \int_0^{\infty} (1 - F_X(z))dz -$$

$$- \int_{-\infty}^{0} F_X(z)dz = EX. \ \square$$

E3. *Linearidade.*

(i) *Se EX está bem definida, então $E(aX + b) = aEX + b$ para todo $a, b \in \mathbb{R}$ (convenção: $0 \cdot \infty = 0$).*

(ii) *$E(aX + bY) = aEX + bEY$, quando o termo à direita da igualdade tem sentido.* (Sobre a restrição: se, por exemplo, $EX = +\infty$, então $0 = E(X - X) \neq EX - EX$, pois $+\infty - \infty$ não tem sentido.)

Prova. (i) Se $a=0$, então $E(aX+b)=Eb=b=aEX+b$. Se $a>0$, então

$$F_{aX+b}(x)=P(aX+b\le x)=P\left(X\le\frac{x-b}{a}\right)=F_X\left(\frac{x-b}{a}\right),$$

logo

$$E(aX+b)=\int_0^\infty (1-F_{aX+b}(x))dx-\int_{-\infty}^0 F_{aX+b}(x)dx=$$

$$=\int_0^\infty\left(1-F_X\left(\frac{x-b}{a}\right)\right)dx-\int_{-\infty}^0 F_X\left(\frac{x-b}{a}\right)dx=$$

$$=\left(\text{faça } y=\frac{x-b}{a}\right)=$$

$$=a\int_{-b/a}^\infty (1-F_X(y))dy-a\int_{-\infty}^{-b/a}F_X(y)dy=$$

$$=a\int_0^\infty (1-F_X(y))dy-a\int_{-\infty}^0 F_X(y)dy+$$

$$+a\int_{-b/a}^0 (1-F_X(y))dy+a\int_{-b/a}^0 F_X(y)dy=$$

$$=aEX+a\int_{-b/a}^0 dy=aEX+b.$$

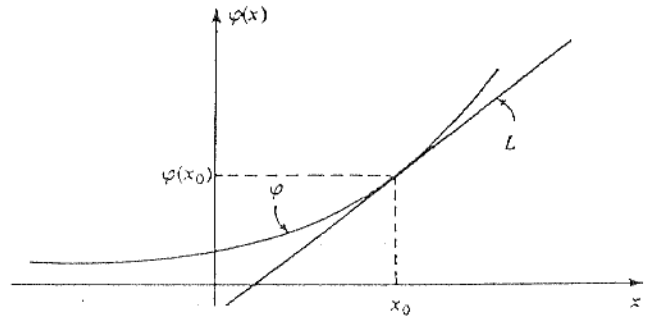
O caso $a<0$ é análogo, e (i) está provado. Para (ii), resta provar $E(X+Y)=EX+EY$ se o termo à direita tem sentido. Veremos mais tarde quando consideramos esperanças de funções de vetores aleatórios. □

E4. *Desigualdade de Jensen. Seja φ uma função convexa definida na reta. Se a variável aleatória X é integrável, então*

$$E\varphi(X)\ge\varphi(EX).$$

(*Notação*: $E\varphi(X)=E\{\varphi(X)\}$.)

Prova. Dado x_0 e o ponto $\varphi(x_0)$ do gráfico de φ, pela convexidade existe uma reta L que passa por $\varphi(x_0)$ e deixa a curva φ acima dela:



Seja para algum $\lambda\in\mathbb{R}$, $y-\varphi(x_0)=\lambda(x-x_0)$ a equação desta reta L. Então

$$\varphi(x)\ge L(x)=\varphi(x_0)+\lambda(x-x_0),\ \forall x.$$

Portanto,

$$E\varphi(X)\overset{E2}{\ge}EL(X)\overset{E3}{=}\varphi(x_0)+\lambda(EX-x_0).$$

Tomando-se $x_0=EX$, vem $E\varphi(X)\ge\varphi(EX)$. □

(*Exercício.* Mostre que se φ é côncava, então $E\varphi(X)\le\varphi(EX)$.)

Conseqüências das propriedades. (1)E2 diz que $X\le Y$ implica $EX\le EY$. Em particular, se $X\ge 0$ então $EX\ge 0$.
Sejam X e Y variáveis aleatórias tais que $Y\ge 0$, Y é integrável, e $|X|\le Y$. Então $0\le|X|\le Y$ implica

$$0\le E|X|\le EY<+\infty,$$

i.e., X é integrável. Em outras palavras, se X é dominada por uma variável aleatória integrável, então também X é integrável. Em particular, *se X é limitada, então ela é integrável*, pois $|X|\le b<+\infty$ implica $E|X|\le b<\infty$.

(2) *Critério para integrabilidade. Seja X uma variável aleatória qualquer. Então* $\sum_{n=1}^\infty P(|X|\ge n)\le E|X|\le 1+\sum_{n=1}^\infty P(|X|\ge n)$, *e portanto X é integrável se, e só se,* $\sum_{n=1}^\infty P(|X|\ge n)<\infty$.

Prova. Se $x \geq 0$, seja $[x]$ a parte inteira de x (o maior inteiro menor que ou igual a x). Então a variável aleatória $[|X|]$ assume o valor k quando $k \leq |X| < k+1$ e

$$0 \leq [|X|] \leq |X| \leq [|X|] + 1,$$

logo, por E2 e E3,

$$0 \leq E[|X|] \leq E|X| \leq 1 + E[|X|].$$

Mas pelo Corolário 2 da Proposição 3.1,

$$E[|X|] = \sum_{n=1}^{\infty} P([|X|] \geq n) = \sum_{n=1}^{\infty} P(|X| \geq n),$$

logo

$$\sum_{n=1}^{\infty} P(|X| \geq n) \leq E|X| \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n). \ \square$$

(3) A desigualdade de Jensen diz que $E\varphi(X) \geq \varphi(EX)$, se φ é função convexa e EX é finita. Por exemplo:

(a) Seja $\varphi(x) = |x|$. Então $E|X| \geq |EX|$ (o que é também conseqüência de E2 e E3: $-|X| \leq X \leq |X| \Rightarrow -E|X| \leq EX \leq E|X|$).

(b) Seja $\varphi(x) = x^2$. Então $EX^2 \geq (EX)^2$.

(c) Seja $\varphi(x) = |x|^p$, onde $p \geq 1$. Então $E|X|^p \geq |EX|^p$. Fazendo $Y = |X|$ e aplicando Jensen a Y, obtemos uma desigualdade mais refinada:

$$E|X|^p \geq (E|X|)^p \geq (\text{pelo item } a) \geq |EX|^p.$$

Observação. Para a validade da desigualdade de Jensen, basta que a função φ seja convexa em um intervalo (a, b) tal que $P(a < X < b) = 1$ (a prova é a mesma, mas veja o exercício 13). Por exemplo, se X é uma variável aleatória positiva ($X > 0$ ou, mais geralmente, $P(X > 0) = 1$), podemos aplicar Jensen à função $\varphi(x) = \frac{1}{x}$, com $(a, b) = (0, \infty)$. Neste caso, a conclusão é que

$$E\left(\frac{1}{X}\right) \geq \frac{1}{EX}.$$

Sob a mesma condição $P(X > 0) = 1$, podemos aplicar Jensen à função côncava $\varphi(x) = \log x$, obtendo

$$E \log X \leq \log(EX).$$

§3.4 Esperanças de funções de variáveis aleatórias

Seja X uma variável aleatória, $\varphi(x)$ uma função real mensurável, $Y = \varphi(X)$. Então Y é uma variável aleatória cuja esperança é dada por

$$EY \overset{def}{=} \int y dF_{\varphi(X)}(y) = \int_0^{\infty} [1 - F_{\varphi(X)}(y)]dy - \int_{-\infty}^{0} F_{\varphi(X)}(y)dy,$$

pela Proposição 3.1.

Para usar estas fórmulas, é preciso obter a distribuição de $Y = \varphi(X)$, o que às vezes não é fácil (estamos supondo que conhecemos a distribuição de X). Mas lembrando-se que a esperança de Y é uma média ponderada dos valores possíveis de Y, onde os "pesos" são determinados pela distribuição de Y, surge uma pergunta natural: será que a esperança de $\varphi(X)$ é uma média ponderada dos valores de $\varphi(X)$ para os valores possíveis de X, onde os pesos são determinados pela distribuição de X? Ou seja, será que $\int y dF_{\varphi(X)}(y) = \int \varphi(x)\, dF_X(x)$? A resposta é afirmativa, como é fácil ver no caso discreto:

Seja X uma variável aleatória discreta com função de probabilidade $p(x_i)$, onde $\sum_i p(x_i) = 1$. É óbvio que a variável aleatória $Y = \varphi(X)$ é também discreta e toma somente os valores $\varphi(x_i)$. Então sejam $y_1, y_2, \ldots$ os valores possíveis de Y, supondo por conveniência que os y_j sejam distintos, de modo que $y_j = \varphi(x_i)$ para pelo menos um i (é possível que exista apenas um y_j, como, por exemplo, se $\varphi(x) = c$ constante). Então temos

$$P(Y = y_j) = \sum_{i:\, \varphi(x_i) = y_j} p(x_i).$$

Obtemos agora a esperança de Y:

$$EY = \int y dF_Y(y) = \sum_j y_j P(Y = y_j) = \sum_j \left(y_j \sum_{i:\, \varphi(x_i) = y_j} p(x_i) \right) =$$

$$= (\text{substituindo}) = \sum_j \sum_{i:\, \varphi(x_i) = y_j} \varphi(x_i) p(x_i) \overset{(*)}{=} \sum_i \varphi(x_i) p(x_i) =$$

$$= \int \varphi(x) dF_X(x),$$

onde a equação $(*)$ é válida desde que a ordem dos termos da série não afete o valor da soma (e, em particular, se EY está bem definida).

Passaremos agora ao caso geral.

TEOREMA 3.1. *Seja X uma variável aleatória, $\varphi(x)$ uma função real mensurável. Então*

$$E\varphi(X) \stackrel{def}{=} \int y dF_{\varphi(X)}(y) = \int \varphi(x) dF_X(x),$$

onde a existência de uma das integrais implica a existência da outra e a igualdade das duas.

Prova (parcial). Prova-se o caso geral com a Teoria da Medida. Mas já provamos para $\varphi(x) = |x|$ e podemos provar o teorema para polinômios usando somente a integral de Riemann, baseando-nos na Proposição 3.1. Mostremos que

$$EX^k = \int x^k dF_X(x), \quad \text{para } k = 1, 2, \ldots$$

Para conveniência de notação, seja $F = F_X$. Primeiro, seja k par. Então

$$\begin{aligned} EX^k &= \int_0^\infty P(X^k > t)dt \quad (\text{pois } X^k \geq 0) = \\ &= \int_0^\infty P(X > \sqrt[k]{t})dt + \int_0^\infty P(X < -\sqrt[k]{t})dt = \\ &= \int_0^\infty [1 - F(\sqrt[k]{t})]dt + \int_0^\infty F((-\sqrt[k]{t})-)dt = \\ &= (\text{fazendo } t = s^k) = \\ &= \int_0^\infty [1 - F(s)]ks^{k-1}\,ds + \int_0^\infty F((-s)-)ks^{k-1}\,ds. \end{aligned}$$

Como as funções monótonas $F((-s)-)$ e $F(-s)$ são iguais exceto em um número finito ou enumerável de pontos, podemos desprezar o segundo sinal negativo. Fazendo $u = -s$ na segunda integral, temos

$$EX^k = \int_0^\infty [1 - F(s)]ks^{k-1}\,ds - \int_{-\infty}^0 F(u)ku^{k-1}\,du. \qquad (3.5)$$

Pelo método de prova da Proposição 3.1 (integração por partes), temos

$$\begin{aligned} \int_{-\infty}^\infty x^k dF(x) &= \int_0^\infty [1 - F(x)]dx^k - \int_{-\infty}^0 F(x)dx^k = \\ &= k\left\{\int_0^\infty [1 - F(x)]x^{k-1}\,dx - \int_{-\infty}^0 F(x)x^{k-1}\,dx\right\}. \end{aligned}$$

Logo $EX^k = \int x^k dF_X(x)$ para k par. Para k ímpar, a prova é análoga.

Portanto o teorema está provado se φ é polinômio, pela linearidade da esperança e da integral de Stieltjes. □

COROLÁRIO. $EX^k = k\{\int_0^\infty [1 - F_X(x)]x^{k-1}\,dx - \int_{-\infty}^0 F_X(x)x^{k-1}\,dx\}$, *para* $k = 1, 2, \ldots$. (É a fórmula (3.5).)

EXEMPLO 5. O corolário facilita muito na distribuição exponencial. Se $X \sim \exp(\lambda)$, então $P(X > x) = e^{-\lambda x}$, $x \geq 0$, e

$$EX^k = k\int_0^\infty x^{k-1} e^{-\lambda x}\,dx.$$

Podemos calcular todos os momentos (veja o §3.5) por iteração, sem integrar por partes. Com efeito, já vimos que $EX = \frac{1}{\lambda}$; portanto,

$$\begin{aligned} EX^2 &= 2\int_0^\infty xe^{-\lambda x}\,dx = \frac{2}{\lambda}\int_0^\infty x\lambda e^{-\lambda x}\,dx = \frac{2}{\lambda}\int_0^\infty xf(x)dx = \\ &= \frac{2}{\lambda}EX = \frac{2}{\lambda^2}, \text{ e} \\ EX^3 &= 3\int_0^\infty x^2 e^{-\lambda x}\,dx = \frac{3}{\lambda}\int_0^\infty x^2\lambda e^{-\lambda x}\,dx = \\ &= \frac{3}{\lambda}EX^2 = \frac{6}{\lambda^3}. \end{aligned}$$

Por indução,

$$EX^k = \frac{k!}{\lambda^k}.$$

EXEMPLO 6. Voltando ao exemplo 3, suponha que $X \sim U[0, 1]$ e $Y = \min\left(X, \frac{1}{2}\right)$. Então $Y = \varphi(X)$, onde

$$\varphi(x) = \min\left(x, \frac{1}{2}\right) = \begin{cases} x, & \text{se } x < \frac{1}{2} \\ \frac{1}{2}, & \text{se } x \geq \frac{1}{2}. \end{cases}$$

Calculemos a esperança de Y, usando o Teorema 3.1 e uma propriedade da integral de Stieltjes no caso contínuo (§3.1, item 7):

$$EY = \int \varphi(x) dF_X(x) = \int \varphi(x) f_X(x) dx = \int_0^1 \varphi(x) dx =$$

$$= \int_0^{1/2} x dx + \int_{1/2}^1 \frac{1}{2} dx = \frac{1}{8} + \frac{1}{4} = \frac{3}{8},$$

assim concordando com o resultado já obtido.

Notemos que neste exemplo, a existência da densidade de X simplificou os cálculos. É bom salientar o significado do teorema nos casos discreto e contínuo:

Caso discreto. Se X tiver função de probabilidade $p(x_i)$, então

$$E\varphi(X) = \sum_i \varphi(x) p(x_i).$$

Caso contínuo. Se X tiver densidade $f(x)$, então

$$E\varphi(X) = \int \varphi(x) f(x) dx.$$

§3.5 Momentos

Seja X uma variável aleatória. O valor $E(X - b)^k$, se existe, é chamado *k-ésimo momento de X em torno de b*, para $b \in \mathbb{R}$, $k = 1, 2, 3, \ldots$.

O k-ésimo momento em torno de zero, EX^k, é chamado simplesmente de *k-ésimo momento* de X ou *momento de ordem k* de X.

Se X é integrável, então o k-ésimo momento em torno da média, $E(X - EX)^k$, se chama *k-ésimo momento central* de X.

É claro que o primeiro momento é a esperança e o primeiro momento central é nulo: $E(X - EX) = 0$. O segundo momento central é chamado *variância* de X:

$$\begin{aligned} \operatorname{Var} X &\stackrel{def}{=} E(X - EX)^2 = E(X^2 - 2XEX + (EX)^2) = \\ &= (\text{por linearidade}) = EX^2 - 2EXEX + (EX)^2 = \\ &= EX^2 - (EX)^2. \end{aligned}$$

Notação. $\operatorname{Var} X = V(X) = \sigma_X^2 = \sigma^2(X)$. $\sigma_X = \sqrt{\operatorname{Var} X}$ é o *desvio-padrão* de X.

Para $t > 0$, $E|X|^t$ é chamado *t-ésimo momento absoluto* de X. Os momentos absolutos possuem a seguinte propriedade de monotonia:

PROPOSIÇÃO 3.2. *Seja X uma variável aleatória. A função*

$$f(t) = E^{1/t}|X|^t$$

é não decrescente em t para $t > 0$. (Notação: $E^{1/t}|X|^t = [E(|X|^t)]^{1/t}$).

Prova. Se $0 < s < t$, a função $\varphi(y) = |y|^{t/s}$ é convexa, pois $t/s > 1$. Se Y é integrável, a desigualdade de Jensen implica que

$$E|Y|^{t/s} \geq |EY|^{t/s}.$$

Faça $Y = |X|^s$: se $|X|^s$ é integrável, então

$$E|X|^t \geq E^{t/s}|X|^s,$$

i. e., $E^{1/t}|X|^t \geq E^{1/s}|X|^s$.

Se $|X|^s$ não é integrável, então $|X|^t$ também não o é, pois

$$s < t \Rightarrow |X|^s \leq 1 + |X|^t$$

e, portanto

$$E|X|^s = +\infty \Rightarrow +\infty \leq 1 + E|X|^t \Rightarrow E|X|^t = +\infty.$$

Em todo caso, temos

$$E^{1/s}|X|^s \leq E^{1/t}|X|^t \quad \text{para } 0 < s < t,$$

i. e., $E^{1/t}|X|^t$ é não-decrescente em t, para $t > 0$. □

COROLÁRIO. *Se $E|X|^t$ é finita para algum $0 < t < \infty$, então $E|X|^s$ é finito para todo s tal que $0 < s < t$.* (Por exemplo, se $EX^2 < \infty$ então X é integrável. Além disso, se o k-ésimo momento é finito, então todos os momentos de ordem menor que k também são finitos.)

Propriedades (da esperança, variância, momentos – continuação).

E5. *Se* $X = c$, *Var* $X = 0$ (*i.e., uma constante não varia*).

Prova. $EX = c$. Var $X = E(X - c)^2 = E(0) = 0$. □

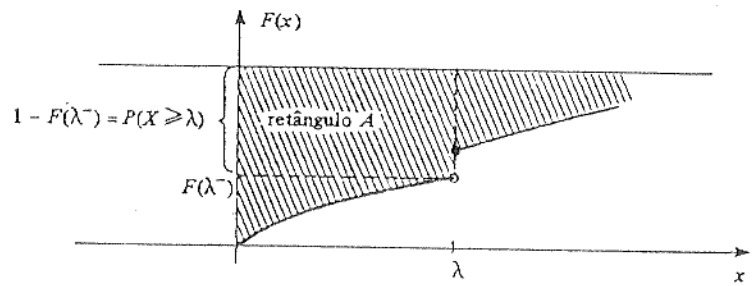
E6. *Var* $(X + b) =$ *Var* X *e Var* $(aX + b) = a^2$ *Var* X, *para toda* $a, b \in \mathbb{R}$.

Prova. $E(aX + b) = aEX + b$,

$$Var\,(aX + b) = E(aX + b - aEX - b)^2 = E(a^2(X - EX)^2) = $$
$$= a^2\, Var\, X. \square$$

E7. *Desigualdade "básica" (ou desigualdade generalizada de Tchebychev). Seja* X *uma variável aleatória não-negativa* $(X \geq 0)$. *Para todo* $\lambda > 0$, $P(X \geq \lambda) \leq \frac{1}{\lambda} EX$.

Prova. Consideremos o seguinte desenho, admitindo a possibilidade de um salto de $F = F_X$ em λ:



Indicando com I a região hachurada, e como $X \geq 0$, temos

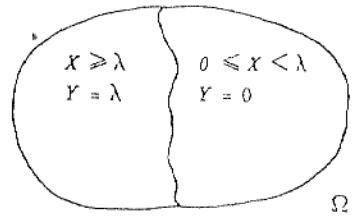
$$EX = \int_0^\infty P(X > x)dx = \text{área } I \geq \text{área (retângulo } A) =$$
$$= \lambda P(X \geq \lambda),$$

portanto

$$P(X \geq \lambda) \leq \frac{1}{\lambda} EX. \square$$

Observação. Eis uma prova alternativa, que será generalizada adiante na prova da desigualdade de Kolmogorov (Capítulo 5):

Seja $Y = \lambda I_{[X \geq \lambda]}$. Então $0 \leq Y \leq X$, como vemos considerando os eventos complementares $[X \geq \lambda]$ e $[0 \leq X < \lambda]$:



Portanto, $EY \leq EX$. Mas Y é uma variável aleatória discreta e $EY = 0 \cdot P(Y = 0) + \lambda P(Y = \lambda) = \lambda P(X \geq \lambda)$. Resumindo, temos

$$\lambda P(X \geq \lambda) \leq EX \quad \text{ou} \quad P(X \geq \lambda) \leq \frac{1}{\lambda} EX. \square$$

Algumas conseqüências são:

(a) *Desigualdade (clássica) de Tchebychev (ou de Bienaymé-Tchebychev). Se* X *é integrável,*

$$P(|X - EX| \geq \lambda) \leq \frac{\text{Var } X}{\lambda^2}, \quad \forall \lambda > 0.$$

Prova. $P(|X - EX| \geq \lambda) = P((X - EX)^2 \geq \lambda^2) \leq \frac{1}{\lambda^2} E(X - EX)^2 =$

$$= \frac{\text{Var } X}{\lambda^2} \quad \square$$

(b) *Desigualdade de Markov. Seja* X *uma variável aleatória qualquer. Então para todo* $t > 0$,

$$P(|X| \geq \lambda) \leq \frac{E|X|^t}{\lambda^t}, \quad \forall \lambda > 0.$$

Prova. $P(|X| \geq \lambda) = P(|X|^t \geq \lambda^t) \leq \frac{1}{\lambda^t} E|X|^t$. □

(c) *Se* $Z \geq 0$ e $EZ = 0$, *então* $P(Z = 0) = 1$ (*i.e.,* $Z = 0$ quase certamente).

Prova. $P\left(Z \geq \frac{1}{n}\right) \leq nEZ = 0$. Mas $[Z > 0] = \bigcup_n \left[Z \geq \frac{1}{n}\right]$, **logo**

$P(Z>0)=\lim_{n\to\infty} P\left(Z\geq\frac{1}{n}\right)=0$ (notemos que os eventos $\left[Z\geq\frac{1}{n}\right]$ crescem com n). Portanto, $P(Z=0)=1-P(Z>0)=1$. □

Observação. O item (*c*) implica que, quando Var $X=0$, temos $E(X-EX)^2=0$ e $P((X-EX)^2=0)=1$, logo $P(X=EX)=1$. Em outras palavras, se Var $X=0$ então X é constante, com probabilidade 1 (é constante quase certamente).

E8. *Se X e Y são variáveis aleatórias em $(\Omega, \mathscr{A}, \mathscr{P})$ tais que $E|X|^t<\infty$ e $E|Y|^t<\infty$, então $E|X+Y|^t<\infty$.*

Como $E|X|^t<\infty$ óbviamente implica $E|aX|^t<\infty$, $\forall a\in\mathbb{R}$, esta propriedade diz que a classe das variáveis aleatórias em $(\Omega, \mathscr{A}, P)$ possuidoras de t-ésimo momento absoluto finito, é um espaço vetorial ou *espaço linear*. (Com t substituído por p, estes são os *espaços* L^p de Análise.)

Destaquemos dois casos particulares desta propriedade, os correspondentes a $t=1$ e $t=2$: (i) *se X e Y são integráveis, então $X+Y$ é integrável* e (ii) *se X e Y têm variâncias finitas, então $X+Y$ também o tem* (lembremos que X tem variância finita se, e somente se, $EX^2<\infty$).

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

$$E|X+Y|^t\leq 2^t(E|X|^t+E|Y|^t).\ \square$$

Consideremos agora dois resultados que são de interesse para a Estatística. Suponhamos que se deseje escolher uma constante real c para "predizer" o valor de uma variável aleatória X. Qual c é o melhor preditor? Se queremos minimizar nosso erro absoluto médio (i. e., a média de $|X-c|$), o melhor preditor é a mediana (veja a definição adiante). Mas se queremos minimizar o erro quadrático médio $E(X-c)^2$, o melhor preditor é a média:

PROPOSIÇÃO 3.3. *Seja X integrável, $\mu=EX$. Então μ minimiza $E(X-c)^2$, $c\in\mathbb{R}$,* i. e.,

$$\text{Var } X=E(X-\mu)^2=\min_{c\in\mathbb{R}} E(X-c)^2.$$

Prova. $(X-c)^2=(X-\mu+\mu-c)^2=(X-\mu)^2+2(\mu-c)(X-\mu)+$ $+(\mu-c)^2$, logo (pela linearidade da esperança)

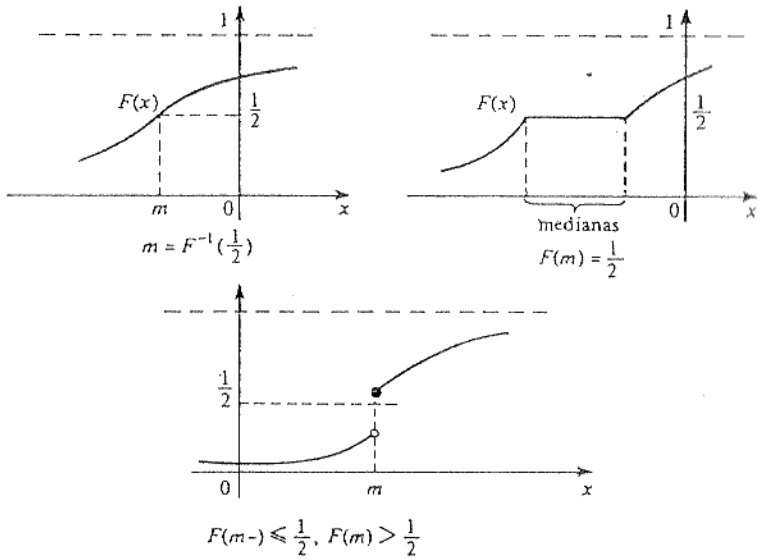
$$E(X-c)^2=E(X-\mu)^2+2(\mu-c)(EX-\mu)+$$
$$+(\mu-c)^2=\text{Var } X+(\mu-c)^2.$$

Conclusão: $E(X-c)^2\geq E(X-\mu)^2$, $\forall c\in\mathbb{R}$. □

PROPOSIÇÃO 3.4. *Seja X uma variável aleatória, e seja m uma mediana de X. Então m minimiza $E|X-c|$, $c\in\mathbb{R}$, i. e.,*

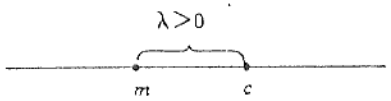
$$E|X-m|=\min_{c\in\mathbb{R}} E|X-c|.$$

Observação. Por definição, m é uma *mediana* de X se $P(X\geq m)\geq 1/2$ e $P(X\leq m)\geq 1/2$. Para obter uma mediana, basta considerar a função de distribuição de X. Por exemplo, consideremos as seguintes funções de distribuição F:



Prova da Proposição. Notemos que X é integrável se, e somente se, $\forall c\in\mathbb{R}$, $X-c$ também 0 é, pela linearidade da esperança. Portanto, se $E|X|=+\infty$, então $E|X-\mu|=+\infty$ $\forall c\in\mathbb{R}$ e a proposição vale trivialmente. Consideremos, então, o caso de X *integrável.*

Suponha $m < c$ (o caso $c < m$ é análogo); desejamos provar que $E|X-c| \geq E|X-m|$. Seja $\lambda = c - m$:



se $x \leq m$, $|x-c| = |x-m| + \lambda$; se $x > m$, $|x-c| \geq |x-m| - \lambda$. Então,

$$X \leq m \Rightarrow |X-c| - |X-m| = \lambda,$$
$$X > m \Rightarrow |X-c| - |X-m| \geq -\lambda.$$

Como m é mediana de X, temos $P(X \leq m) \geq \frac{1}{2}$ e $P(X > m) = 1 - P(X \leq m) \leq \frac{1}{2}$. Então a variável aleatória $Y = |X-c| - |X-m|$ tem esperança não-negativa, pois toma o valor $\lambda > 0$ com probabilidade $\geq \frac{1}{2}$, e com probabilidade $\leq \frac{1}{2}$ toma valores $\geq -\lambda$. A prova formal desse fato utiliza um argumento do tipo utilizado na prova alternativa da desigualdade "basica":

Como $Y \geq \lambda I_{[X \leq m]} - \lambda I_{[X > m]}$, temos

$$EY \geq \lambda E I_{[X \leq m]} - \lambda E I_{[X > m]} = \lambda P(X \leq m) - \lambda P(X > m) =$$
$$= \lambda(P(X \leq m) - P(X > m)) \geq \lambda\left(\frac{1}{2} - \frac{1}{2}\right) = 0.$$

Portanto, pela definição de Y e a linearidade da esperança, temos $E|X-c| \geq E|X-m|$. □

§3.6 Esperanças de funções de vetores aleatórios

TEOREMA 3.2. *Seja $\underset{\sim}{X} = (X_1, \ldots, X_n)$ um vetor aleatório e $\varphi: \mathbb{R}^n \to \mathbb{R}$ mensurável a Borel. Então*

$$E\varphi(\underset{\sim}{X}) \overset{def}{=} \int y dF_{\varphi(\underset{\sim}{X})}(y) = \int_{\mathbb{R}^n} \varphi dF_{\underset{\sim}{X}} =$$
$$= \int \cdots \int \varphi(x_1, \ldots, x_n) dF_{\underset{\sim}{X}}(x_1, \ldots, x_n),$$

onde a última integral é uma integral n-dimensional de Stieltjes, assim como a penúltima (notação compacta).

Prova. Teoria da Medida. □

Observações. (1) Você não precisa entender a integral de Stieltjes no $\mathbb{R}^n$ (a integração é feita em relação à medida de Lebesgue-Stieltjes gerada por $F_{\underset{\sim}{X}}$, ou seja, em relação à distribuição de $\underset{\sim}{X}$). Basta saber que a integral se simplifica nos casos discreto e contínuo, como no caso unidimensional:

Caso discreto. Se $\underset{\sim}{X}$ for discreto, tomando os valores $x_i = (x_{i_1}, \ldots, x_{i_n})$ com probabilidade $p(\underset{\sim}{x}_i)$, onde $\sum_i p(\underset{\sim}{x}_i) = 1$, então

$$E\varphi(\underset{\sim}{X}) = \sum_i \varphi(\underset{\sim}{x}_i)\, p(\underset{\sim}{x}_i).$$

Caso contínuo. Se $\underset{\sim}{X}$ for contínuo com densidade $f(x_1, \ldots, x_n)$, então

$$E\varphi(\underset{\sim}{X}) = \int \cdots \int \varphi(x_1, \ldots, x_n) f(x_1, \ldots, x_n) dx_1 \ldots dx_n.$$

(2) Podemos terminar agora a prova da propriedade E3, a linearidade da esperança. Resta provar que $E(X+Y) = EX + EY$, contanto que o termo à direita tenha sentido. Por isso, sejam $\varphi(x, y) = x + y$, $\varphi_1(x, y) = x$, $\varphi_2(x, y) = y$. Pelo teorema,

$$E(X+Y) = E\varphi(X, Y) = \iint (x+y) dF_{X,Y}(x, y).$$

Agora aplicamos a linearidade da integral múltipla de Stieltjes, obtendo

$$E(X+Y) = \iint x dF_{X,Y}(x, y) + \iint y dF_{X,Y}(x, y) =$$
$$= E\varphi_1(X, Y) + E\varphi_2(X, Y) = EX + EY.$$

(*Exercício.* Verifique a linearidade da esperança para combinações lineares de n variáveis: $E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i EX_i$.)

(3) No caso da independência das X_i, a integral n-dimensional se simplifica, tornando-se integral iterada (n iterações, cada uma sendo

uma integral de Stieltjes na reta). Consideremos primeiro o caso contínuo:

Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes, com $\underset{\sim}{X} = (X_1, \ldots, X_n)$, e suponha que $X_1, \ldots, X_n$ tenham (respectivamente) densidades $f_1, \ldots, f_n$. Então a densidade conjunta é o produto das densidades f_i, e

$$E\varphi(\underset{\sim}{X}) = \int \cdots \int \varphi(x_1, \ldots, x_n) f_1(x_1) \ldots f_n(x_n) dx_1 \ldots dx_n =$$

$$= \int \cdots \left[\int \varphi(x_1, \ldots, x_n) f_1(x_1) dx_1 \right] f_2(x_2)\, dx_2 \ldots f_n(x_n) dx_n.$$

Escrevendo $f_i(x_i)dx_i = dF_{X_i}(x_i)$, chegamos à seguinte fórmula que é válida para $X_1, \ldots, X_n$ independentes no *caso geral*:

$$E\varphi(\underset{\sim}{X}) = \int \cdots \int \varphi(x_1, \ldots, x_n) dF_{X_1}(x_1) \ldots dF_{X_n}(x_n).$$

(Não provaremos a fórmula no caso geral.)

Como conseqüência imediata temos que se as X_i são independentes, a esperança do produto é o produto das esperanças:

PROPOSIÇÃO 3.5. *Se $X_1, \ldots, X_n$ são variáveis aleatórias independentes e integráveis, então $\prod_{i=1}^{n} X_i$ é integrável e*

$$E(X_1 X_2 \ldots X_n) = \prod_{i=1}^{n} EX_i.$$

Prova. Basta provar para $n = 2$ (e completar com indução). Seja $\varphi(x, y) = xy$, então a independência de X e Y implica

$$EXY = E\varphi(X, Y) = \int\int \varphi(x, y) dF_X(x) dF_Y(y) =$$

$$= \int \left[\int x dF_X(x) \right] y dF_Y(y) = \int (EX) y dF_Y(y) = EX \cdot EY. \quad \square$$

Adverténcia. *$EXY = EX \cdot EY$ não implica X e Y independentes*, como vemos no seguinte exemplo. Sejam X e Y variáveis aleatórias tomando os valores -1, 0, 1, com distribuição conjunta definida por $p(-1, -1) = p(-1, 1) = p(1, -1) = p(1, 1) = p(0, 0) = \frac{1}{5}$, i. e., a função de probabilidade conjunta é a da seguinte tabela.

Y \ X	-1	0	1
-1	$1/5$	0	$1/5$
0	0	$1/5$	0
1	$1/5$	0	$1/5$

Então,

$$EX = (-1) \cdot \frac{2}{5} + 0 \cdot \frac{1}{5} + 1 \cdot \frac{2}{5} = 0 = EY, \text{ e}$$

$$EXY = \sum_{i,j} ij\, p(i, j) = \frac{1}{5}(1 - 1 - 1 + 1 + 0) = 0.$$

Portanto, $EXY = EX \cdot EY$. Mas X e Y não são independentes. Temos, por exemplo,

$$P(X = 0, Y = 0) = p(0, 0) = \frac{1}{5} \neq \frac{1}{25} = \frac{1}{5} \cdot \frac{1}{5} = P(X = 0) \cdot P(Y = 0).$$

A diferença entre os valores EXY e $EX \cdot EY$ será chamada *covariância* entre X e Y. Formalmente, sejam X e Y variáveis aleatórias integráveis. Então a covariância entre X e Y é definida por

$$\operatorname{Cov}(X, Y) = E[(X - EX)(Y - EY)],$$

se esta esperança existe. Por linearidade, temos

$$\operatorname{Cov}(X, Y) = E(XY - YEX - XEY + EX \cdot EY) = EXY - EX \cdot EY,$$

de modo que existe a covariância entre duas variáveis integráveis se, e somente se, existe a esperança EXY.

Se $\operatorname{Cov}(X, Y) = 0$, dizemos que X e Y são *não-correlacionadas*. Se X e Y são independentes e integráveis, então são não-correlacionadas, pois neste caso $EXY = EX \cdot EY$ pela Proposição 3.5. Mas acabamos de ver que a igualdade $EXY = EX \cdot EY$ não implica a independência, ou seja, covariância zero não necessariamente implica independência.

Observação. Há certos casos especiais em que não correlação implica independência. Talvez o mais importante seja o da normal: se X e Y possuem distribuição conjunta normal bivariada e são não-correlacionadas, então $\rho = 0$ (isto será visto no §4.5). E já vimos no exemplo 15, §2.5, que X e Y são independentes se $\rho = 0$.

Para um outro exemplo de independência como conseqüência de covariância zero, veja o exercício 26.

Vejamos agora que se as variáveis aleatórias $X_1, \ldots, X_n$ são não-correlacionadas (2 a 2), então a variância da soma é a soma das variâncias.

PROPOSIÇÃO 3.6. *Sejam $X_1, \ldots, X_n$ variáveis aleatórias integráveis tais que $Cov(X_i, X_j) = 0$ para $i \neq j$. Então*

$$Var(X_1 + \ldots + X_n) = \sum_{i=1}^{n} Var\, X_i.$$

Prova. $Var(X_1 + \ldots + X_n) = E(X_1 + \ldots + X_n - E(X_1 + \ldots + X_n))^2 =$
$= E((X_1 - EX_1) + \ldots + (X_n - EX_n))^2 =$
$= E\left[\sum_{i=1}^{n} (X_i - EX_i)^2 + 2\sum_{i<j} (X_i - EX_i)(X_j - EX_j)\right] =$
$= \sum_{i=1}^{n} \text{Var}\, X_i + 2\sum_{i<j} \text{Cov}(X_i, X_j) = \sum_{i=1}^{n} \text{Var}\, X_i.$ □

COROLÁRIO. *Se $X_1, \ldots, X_n$ são independentes e integráveis, então*

$$Var(X_1 + \ldots + X_n) = \sum_{i=1}^{n} Var\, X_i.$$

Observações. (1) Salientamos um resultado que aparece na prova acima: *se, $X_1, \ldots, X_n$ são integráveis, então*

$$Var(X_1 + \ldots + X_n) = \sum_{i=1}^{n} Var\, X_i + 2\sum_{i<j} Cov(X_i, X_j).$$

(2) Já vimos no §2.6 um exemplo da propriedade enunciada pela proposição: se $X_1, \ldots, X_n$ são independentes e normais, com $X_i \sim N(\mu_i, \sigma_i^2)$, então a soma é também normal, e para obter os parâmetros basta somar os parâmetros das parcelas: $\sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$. Seria bom você verificar agora, se ainda não o fez, que $X \sim N(\mu, \sigma^2) \Rightarrow$ $\Rightarrow EX = \mu$, Var $X = \sigma^2$.

Suponhamos agora que X e Y sejam variáveis aleatórias integráveis, com variâncias positivas e finitas ($0 < \sigma_X^2 < \infty$, $0 < \sigma_Y^2 < \infty$). A variável aleatória $\dfrac{X - EX}{\sigma_X}$ é uma *padronização* de X (também chamada *redução* ou *normalização* de X), pois expressa o valor de X em unidades padronizadas, i. e., desvios-padrão. Notemos que esta variável aleatória padronizada possui esperança zero e variância um. Além disso, não depende da escala nem da locação de X, no sentido de que $Z = aX + b$ possui a mesma padronização que X, se $a > 0$ e $b \in \mathbb{R}$. No mesmo estilo, a covariância entre as variáveis padronizadas também não depende da escala nem da locação de X e Y; é uma espécie de covariância padronizada. Chama-se *coeficiente de correlação* entre X e Y e indica-se com $\rho_{X,Y}$ ou $\rho(X, Y)$:

$$\rho_{X,Y} \overset{def}{=} \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = E\left[\left(\frac{X - EX}{\sigma_X}\right)\left(\frac{Y - EY}{\sigma_Y}\right)\right].$$

(*Exercício.* Verifique que $\rho(X, Y) = \rho(aX + b, cY + d)$ para $a > 0$, $c > 0$, i. e., o coeficiente de correlação é independente da escala e locação das variáveis.)

Podemos dizer que em certo sentido, $\rho_{X,Y}$ representa a dependência linear entre X e Y, como vemos pela seguinte proposição.

PROPOSIÇÃO 3.7. *Sejam X e Y variáveis aleatórias com variâncias finitas e positivas. Então:*

(a) $-1 \leq \rho(X, Y) \leq 1$.
(b) $\rho(X, Y) = 1$ *se, e somente se,* $P(Y = aX + b) = 1$ *para algum* $a > 0$, $b \in \mathbb{R}$.
(c) $\rho(X, Y) = -1 \Leftrightarrow P(Y = aX + b) = 1$ *para algum* $a < 0$, $b \in \mathbb{R}$.

Prova. (a) Como

$$0 \leq \left(\frac{X - EX}{\sigma_X} - \frac{Y - EY}{\sigma_Y}\right)^2,$$

temos

$$0 \leq E\left(\frac{X - EX}{\sigma_X} - \frac{Y - EY}{\sigma_Y}\right)^2 = E\left(\frac{X - EX}{\sigma_X}\right)^2 + E\left(\frac{Y - EY}{\sigma_Y}\right)^2 -$$
$$- \frac{2}{\sigma_X \sigma_Y} E[(X - EX)(Y - EY)] = \frac{\text{Var}\, X}{\sigma_X^2} + \frac{\text{Var}\, Y}{\sigma_Y^2} -$$
$$- \frac{2\,\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = 2 - 2\rho(X, Y).$$

Logo $\rho(X, Y) \leq 1$.

Substituindo o sinal "$-$" por "$+$" na expressão acima, temos $0 \leq 2 + 2\rho(X, Y)$, i. e., $\rho(X, Y) \geq -1$.

(b) e (c). Se $\rho(X, Y) = 1$, então (pela prova de (a))

$$E\left(\frac{X - EX}{\sigma_X} - \frac{Y - EY}{\sigma_Y}\right)^2 = 0,$$

i. e., $P\left(\frac{X - EX}{\sigma_X} = \frac{Y - EY}{\sigma_Y}\right) = 1$ (conseqüência (c) da desigualdade "básica" – propriedade E7). Em outras palavras, $Y = \frac{\sigma_Y}{\sigma_X}(X - EX)$ quase certamente, o que prova a necessidade em (b), com $a = \frac{\sigma_Y}{\sigma_X}$ e $b = EY - \frac{\sigma_Y}{\sigma_X} EX$.

Se $\rho(X, Y) = -1$, então $E\left(\frac{X - EX}{\sigma_X} + \frac{Y - EY}{\sigma_Y}\right)^2 = 0$ e $Y = EY - \frac{\sigma_Y}{\sigma_X}(X - EX)$ com probabilidade 1. Neste caso, $a = -\frac{\sigma_Y}{\sigma_X}$ e $b = EY + \frac{\sigma_Y}{\sigma_X} EX$.

Por outro lado, se $P(Y = aX + b) = 1$ para algum $a \neq 0$, temos

$$\rho(X, Y) = E\left[\left(\frac{X - EX}{\sigma_X}\right)\left(\frac{aX + b - aEX - b}{\sqrt{a^2\sigma_X^2}}\right)\right] =$$

$$= \frac{a}{|a|} E\left(\frac{X - EX}{\sigma_X}\right)^2 = \frac{a}{|a|} = \text{sinal } (a) = \pm 1. \ \square$$

Uma alta correlação entre X e Y, i. e., $\rho_{X,Y}$ próximo de 1, significa que o valor de Y tende a acompanhar o de X (quanto maior X, maior também é, geralmente, Y). Por outro lado, uma correlação negativa forte ($\rho_{X,Y}$ próximo de -1) significa a tendência oposta, ou seja, quanto maior X, menor Y, e vice-versa.

Já vimos, nesta seção, um exemplo de um par de variáveis aleatórias dependentes, mas que tinham covariância zero e, portanto, coeficiente de correlação zero. Nesse exemplo, em que a distribuição conjunta estava concentrada nos cinco pontos $(-1, -1)$, $(-1, 1)$, $(0, 0)$, $(1, -1)$, $(1, 1)$, cada um tendo probabilidade $\frac{1}{5}$, vemos que há uma simetria da distribuição conjunta em relação aos dois eixos. Essa simetria garante a falta de correlação sem ao mesmo tempo garantir a independência. É claro que o valor de Y não possui nem a tendência a acompanhar o de X, nem a tendência oposta.

Observação. Quando X e Y têm distribuição normal bivariada, ρ é o coeficiente de correlação, como será mostrado no §4.5.

§3.7 Teoremas de convergência

As provas desta seção podem ser omitidas em uma primeira leitura. Mas os dois teoremas – da Convergência Monótona e Dominada – são de grande importância e utilidade na Probabilidade, Teoria da Medida, e Matemática em geral. (Para exemplos da utilização de métodos probabilísticos na demonstração de resultados de Análise, veja as conseqüências 2 e 3 a seguir.)

Os teoremas serão dados do ponto de vista probabilístico. Para tanto, sejam $X, X_1, X_2, \ldots$ variáveis aleatórias definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$, e suponha que X_n convirja para X quando $n \to \infty$. A convergência aqui é pontual (recorde que variáveis aleatórias são funções definidas no espaço amostral), i.e., $X_n(\omega) \to X(\omega)$ para todo $\omega \in \Omega$. (Estudaremos outros tipos de convergência nos Capítulos 5 e 6.) A questão é a seguinte: sob que condições a *esperança do limite é o limite das esperanças*? Isto é, queremos saber quando $EX = \lim_{n \to \infty} EX_n$. Não é verdade, infelizmente, que $EX_n \to EX$ sempre. Um contra-exemplo simples é o seguinte: suponha que $X \sim$ Cauchy-padrão e seja

$$X_n = XI_{[-n \leq X \leq n]} = \begin{cases} X & \text{se } -n \leq X \leq n \\ 0 & \text{se } |X| > n. \end{cases}$$

X_n é uma variável aleatória de Cauchy *truncada*.

Então $X_n \to X$, pois $X_n(\omega) = X(\omega)$ para $n \geq |X(\omega)|$. X_n é integrável, porque limitada, e $EX_n = 0$, por simetria (veja o exercício 1). Mas EX_n não converge para EX, porque EX não existe. (Para um outro exemplo, em que $X_n \to X$ e EX existe, mas $EX_n \not\to EX$, veja o exercício 37.)

Daremos duas respostas à questão colocada acima: a esperança do limite é o limite das esperanças (a) quando as variáveis são não-negativas e a seqüência cresce monotonamente e (b) quando a se-

qüência é uniformemente limitada (i.e., dominada) por uma variável aleatória integrável.

TEOREMA 3.3. *Teorema da Convergência Monótona.*

Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias em $(\Omega, \mathscr{A}, P)$. Se $0 \le X_n \uparrow X$, i.e., $X_n(\omega) \ge 0$ e $X_n(\omega) \uparrow X(\omega)$ para todo $\omega \in \Omega$, então $EX_n \uparrow EX$.

Prova. Pela propriedade E2, temos $0 \le EX_n \le EX$ e $EX_n \uparrow$. Logo $\lim_{n\to\infty} EX_n \le EX$ e basta provar que $\lim EX_n \ge EX - \varepsilon$ para todo $\varepsilon > 0$. Para isso, vamos aproximar X por meio de uma variável aleatória discreta Y tal que $|X - Y| \le \varepsilon$, onde $\varepsilon > 0$ é fixo.

Definamos o evento $B_n = [n\varepsilon < X \le (n+1)\varepsilon]$, $n = 0, 1, 2, \ldots$, e a variável aleatória $Y = \sum_{n=0}^{\infty} n\varepsilon I_{B_n}$. Em outras palavras,

$$Y(\omega) = \begin{cases} n\varepsilon \text{ se } n\varepsilon < X(\omega) \le (n+1)\varepsilon \\ 0 \text{ se } X(\omega) = 0. \end{cases}$$

Logo $X - \varepsilon \le Y \le X$ e $EX - \varepsilon \le EY \le EX$ (vale também caso $EX = +\infty$).

Vamos provar agora que $EY \le \lim EX_n$, i.e., $\lim EX_n \ge EX - \varepsilon$, assim terminando a prova. Para tanto, seja

$$A_k = [X_k \ge Y].$$

Observemos que $A_k \uparrow \Omega$ (pois $X_k(\omega) \ge Y(\omega) \Rightarrow X_{k+1}(\omega) \ge Y(\omega)$, pela monotonia de X_k, portanto $A_k \uparrow$. Mas a convergência de X_k para X implica que $X_k(\omega) \ge Y(\omega)$ para k suficientemente grande: notemos que $Y(\omega) < X(\omega)$ a menos que $X(\omega) = 0$. Logo $\Omega = \cup A_k = \lim A_k$.) Portanto,

$$B_n \cap A_k \uparrow B_n \text{ quando } k \to \infty \ (n \text{ fixo}).$$

Ora, a variável aleatória YI_{A_k} é discreta e

$$Y(\omega)\, I_{A_k}(\omega) = \begin{cases} Y(\omega) \le X_k(\omega), \text{ se } \omega \in A_k \\ 0 \le X_k(\omega), \text{ se } \omega \notin A_k, \end{cases}$$

logo $0 \le YI_{A_k} \le X_k$ e $0 \le EYI_{A_k} \le EX_k$. Para calcular EYI_{A_k}, é preciso notar que

$$Y(\omega) I_{A_k}(\omega) = \begin{cases} n\varepsilon \text{ se } \omega \in B_n \cap A_k, \ n = 0, 1, 2, \ldots \\ 0 \text{ se } \omega \notin \bigcup_{n=0}^{\infty} (B_n \cap A_k). \end{cases}$$

Portanto,

$$EX_k \ge EYI_{A_k} = \sum_{n=0}^{\infty} n\varepsilon\, P(B_n \cap A_k) \ge \sum_{n=0}^{m} n\varepsilon\, P(B_n \cap A_k), \ \forall m.$$

Mas $P(B_n \cap A_k) \uparrow P(B_n)$ quando $k \to +\infty$, logo

$$\lim_{k\to\infty} EX_k \ge \lim_{k\to\infty} \sum_{n=0}^{m} n\varepsilon\, P(B_n \cap A_k) = \sum_{n=0}^{m} n\varepsilon\, P(B_n), \ \forall m.$$

Portanto,

$$\lim_{k\to\infty} EX_k \ge \sum_{n=0}^{\infty} n\varepsilon\, P(B_n) = EY. \ \square$$

TEOREMA 3.4. *Teorema da Convergência Dominada.*

Sejam $Y, X, X_1, X_2, \ldots$ variáveis aleatórias em $(\Omega, \mathscr{A}, P)$ tais que Y é integrável, $|X_n| \le Y \ \forall n$, e $X_n \to X$ (i.e., $X_n(\omega) \to X(\omega) \ \forall \omega$). Então X e X_n são integráveis e $EX_n \to EX$.

Prova. Como X_n e X são dominadas por Y ($|X| = \lim |X_n| \le Y$), a integrabilidade delas é conseqüência da propriedade E2.

Faça $Y_n = \inf_{k \ge n} X_k$, então $Y_n \uparrow X$ quando $n \to \infty$ (pois $X(\omega) = \lim X_n(\omega) = \liminf_{n\to\infty} X_n(\omega) = \lim_{n\to\infty} (\inf_{k\ge n} X_k(\omega)) = \lim_{n\to\infty} Y_n$, sendo óbvio que Y_n cresce com n). Logo, temos $(Y_n + Y) \uparrow (X + Y)$ quando $n \to \infty$.

Mas $X_n \ge -Y \ \forall n \Rightarrow Y_n \ge -Y \Rightarrow Y_n + Y \ge 0$, e o Teorema da Convergência Monótona implica que $E(Y_n + Y) \uparrow E(X + Y)$. Por linearidade e integrabilidade, temos

$$EY_n \uparrow EX. \tag{3.6}$$

(*Observação*: Y_n é variável aleatória, pois $[Y_n < a] = \bigcap_{k\ge n} [X_k < a]$.)

De modo análogo, se $Z_n(\omega) \overset{def}{=} \sup_{k\ge n} X_k(\omega)$ então $Z_n(\omega) \downarrow X(\omega)$ quando $n \to \infty$, i.e., $(Y - Z_n) \uparrow (Y - X)$.

Mas $X_n \le Y \ \forall n \Rightarrow Z_n \le Y \Rightarrow Y - Z_n \ge 0$, e pelo Teorema da Convergência Monótona, $E(Y - Z_n) \uparrow E(Y - X)$, de modo que

$$EZ_n \downarrow EX. \tag{3.7}$$

Já que

$$Y_n = \inf_{k\ge n} X_k \le X_n \le \sup_{k\ge n} X_k = Z_n,$$

temos $EY_n \le EX_n \le EZ_n$, o que, combinado com (3.6) e (3.7), implica $EX_n \to EX$. □

Conseqüências dos teoremas de convergência.

(1) Seja X uma variável aleatória. Se $E|X|^t < \infty$ para algum $t > 0$, então a função g definida por $g(s) = E|X|^s$ é contínua no intervalo $(0, t]$.

Prova. Suponha $s_n \to s$, onde $s,\ s_n \in (0, t]$. Então $|X|^{s_n} \to |X|^s$, e para aplicar o Teorema da Convergência Dominada basta verificar se as variáveis $|X|^{s_n}$ são dominadas por uma variável aleatória integrável. Mas $|X|^{s_n} \le |X|^t + 1$ e $E(|X|^t + 1) = E|X|^t + 1 < \infty$. Logo $E|X|^{s_n} \to E|X|^s$. Portanto, $g(s_n) \to g(s)$ para toda seqüência $(s_n)_{n \ge 1}$ que converge para s. Por isso, g é contínua em $(0, t]$. □

(2) *Teorema de Arzelà* (Veja Apostol [2], Teorema 13-17.)

Sejam $f, f_1, f_2, \ldots$ funções reais (mensuráveis a Borel) definidas no intervalo $[a, b]$, $a < b$, e integráveis a Riemann. Se $f_n \to f$ em toda parte, e se $|f_n| \le M < \infty$ para todo n (i.e., as f_n são uniformemente limitadas), então

$$\int_a^b f_n(x)\,dx \to \int_a^b f(x)\,dx \text{ quando } n \to \infty.$$

Prova. (O teorema é verdadeiro sem a hipótese de mensurabilidade a Borel, mas a prova é mais difícil.) Sejam $\Omega = [a, b]$, $\mathscr{A} = \mathscr{B}_{[a,b]}$ = borelianos de $[a, b]$, e P = probabilidade uniforme em $[a, b]$, i.e., $P(A) = \dfrac{\text{comprimento } (A)}{b - a}$, $A \in \mathscr{A}$.

Definamos variáveis aleatórias X_n e X:

$$X_n(\omega) = (b - a) f_n(\omega);\ X(\omega) = (b - a) f(\omega)$$

(são variáveis aleatórias porque são mensuráveis). Então $X_n(\omega) \to X(\omega)$, $\forall \omega \in \Omega$, e as X_n são dominadas por $(b - a)M$, que é integrável. Pelo Teorema da Convergência Dominada, $EX_n \to EX$. Mas $EX_n = \int_a^b f_n(x)\,dx$ e $EX = \int_a^b f(x)\,dx$, logo

$$\int_a^b f_n(x)\,dx \to \int_a^b f(x)\,dx.$$

(Para ver que $EX_n = \int_a^b f_n(x)\,dx$, seja Z a variável aleatória identidade: $Z(\omega) = \omega$. Então $Z \sim U[a, b]$ e $f_Z(x) = \dfrac{1}{b - a} I_{[a,b]}(x)$. Como $X_n(\omega) = X_n(Z(\omega)) = (b - a) f_n(Z(\omega))$, temos

$$EX_n = (b - a)\,Ef_n(Z) = (b - a) \int f_n(z) f_Z(z)\,dz = \int_a^b f_n(z)\,dz.)\ \square$$

(3) *Convergência de séries.* Se $a_{mn} \ge 0$ para $m,\ n = 1, 2, 3, \ldots$, e se $a_{mn} \uparrow a_m$ quando $n \to \infty$, para todo m, então

$$\sum_{m=1}^{\infty} a_{mn} \uparrow \sum_{m=1}^{\infty} a_m.$$

Prova. Escolhamos $p_m > 0$ tal que $\sum_{m=1}^{\infty} p_m = 1$. Definamos $\Omega = \{1, 2, \ldots\}$, $p(m) = p_m$, $X_n(m) = \dfrac{a_{mn}}{p_m}$, $X(m) = \dfrac{a_m}{p_m}$. Então

$$0 \le X_n(m) \underset{n \to \infty}{\uparrow} X(m)\ \forall m,$$

logo o Teorema da Convergência Monótona implica $EX_n \uparrow EX$. Mas $EX_n = \sum_{m=1}^{\infty} X_n(m) p_m = \sum_{m=1}^{\infty} a_{mn}$ e $EX = \sum_{m=1}^{\infty} X(m) p_m = \sum_{m=1}^{\infty} a_m$. □

EXERCÍCIOS DO CAPÍTULO 3

§ 3.2

1. Dizemos que a distribuição de X é *simétrica* (em torno de zero) se $P(X \ge x) = P(X \le -x)$ para todo $x \in \mathbb{R}$, e que sua distribuição é *simétrica em torno de* μ se

$$P(X \ge \mu + x) = P(X \le \mu - x)\ \forall x \in \mathbb{R}.$$

(a) Prove: se a distribuição de X é simétrica em torno de μ e se X é integrável, então $EX = \mu$. (*Sugestão.* Prove primeiro para $\mu = 0$.)

(b) Suponha que X possua densidade $f(x)$. Mostre que se $f(\mu + x) = f(\mu - x)\ \forall x$, então a distribuição de X é simétrica em torno de μ. Enuncie e prove a propriedade análoga para o caso discreto.

(c) Qual o ponto de simetria das distribuições das seguintes variáveis aleatórias? Dê a esperança de cada variável aleatória, se existir.

(i) $X \sim N(\mu, \sigma^2)$.

(ii) $X \sim$ Cauchy (M, b), i. e., $f(x) = \dfrac{b}{\pi[b^2 + (x - M)^2]}$, $x \in \mathbb{R}$.

(iii) $X \sim U[a, b]$.

(iv) $X \sim b\left(n, \dfrac{1}{2}\right)$.

(v) X tal que F_X é a função de Cantor.

(vi) X tendo *distribuição de Laplace* (ou *exponencial dupla*):

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x-\mu|}, \quad x \in \mathbb{R}.$$

2. Seja X uma variável aleatória tendo *distribuição logística* com densidade

$$f(x) = \frac{e^{-x}}{(1 + e^{-x})^2}, \quad x \in \mathbb{R}.$$

(a) Prove que a distribuição de X é simétrica em torno de zero (veja o exercício 1).

(b) Determine se X tem esperança finita. Se tiver, ache o valor.

(c) Obtenha a densidade de $Y = e^X$ e ache EY.

3. Calcule EX se X possui densidade $f(x)$, onde:

(a) f é a densidade dada no exercício 6 do Capítulo 2.

(b) $f(x) = \dfrac{1}{(1 + x)^2}$, se $x > 0$; $f(x) = 0$, se $x \leq 0$.

4. (a) Se $X \sim \Gamma(\alpha, \beta)$, onde $\alpha > 0$ e $\beta > 0$, qual é EX?

(b) Dizemos que X tem distribuição de Weibull se possui densidade

$$f(x) = \begin{cases} \lambda\alpha x^{\alpha-1} e^{-\lambda x^\alpha}, & x > 0 \\ 0, & x \leq 0. \end{cases}$$

Calcule EX neste caso.

5. Sejam X e Y variáveis aletatórias independentes com distribuição comum $U[0, 1]$. Calcule EZ e EW, onde $Z = \min(X, Y)$ e $W = \max(X, Y)$.

6. Um jogador vai lançando uma moeda honesta. Ele pára depois de lançar ou duas caras sucessivas ou duas coroas sucessivas. Qual a esperança do número de lançamentos?

7. Uma urna contém n bolas numeradas 1, 2, ..., n. Uma pessoa tira uma bola e a devolve, tira outra e a devolve, continuando até tirar uma bola pela segunda vez. Seja X o número total de retiradas necessárias para obter essa repetição.

(a) Ache a distribuição de X. (*Sugestão.* Ache $P(X > k)$.)

(b) Mostre que

$$EX = 2 + \left(1 - \frac{1}{n}\right) + \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) + \ldots + \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\ldots\left(1 - \frac{n-1}{n}\right).$$

8. Sejam X e Y variáveis aleatórias. Se $F_X(x) \leq F_Y(x)$ para todo $x \in \mathbb{R}$, dizemos que X é *estocasticamente maior* que Y. Prove que se X é estocasticamente maior que Y, então $EX \geq EY$ (se EX e EY existem).

9. Dois componentes eletrônicos vão ser testados simultaneamente. Suponha que a "vida" em horas de cada componente é exponencialmente distribuída com parâmetro λ, e que as vidas dos componentes são independentes. Calcule:

(a) A esperança do tempo até a primeira falha de um dos componentes.

(b) A esperança do tempo até ambos os componentes falharem.

10. Dois jogadores lançam moedas simultaneamente até obterem o primeiro casamento (i.e., ou duas caras ou duas coroas). Se os dois lançam "cara" simultaneamente, ganha o jogador I; se ambos lançam "coroa", ganha o jogador II. Por exemplo, se os dois obtêm "cara" no primeiro lançamento, então o jogo termina e o jogador I ganha o jogo. Suponha que a moeda do jogador I seja honesta, mas que a moeda do outro não necessariamente o seja, tendo probabilidade p de "cara", $0 < p < 1$.

(a) Calcule a esperança do número de lançamentos (i.e, o número de vezes que o jogador I lança a moeda até terminar o jogo).

(b) Ache a probabilidade do jogador I ganhar o jogo (mais cedo ou mais tarde).

11. Jogadores I e II têm Cr$ 200 cada um. Lança-se uma moeda com probabilidade p de dar cara ($0 < p < 1$). Se der cara, o jogador I recebe Cr$ 100 do II; se der coroa, I paga Cr$ 100 ao II. Con-

tinua-se lançando a moeda, independentemente, até um dos jogadores perder tudo, i.e., até um deles ficar com os Cr$ 400. Determine EN, onde N é o número de lançamentos até terminar o jogo.

§3.3

12. Prove que o critério para integrabilidade ainda vale se "$\geq$" é substituído por "$>$", i.e., X é integrável se, e somente se, $\sum_{n=1}^{\infty} P(|X| > n) < \infty$.

13. Verifique que a desigualdade de Jensen ainda vale se a função φ é convexa em um intervalo (a, b) tal que $P(a < X < b) = 1$, onde admitimos a possibilidade de $a = -\infty$ ou $b = +\infty$. (*Sugestão.* Para provar que $a < EX < b$ – note que a propriedade E3 só implica $a \leq EX \leq b$ – use o seguinte resultado do § 3.5: se $Z \geq 0$ e $EZ = 0$, então $P(Z = 0) = 1$.)

14. Sejam X e Y variáveis aleatórias com densidade conjunta $f(x, y)$. Verifique se $E(aX + bY) = aEX + bEY$ para este caso. (Suponha EX e EY finitas.)

§3.4

15. Suponha que $X \sim U[0, 1]$. Determine os valores de $t (t \in \mathbb{R})$ tais que $E(X^t)$ é finita. Qual o valor da esperança nesse caso?

16. Calcule $E(e^X)$, onde X tem densidade logística $f(x) = \dfrac{e^{-x}}{(1 + e^{-x})^2}$, $x \in \mathbb{R}$, utilizando o Teorema 3.1. Compare com o resultado obtido no exercício 2(*c*).

17. Calcule EY nos exercícios 6 e 7 do Capítulo 2.

18. Seja X o tempo de espera até o primeiro sucesso em uma seqüência de ensaios de Bernoulli tendo probabilidade p de sucesso em cada ensaio. Calcule EX^2.

§3.5

19. Suponha que a variável aleatória X tenha a seguinte densidade "triangular":

$$f(x) = \begin{cases} 1 + x \text{ se } -1 \leq x \leq 0 \\ 1 - x \text{ se } 0 < x \leq 1 \\ 0 \quad \text{se } x < -1 \text{ ou } x > 1. \end{cases}$$

Calcule EX e Var X.

20. (a) Prove: se a variável aleatória X é limitada, então tem momentos finitos de toda ordem.
    (b) Seja A um evento aleatório. Calcule todos os momentos absolutos da variável aleatória $X = I_A$.
    (c) Demonstre: se $X \sim N(\mu, \sigma^2)$, então todos os momentos absolutos de X são finitos.
    (d) Seja $X \sim$ Cauchy (0, 1). Quais são os momentos absolutos finitos de X?

21. Obtenha as variâncias de Z e W no exercício 5.

22. Através de experimentos estatísticos, determina-se que a duração de um certo tipo de chamada telefônica satisfaz a relação $P(T > t) = ae^{-\lambda t} + (1 - a)e^{-\xi t}$, $t \geq 0$, onde $0 \leq a \leq 1$, $\lambda > 0$, $\xi > 0$. Ache a média e a variância de T.

23. Prove que se X assume valores somente no intervalo $[a, b]$, então $a \leq EX \leq b$ e $\text{Var } X \leq \dfrac{(b-a)^2}{4}$. (*Sugestão.* Faça primeiro para $a = 0$, $b = 1$). Exiba uma variável aleatória que atinge a variância máxima.

24. Calcule a variância da variável aleatória X, sob as seguintes condições:
    (a) $X \sim$ Poisson (λ), onde $\lambda > 0$.
    (b) $X \sim b(n, p)$, onde $0 < p < 1$.
    (c) $X \sim N(\mu, \sigma^2)$, onde $\mu \in \mathbb{R}$ e $\sigma^2 > 0$.
    (d) $X \sim \Gamma(\alpha, \beta)$, onde $\alpha > 0$ e $\beta > 0$.
    (e) $X \sim U[a, b]$, onde $a < b$.

25. Demonstre que a desigualdade de Jensen é estrita, i.e., $E\varphi(X) > \varphi(EX)$, se a função φ é estritamente convexa e X não é constante. (*Sugestão.* Reveja a prova de E4 e use a conseqüência (*c*) de E7. Observe que, pela convexidade estrita, as curvas φ e L da prova de E4 têm apenas um ponto em comum.)

§3.6

26. (a) Sejam X e Y variáveis aleatórias que só assumem os valores 0 e 1. Mostre que se $EXY = EX \cdot EY$, então X e Y são independentes.
(b) Prove: se X assume apenas os valores a e b, Y assume apenas os valores c e d, e $\text{Cov}(X, Y) = 0$, então X e Y são independentes.

27. Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes com $EX_j = \mu_j$ e $\text{Var } X_j = \sigma_j^2$. Considere combinações lineares $Y = \sum_{j=1}^{n} p_j X_j$, onde $p_j \geq 0$ e $\sum_{j=1}^{n} p_j = 1$. Prove que $\text{Var} Y$ é minimizada pela escolha de

$$p_j = \frac{\sigma_j^{-2}}{\sum_{i=1}^{n} \sigma_i^{-2}}, \quad j = 1, \ldots, n.$$

28. Se X e Y são variáveis aleatórias independentes com variâncias finitas, demonstre que

$$\text{Var}(XY) = \text{Var}(X)\ \text{Var}(Y) + (EX)^2\ \text{Var}(Y) + (EY)^2 \text{Var}(X).$$

29. Sejam X e Y variáveis aleatórias com variâncias finitas. Mostre que se $\text{Var } X \neq \text{Var } Y$, então $X + Y$ e $X - Y$ não são independentes.

30. Seja X uma variável aleatória tendo distribuição $b(n, p)$. Mostre que X tem a mesma distribuição que $X_1 + \ldots + X_n$, onde as X_i são variáveis aleatórias independentes e identicamente distribuídas que assumem apenas os valores 0 e 1. (Qual é $P(X_i = 1)$?) Utilize esse resultado para calcular a esperança e a variância de X.

31. Demonstre que a covariância é *bilinear*:

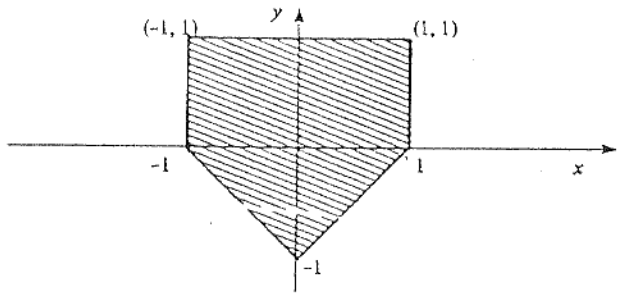
$$\text{Cov}\left(\sum_{i=1}^{m} a_i X_i,\ \sum_{j=1}^{n} b_j Y_j\right) = \sum_{j=1}^{n} \sum_{i=1}^{m} a_i b_j \text{Cov}(X_i, X_j),$$

onde os a_i e b_j são números reais. (Suponha que as X_i e Y_j possuam variâncias finitas.)

32. Seja $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ um vetor aleatório $(m+n)$ – dimensional tal que $\text{Var}(X_i) = \text{Var}(Y_j) = 1$, $\rho(X_i, X_j) = \rho_1$ e $\rho(Y_i, Y_j) = \rho_2$ $\forall i \neq j$, e $\rho(X_i, Y_j) = \rho_3\ \forall i, j$. Se $U = X_1 + \ldots X_m$ e $V = Y_1 + \ldots + Y_n$, então prove que o coeficiente de correlação entre U e V, $\rho(U, V)$, é igual a

$$\frac{\sqrt{mn}\rho_3}{\sqrt{1 + (m-1)\rho_1}\ \sqrt{1 + (n-1)\rho_2}}.$$

33. Seja (X, Y) uniformemente distribuído na seguinte região:



Calcule $\text{Cov}(X, Y)$.

34. Sejam X e Y variáveis aleatórias independentes com distribuição comum $U[0, 1]$, e sejam $X_{(1)} = \min(X, Y)$, $X_{(2)} = \max(X, Y)$. Calcule o coeficiente de correlação $\rho(X_{(1)}, X_{(2)})$.

35. Sejam X, Y e Z independentes com distribuição comum $U[0, 1]$. Calcule a esperança e a variância de $W = (X + Y) \cdot Z$.

36. Seja ρ o coeficiente de correlação entre X e Y. Determine $\rho(Z, W)$ em função de ρ, se $Z = aX + b$ e $W = cY + d$, onde $a \neq 0$, $c \neq 0$.

§3.7

37. Exiba um exemplo de uma seqüência tal que $X_n(\omega) \to X(\omega)\ \forall \omega \in \Omega$, com EX e EX_n finitas, mas $EX_n \not\to EX$. (*Sugestão.* Seja $Y \sim U[0, 1]$ e defina $X_n = nI_{[0 < Y < 1/n]}$.)

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE CIENCIAS

CAPÍTULO IV

DISTRIBUIÇÃO E ESPERANÇA CONDICIONAIS

§4.1 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

$$P(X \in B \mid A) = \frac{P([X \in B] \cap A)}{P(A)},$$

para $B \in \mathcal{B}$, a σ-álgebra dos borelianos na reta. Isto realmente define uma distribuição na reta, i.e., uma probabilidade nos borelianos, pois verificam-se os axiomas:

Axioma 1. $P(X \in B \mid A) \geq 0$ é quociente de probabilidades).

Axioma 2. $P(X \in \mathbb{R} \mid A) = 1$ (pois $P(X \in \mathbb{R} \mid A) = \frac{P(\Omega \cap A)}{P(A)} = 1$).

Axioma 3'. Se $B_1, B_2, \ldots$ são borelianos disjuntos 2 a 2, então $P(X \in \bigcup_n B_n \mid A) = \sum_n P(X \in B_n \mid A)$. (Pois

$$P(X \in \bigcup_n B_n \mid A) = \frac{P([X \in \cup\, B_n] \cap A)}{P(A)},$$

com

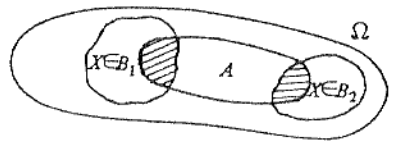
$$[X \in \bigcup_n B_n] \cap A = (\bigcup_n [X \in B_n]) \cap A = \bigcup_n ([X \in B_n] \cap A).$$

Como os B_n são disjuntos 2 a 2, os eventos $[X \in B_n] \cap A$ também o são, logo

$$P(X \in \cup\, B_n \mid A) = \frac{\sum_n P([X \in B_n] \cap A)}{P(A)} = \sum_n \frac{P([X \in B_n] \cap A)}{P(A)} =$$
$$= \sum_n P(X \in B_n \mid A).)$$

Portanto, $P(X \in B \mid A)$, $B \in \mathcal{B}$, define uma probabilidade em $\mathcal{B}$.

Podemos interpretar a distribuição condicional de X dado A como a nova distribuição que se atribui a X quando se sabe da ocorrência do evento A. Pensemos nas probabilidades dos eventos no seguinte diagrama como proporcionais às suas áreas:



Neste caso, $P(X \in B_1 \mid A)$, a probabilidade condicional de X pertencer a B_1 dado A, é a proporção do evento A em que $X \in B_1$, com uma interpretação análoga para B_2.

A função de distribuição associada à distribuição condicional é chamada *função de distribuição condicional de X dado A*:

$$F_X(x \mid A) = P(X \leq x \mid A) = \frac{P([X \leq x] \cap A)}{P(A)}, \quad x \in \mathbb{R}.$$

A *esperança condicional de X dado A* é a esperança da distribuição condicional, definida por

$$E(X \mid A) = \int x dF_X(x \mid A),$$

se esta esperança existe.

Agora suponhamos que os eventos aleatórios $A_1, A_2, \ldots$ formem uma partição (finita ou enumerável) de Ω i.e., que os A_n sejam disjuntos 2 a 2 e $\bigcup_n A_n = \Omega$. Pelo Teorema da Probabilidade Total, temos

$$P(X \in B) = \sum_n P(A_n) P(X \in B \mid A_n), \quad \forall B \in \mathcal{B},$$

$$F_X(x) = P(X \leq x) = \sum_n P(A_n)\, P(X \leq x \mid A_n) = \sum_n P(A_n)\, F_X(x \mid A_n), \quad \forall x,$$

e se a esperança de X existe,

$$EX = \int x\, dF_X(x) = \int x\, d(\sum_n P(A_n) F_X(x \mid A_n)) =$$
$$= \text{(pela linearidade da integral de Stieltjes)} =$$
$$= \sum_n \{P(A_n) \int x dF_X(x \mid A_n)\} = \sum_n P(A_n)\, E(X \mid A_n).$$

Em outras palavras, a distribuição de X (função de distribuição de X, esperança de X) é uma média ponderada da distribuição condicional (função de distribuição condicional, esperança condicional) dado A_n, onde os pesos são as probabilidades dos membros A_n da partição.

EXEMPLO 1. Um exemplo simples é o seguinte: seja $X \sim U[-1, 1]$, e sejam $A_1 = [X \geq 0]$, $A_2 = [X < 0] = A_1^c$. Qual a distribuição condicional de X dado A_1? Em outras palavras, dado que $X \geq 0$ qual a nova distribuição que se atribui a X? A resposta intuitiva é: uniforme em $[0, 1]$. Verifiquemos a resposta: já que para todo boreliano B, $P(X \in B) = \frac{1}{2} \times$ comprimento $(B \cap [-1, 1])$, temos

$$P(X \in B \mid A_1) = \frac{P([X \in B] \cap [X \geq 0])}{P(X \geq 0)} = \frac{P(X \in B \cap [0, \infty))}{P(X \geq 0)} =$$

$$= \frac{\frac{1}{2} \text{ comprimento } (B \cap [0, 1])}{\frac{1}{2}} =$$

$$= \text{comprimento } (B \cap [0, 1]),$$

satisfazendo assim a definição da distribuição $U[0, 1]$. Vamos usar a notação $X \mid A_1 \sim U[0, 1]$ para indicar o fato da distribuição condicional de X dado A_1 ser uniforme em $[0, 1]$. Podemos mostrar por um método análogo que $X \mid A_2 \sim U[-1, 0]$.

Logo temos $E(X \mid A_1) = \frac{1}{2}$, $E(X \mid A_2) = -\frac{1}{2}$ e, verificando a fórmula que expressa a esperança como média ponderada da esperança condicional,

$$0 = EX = \frac{1}{2} \cdot P(A_1) - \frac{1}{2} P(A_2) = \frac{1}{4} - \frac{1}{4} = 0.$$

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 $y_1, y_2, \ldots$, onde admitimos que esta seqüência dos valores possíveis de Y seja finita ou enumerável. (Ressaltamos que estamos supondo que X e Y sejam definidas no mesmo espaço de probabilidade.) Então

os eventos $A_n = [Y = y_n]$ formam uma partição de Ω. Neste caso, a distribuição

$$P(X \in B \mid Y = y_n) = P(X \in B \mid A_n), \quad B \in \mathcal{B},$$

é chamada *distribuição condicional de X dado que $Y = y_n$*, e valem as fórmulas

$$P(X \in B) = \sum_n P(Y = y_n)\, P(X \in B \mid Y = y_n), \quad B \in \mathcal{B},$$

$$F_X(x) = \sum_n P(Y = y_n)\, F_X(x \mid Y = y_n), \quad x \in \mathbb{R},$$

$$EX = \sum_n P(Y = y_n)\, E(X \mid Y = y_n),$$

onde vale a última fórmula se EX existe; em particular, se X é integrável.

Notemos que para B fixo, $P(X \in B \mid Y = y_n)$ é função de y_n, digamos $g(y_n)$. Se definirmos $g(y) = P(X \in B \mid Y = y)$ arbitrariamente para $y \in \{y_n : n \geq 1\}$, por exemplo, $g(y) = P(X \in B)$, então teremos

$$P(X \in B) = \int P(X \in B \mid Y = y)\, dF_Y(y) = \int g(y)\, dF_Y(y),$$

pelo item 6, §3.1 (integral de Stieltjes no caso discreto). As outras fórmulas gozam de interpretações análogas, logo temos

$$\begin{aligned} P(X \in B) &= \int P(X \in B \mid Y = y)\, dF_Y(y), \\ F_Y(y) &= \int F_X(x \mid Y = y)\, dF_Y(y), \\ EX &= \int E(X \mid Y = y)\, dF_Y(y). \end{aligned} \tag{4.1}$$

Essas fórmulas vão valer também no caso geral (Y não necessariamente discreta), como veremos adiante. Salientamos que a esperança precisa existir para que valha a última fórmula. De fato, quando X for integrável, $\varphi(y) \stackrel{def}{=} E(X \mid Y = y)$ será finito. Nesse caso, a variável aleatória $\varphi(Y)$ será chamada *esperança condicional de X dada Y* e será indicada por $\varphi(Y) = E(X \mid Y)$. Notemos que $E(X \mid Y = y)$ é um valor particular da variável aleatória $E(X \mid Y)$: é o valor quando $Y = y$. Portanto, a última fórmula pode ser interpretada assim:

$$EX = (\text{pelo Teorema 3.1}) = E\varphi(Y) = E\{E(X \mid Y)\}.$$

Em outras palavras, *a esperança de X é igual à esperança da esperança condicional de X dada Y.*

EXEMPLO 2. 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 $0 \le k \le n$, e que depois disso o jogador II lance a mesma moeda k vezes. Seja X o número de caras obtidas pelo jogador II.

Problema. Determinar a esperança de X, supondo independência de todos os lançamentos.

Solução. Seja Y o número de caras nos n lançamentos do jogador I. Decorre das condições do experimento que $Y \sim b\left(n, \frac{1}{2}\right)$ e que

$$X \mid Y = k \sim b\left(k, \frac{1}{2}\right).$$

Por isso, a esperança condicional de X dado que $Y = k$ é a esperança da distribuição $b\left(k, \frac{1}{2}\right)$:

$$E(X \mid Y = k) = \frac{k}{2},$$

ou seja,

$$E(X \mid Y) = \frac{Y}{2}.$$

Utilizando a fórmula, temos

$$EX = E\{E(X \mid Y)\} = E\left(\frac{Y}{2}\right) = \frac{1}{2}EY = \frac{n}{4}.$$

Neste exemplo, não era preciso calcular a distribuição condicional de X dado que $Y = k$, pois esta foi deduzida das condições do experimento. Tal método vale, em geral, para experimentos de duas etapas: partindo do conhecimento da distribuição do resultado da primeira etapa e da distribuição condicional do resultado da segunda etapa dado o resultado da primeira etapa, obtém-se a distribuição (ou a esperança) do resultado da segunda etapa.

Vejamos agora um exemplo em que há algumas contas para fazer na determinação da distribuição condicional.

EXEMPLO 3. De volta ao *processo de Poisson*. Consideremos partículas que chegam a um contador segundo um processo de Poisson com parâmetro $\lambda > 0$. Recordemos as nossas variáveis aleatórias; X_t é o número de partículas que chegam até o instante $t \ge 0$, T_1 é o tempo de chegada da primeira partícula, T_n é o tempo entre a chegada número $n-1$ e a n-ésima.

Já foi visto que $X_t \sim$ Poisson (λt) e $T_1 \sim \exp(\lambda)$. Ocorre que T_1, $T_2, \ldots$ são independentes com a mesma distribuição exponencial de parâmetro λ (não provaremos isto).

Consideremos o seguinte problema: dado que exatamente uma partícula chegou até o tempo $t > 0$, qual a distribuição condicional do seu tempo de chegada? Traduzindo em termos de variáveis aleatórias, temos o problema:

Qual a distribuição condicional de T_1 dado que $X_t = 1$? A resposta é: uniforme em $[0, t]$, i.e.,

$$T_1 \mid X_t = 1 \sim U[0, t].$$

Verifiquemos esta solução, calculando a função de distribuição condicional. Como $X_t = 1$ implica que a primeira partícula chegou até o tempo t, temos $[X_t = 1] \subset [0 < T_1 \le t]$, logo $P(0 < T_1 \le t \mid X_t = 1) = 1$. Consequentemente, temos

$$F_{T_1}(s \mid X_t = 1) = \begin{cases} 0 & \text{se } s \le 0 \\ 1 & \text{se } s \ge t. \end{cases}$$

Se $0 < s < t$,

$$F_{T_1}(s \mid X_t = 1) = P(T_1 \le s \mid X_t = 1) = \frac{P(T_1 \le s, X_t = 1)}{P(X_t = 1)}.$$

Agora, $T_1 \le s$ e $X_t = 1$ significam que a primeira partícula chegou até o instante s e, já que só uma partícula chegou até o instante t, não chegou outra partícula até o instante t. Em outras palavras, houve uma chegada em $(0, s]$ e nenhuma em $(s, t]$. Estes intervalos são disjuntos, logo, por independência, temos

$$P(T_1 \le s, X_t = 1) = P(X_s = 1, X_t - X_s = 0) = \\ = P(X_s = 1)\, P(X_t - X_s = 0).$$

Como o número de chegadas durante um período de duração t tem distribuição de Poisson de parâmetro λt, segue-se que para $0 < s < t$,

$$F_{T_1}(s \mid X_t = 1) = \frac{P(X_s = 1)\, P(X_t - X_s = 0)}{P(X_t = 1)} =$$

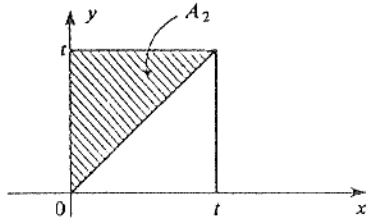
$$= \frac{\lambda s e^{-\lambda s} \cdot e^{-\lambda(t-s)}}{\lambda t e^{-\lambda t}} = \frac{s}{t}.$$

Uma vez que a função de distribuição condicional é a função de distribuição da $U[0, t]$, concluímos que $T_1 \mid X_t = 1 \sim U[0, t]$.

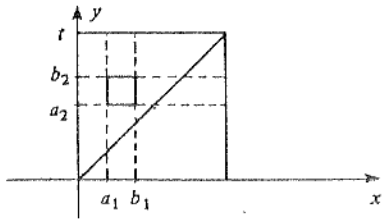
Generalizando um pouco, podemos perguntar o seguinte: dado que chegaram exatamente duas partículas até o tempo $t > 0$, qual a distribuição dos *dois* tempos de chegada? Isto é, qual a distribuição condicional do *vetor* $(T_1, T_1 + T_2)$ dado que $X_t = 2$? (Deve ser óbvio como definir a distribuição condicional de um vetor aleatório (X_1, X_2) dada uma variável aleatória discreta Y:

$$P((X_1, X_2) \in B \mid Y = y_n) = \frac{P((X_1, X_2) \in B,\ Y = y_n)}{P(Y = y_n)},\quad B \in \mathscr{B}^2,$$

onde $\mathscr{B}^2$ é a σ-álgebra de Borel no plano.) A resposta à pergunta é que $(T_1, T_1 + T_2) \mid X_t = 2 \sim U(A_2)$, onde $A_2 = \{(x, y) \in \mathbb{R}^2 : 0 \le x < y \le t\}$, i.e., a distribuição condicional é uniforme em A_2:



Para provar isso, consideremos dois intervalos $(a_1, b_1]$ e $(a_2, b_2]$ tais que $0 < a_1 < b_1 < a_2 < b_2 < t$. Então o retângulo $(a_1, b_1] \times (a_2, b_2]$ está contido em A_2:



Veremos agora que é suficiente provar que

$$P((T_1, T_1 + T_2) \in (a_1, b_1] \times (a_2, b_2] \mid X_t = 2) = \tag{4.2}$$

$$= \frac{2(b_1 - a_1)(b_2 - a_2)}{t^2} = \frac{\text{área (retângulo)}}{\text{área } A_2}.$$

Isto é suficiente porque decorre de (4.2) que

$$P((T_1, T_1 + T_2) \in B \mid X_t = 2) = \frac{\text{área } B}{\text{área } A_2}$$

para todo $B \subset A_2$, B boreliano. (Vale se B é união disjunta enumerável de retângulos do tipo descrito acima, por σ-aditividade. Como todo boreliano de A_2 pode ser aproximado por uma tal união disjunta, segue-se o resultado.) Conseqüentemente,

$$(T_1, T_1 + T_2) \mid X_t = 2 \sim U(A_2).$$

Resta, então, provar a expressão (4.2), cujo primeiro termo é igual a

$$\frac{P(T_1 \in (a_1, b_1],\ T_1 + T_2 \in (a_2, b_2],\ X_t = 2)}{P(X_t = 2)}.$$

Como $X_t = 2$ e $T_1 \in (a_1, b_1]$ e $T_1 + T_2 \in (a_2, b_2]$ se, e somente se, houve uma chegada em $(a_1, b_1]$, outra em $(a_2, b_2]$ e nenhuma nos outros intervalos que compõem $(0, t]$, a saber, $(0, a_1]$, $(b_1, a_2]$ e $(b_2, t]$, temos, por independência,

$$P(T_1 \in (a_1, b_1],\ T_1 + T_2 \in (a_2, b_2],\ X_t = 2) =$$

$$= P(X_{a_1} = 0) P(X_{b_1} - X_{a_1} = 1) P(X_{a_2} - X_{b_1} = 0) P(X_{b_2} - X_{a_2} = 1) P(X_t - X_{b_2} = 0)$$
$$= e^{-\lambda a_1} \lambda(b_1 - a_1) e^{-\lambda(b_1 - a_1)} e^{-\lambda(a_2 - b_1)} \lambda(b_2 - a_2) e^{-\lambda(b_2 - a_2)} e^{-\lambda(t - b_2)} =$$
$$= \lambda^2 (b_1 - a_1)(b_2 - a_2)\, e^{-\lambda t}.$$

Como $P(X_t = 2) = \dfrac{(\lambda t)^2}{2} e^{-\lambda t}$, temos

$$P((T_1, T_1 + T_2) \in (a_1, b_1] \times (a_2, b_2] \mid X_t = 2) = \frac{2(b_1 - a_1)(b_2 - a_2)}{t^2},$$

e (4.2) está provada.

Um argumento análogo mostra que dado que $X_t = n$, a distribuição condicional de $(T_1, T_1 + T_2, \ldots, T_1 + T_2 + \ldots + T_n)$ é $U(A_n)$, onde

$$A_n = \{(x_1, \ldots, x_n) \in \mathbb{R}^n :\ 0 \le x_1 < x_2 < \ldots < x_n \le t\}.$$

A distribuição uniforme nesta pirâmide surge também nas seguintes circunstâncias (daqui a pouco veremos a relação com nosso exemplo): sejam $Y_1, \ldots, Y_n$ variáveis aleatórias independentes, cada uma tendo distribuição $U[0, t]$, $t > 0$. Então, pelo exemplo 22 do Capítulo 2, $U(A_n)$ é a distribuição conjunta das estatísticas de ordem da amostra $Y_1, \ldots, Y_n$, pois a densidade de Y_i é $t^{-1} I_{[0,t]}$ e, portanto, a densidade de $(Y_{(1)}, \ldots, Y_{(n)})$ é

$$f_{Y_{(1)}, \ldots, Y_{(n)}}(y_1, \ldots, y_n) = \begin{cases} \dfrac{n!}{t^n} \text{ Se } 0 \le y_1 < y_2 < \ldots < y_n \le t \\ 0 \text{ caso contrário.} \end{cases}$$

(Notemos que é possível concluir deste resultado que volume $(A_n) = \dfrac{t^n}{n!}$.)

Agora surge uma pergunta bem natural: há alguma relação entre as estatísticas de ordem $Y_{(1)}, \ldots, Y_{(n)}$ e as variáveis condicionais $T_1, \ldots, T_1 + \ldots + T_n$ dado que $X_t = n$? Ou melhor, existe alguma explicação intuitiva para os dois vetores terem a mesma distribuição? Uma possível explicação é a seguinte, baseada no resultado para $n = 1$. Dado que chegaram exatamente n partículas até o instante t, é razoável pensar que cada partícula escolheria seu tempo de chegada uniformemente no intervalo $[0, t]$, como no caso $n = 1$, independentemente das outras partículas. Então os tempos de chegada corresponderiam a n variáveis aleatórias independentes, cada uma distribuída uniformemente em $[0, t]$. Mas os tempos $T_1, T_1 + T_2, \ldots, T_1 + \ldots + T_n$, sendo os tempos de chegada *ordenados*, corresponderiam às estatísticas de ordem de uma amostra da distribuição $U[0, t]$.

Podemos utilizar este resultado para resolver o seguinte problema: se $X \sim$ Poisson (λ_1), $Y \sim$ Poisson (λ_2) e X e Y são independentes, qual a distribuição condicional de X dado que $X + Y = n$? Poderíamos calcular a distribuição condicional diretamente, para obter

$$X \mid X + Y = n \sim b\left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right), \quad n \ge 1$$

(e também para $n = 0$, com a convenção $b(0, p) =$ massa pontual em 0, i.e., $X \sim b(0, p) \Rightarrow P(X = 0) = 1$). Mas o seguinte argumento é mais interessante do ponto de vista probabilístico.

Consideremos um processo de Poisson com parâmetro $\lambda = 1$. Então $X_{\lambda_1} \sim$ Poisson (λ_1), $X_{\lambda_1 + \lambda_2} - X_{\lambda_1} \sim$ Poisson (λ_2) e X_{λ_1} e $X_{\lambda_1 + \lambda_2} - X_{\lambda_1}$ são independentes.

Problema. Qual a distribuição condicional de X_{λ_1} dado que $X_{\lambda_1 + \lambda_2} = n$? (Notemos que as variáveis aleatórias X e Y do problema original estão "mergulhadas" agora no processo de Poisson.)

Solução. Dado que $X_{\lambda_1 + \lambda_2} = n$, os tempos de chegada das primeiras n partículas possuem a distribuição das estatísticas de ordem de uma amostra aleatória de tamanho n da $U[0, \lambda_1 + \lambda_2]$. Portanto, a distribuição do número de chegadas até o tempo λ_1 é a distribuição do número de $Y_{(i)} \le \lambda_1$, onde $Y_{(1)}, \ldots, Y_{(n)}$ são as estatísticas de ordem da amostra $(Y_1, \ldots, Y_n)$, $Y_i \sim U[0, \lambda_1 + \lambda_2]$. Por isso, dado que $X_{\lambda_2 + \lambda_1} = n$, a distribuição condicional de X_{λ_1} é a distribuição do número de $Y_i \le \lambda_1$. Como $P(Y_i \le \lambda_1) = \dfrac{\lambda_1}{\lambda_1 + \lambda_2}$, trata-se de um caso de n ensaios de Bernoulli cem probabilidade $p = \dfrac{\lambda_1}{\lambda_1 + \lambda_2}$ de sucesso (chamemos o i-ésimo ensaio de sucesso se $Y_i \le \lambda_1$, fracasso se $Y_i > \lambda_1$). Logo a distribuição é binomial, com parâmetros n e $\dfrac{\lambda_1}{\lambda_1 + \lambda_2}$. (Observação: para $n = 0$, $P(X_{\lambda_1} = 0 \mid X_{\lambda_1 + \lambda_2} = 0) = 1$.) □

Pelo Teorema da Probabilidade Total (ou pelas fórmulas (4.1)), a distribuição de X é determinada pela distribuição de Y e a distribuição condicional de X dada Y. Então a solução do último problema implica o seguinte resultado: se Y possui distribuição Poisson (λ) e a distribuição condicional de X dado que $Y = n$ é $b(n, p)$, $\forall n \ge 0$, então X tem distribuição Poisson (λp). (Para ver isso, faça $\lambda_1 = \lambda p$ e $\lambda_2 = \lambda(1 - p)$, de modo que $\lambda_1 + \lambda_2 = \lambda$ e $\dfrac{\lambda_1}{\lambda_1 + \lambda_2} = p$.) Vejamos um exemplo.

EXEMPLO 4. Consideremos outro jogo que conta com a participação de dois jogadores, I e II. Neste, 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 (λ), 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 deste ter feito n lançamentos (se $N = 0$, o jogador II não permite nenhum lançamento).

Problema. Se S for o número de caras observadas até o jogador I parar, quais são a distribuição e a esperança de S?

Solução. Como a seqüência de lançamentos é independente de N, a distribuição condicional de S dado que $N = n$ é $b(n, p)$, *i.e.*,

$$S \mid N = n \sim b(n, p).$$

Isto vale ainda no caso em que $n = 0$.

Como $N \sim$ Poisson (λ), S tem distribuição Poisson (λp) e $ES = \lambda p$.

Observação. Se quiséssemos determinar apenas o valor de ES, não era preciso obter a distribuição de S. Com efeito, já vimos que $ES = E\{E(S \mid N)\}$. Como $S \mid N = n \sim b(n, p)$, temos $E(S \mid N = n) = np$, i.e., $E(S \mid N) = Np$. Portanto,

$$ES = E(Np) = pEN = p\lambda.$$

§4.2 Distribuição condicional de X dada Y: caso geral

Nosso objetivo nesta seção é definir distribuição condicional de X dado que $Y = y$ para todo $y \in \mathbb{R}$ e todo par de variáveis aleatórias X e Y definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$.

No §4.1, definimos 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 esta definição era arbitrária, pois o conjunto dos y tais que $P(Y = y) = 0$, i.e., o conjunto $B_0 = \{y_n : n = 1, 2, \ldots\}^c$, também tinha probabilidade zero, no sentido de que $P(Y \in B_0) = 0$. Mas é evidente que essa solução não serve para o caso geral, já que é bem possível que $P(Y = y) = 0$ para todo $y \in \mathbb{R}$ (por exemplo, no caso contínuo).

Mais uma vez, nossa solução será uma aproximação utilizando a definição do caso discreto. Para tanto, seja I um intervalo pequeno de comprimento Δy 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,

$$P(X \in B \mid Y = y) \approx P(X \in B \mid Y \in I) = \frac{P(X \in B,\ Y \in I)}{P(Y \in I)}.$$

Se $P(X \in B \mid Y \in I)$ converge para um limite quando $\Delta y \to 0$, chamemos o limite $P(X \in B \mid Y = y)$. Se $P(Y \in I) = 0$ para alguma vizinhança I, então definamos arbitrariamente a probabilidade condicional, digamos $P(X \in B \mid Y = y) = P(X \in B)$. Então nossa definição (informal, por enquanto; as definições formais serão dadas na seção seguinte) será:

Relação 1. $P(X \in B \mid Y = y) = \lim_{\Delta y \to 0} P(X \in B \mid Y \in I)$, onde I é um intervalo que contém y, de comprimento Δy.

Observação. O limite existe quase certamente, no sentido de que existe para quase todos os valores possíveis de Y. De fato, veremos que a Relação 1 dá o "valor certo" da probabilidade condicional com probabilidade 1 (o "valor certo" é o valor dado pela distribuição condicional regular, a ser definida no §4.3).

Embora essa definição seja construtiva e dê uma receita para se calcular a distribuição condicional, ela não é muito prática (como o leitor descobrirá lendo os exemplos e fazendo os exercícios). O que se costuma fazer é conjeturar – ou, se quiser, "chutar" – a distribuição condicional e depois verificá-la. Mais adiante voltaremos a este assunto.

Essa definição não é a única existente. Há outra, baseada na Teoria da Medida, que é muito mais útil no sentido teórico (pode-se provar muito mais com ela). Para termos uma idéia da base desta definição, consideremos novamente o caso discreto.

Para Y discreta, as fórmulas (4.1) dizem que a distribuição (ou função de distribuição, ou esperança) de X é determinada pela distribuição de Y e a distribuição (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 isso, basta notar que para todo x e y,

$$\begin{aligned} F_{X,Y}(x, y) &= P(X \le x, Y \le y) = \sum_{n:\, y_n \le y} P(X \le x, Y = y_n) = \\ &= \sum_{n:\, y_n \le y} P(Y = y_n)\, P(X \le x \mid Y = y_n) = \\ &= \sum_{n:\, y_n \le y} P(Y = y_n)\, F_X(x \mid Y = y_n) = \\ &= \int_{-\infty}^{y} F_X(x \mid Y = t)\, dF_Y(t), \end{aligned}$$

valendo a última passagem pela forma da integral de Stieltjes no caso discreto (item 6, §3.1).

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. Mas ocorre 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 (a prova não é fácil). Além disso, será visto no §4.3 que esta família está em concordância com a definição de distribuição condicional com limite dada pela Relação 1. Isto significa que a seguinte relação pode ser usada como *definidora* da distribuição condicional de X dada Y:

Relação 2. $F_{X,Y}(x, y) = \int_{-\infty}^{y} F_X(x \mid Y = t)\, dF_Y(t), \quad (x, y) \in \mathbb{R}^2.$

(É claro que a distribuição condicional será a distribuição definida pela função de distribuição condicional.)

Entretanto, o resultado justificado a adoção da Relação 2 como uma segunda definição serve apenas para garantir a existência da distribuição condicional e não contribui para ensinar *como achá-la*. Reiteramos que a melhor maneira de se achar a distribuição condicional é "chutar" – usando alguns princípios que serão introduzidos adiante – e assim obter um "candidato" para a distribuição condicional. Em seguida verifica-se se o candidato realmente é a distribuição condicional. Para tal verificação, é suficiente conferir que o candidato

satisfaz uma das relações acima (Relação 1 ou Relação 2). Este é o procedimento mais indicado para o caso geral. Consideremos agora alguns casos simples em que a solução vem quase de imediato.

Casos de fácil solução.

Caso 1. Y discreta. Neste caso já definimos a distribuição condicional. Para a seqüência finita ou enumerável de valores $y_1, y_2, \ldots$ tais que $P(Y = y_n) > 0$, definimos

$$P(X \in B \mid Y = y_n) = \frac{P(X \in B, Y = y_n)}{P(Y = y_n)}, \quad B \in \mathcal{B}.$$

O conjunto dos outros valores de y tem probabilidade zero, no sentido de que $P(Y \in \{y_n : n = 1, 2, \ldots\}^c) = 0$, e admitimos a definição arbitrária

$$P(X \in B \mid Y = y) = P(X \in B), \quad B \in \mathcal{B},$$

para todo y que não seja um dos y_n. Desta maneira, definimos uma distribuição condicional dado que $Y = y$ para todo $y \in \mathbb{R}$. Verifiquemos que esta distribuição concorda com a dada pelas duas relações.

Relação 2 já foi verificada; de fato, foi derivada justamente no caso discreto. Embora não seja necessário verificar a Relação 1, bastando verificar uma das duas definições, vamos fazê-lo para mostrar a concordância entre as definições.

Como basta verificar a Relação 1 para "quase todo" y, i. e., para um conjunto de y's que contém o valor de Y com probabilidade um, verificaremos para todo y_n. Se $y_n \in I$, onde I é um intervalo de comprimento Δy_n, então

$$P(Y \in I) \to P(Y = y_n) \text{ quando } \Delta y_n \to 0,$$

pela continuidade de probabilidade. Notemos que aqui o índice n é fixo e o comprimento do intervalo contendo y_n é que converge.

(*Observação.* A continuidade de probabilidade foi demonstrada para seqüências monótonas de eventos, enquanto que esta convergência seja mais geral. Mas é conseqüência imediata do resultado anterior, pois esse implica que

$$P\left(y_n - \frac{1}{m} < Y < y_n + \frac{1}{m}\right) \to P(Y = y_n) \text{ quando } m \to \infty,$$

e quando $\Delta y_n < \frac{1}{m}$, temos $I \subset \left(y_n - \frac{1}{m},\ y_n + \frac{1}{m}\right)$ e

$$P(Y = y_n) \le P(Y_n \in I) \le P\left(y_n - \frac{1}{m} < Y < y_n + \frac{1}{m}\right).$$

Pela mesma continuidade de probabilidade, temos

$$P(X \in B,\ Y \in I) \to P(X \in B,\ Y = y_n) \text{ quando } \Delta y_n \to 0.$$

Como $P(Y = y_n) > 0$, segue-se que

$$P(X \in B \mid Y \in I) = \frac{P(X \in B,\ Y \in I)}{P(Y \in I)} \xrightarrow[\Delta y_n \to 0]{} \frac{P(X \in B,\ Y = y_n)}{P(Y = y_n)},$$

como queríamos demonstrar.

Portanto, as duas relações definidoras de distribuição condicional são consistentes com a definição vista no caso discreto.

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

$$P(X \in B \mid Y = y) = P(X \in B),\ B \in \mathcal{B},\ y \in \mathbb{R},$$

ou seja, a distribuição condicional é a própria distribuição (não-condicional) de X. Verifiquemos as duas relações.

Relação 1: Pela independência, $P(X \in B \mid Y \in I) = P(X \in B)$ para todo intervalo I e todo $B \in \mathcal{B}$. Logo, $\forall B \in \mathcal{B}$,

$$P(X \in B \mid Y \in I) \to P(X \in B) \text{ quando } \Delta y \to 0.$$

Relação 2: Nosso candidato para $F_X(x \mid Y = y)$ é $F_X(x)$. Substituindo, temos

$$\int_{-\infty}^{y} F_X(x) dF_Y(t) = F_X(x) \int_{-\infty}^{y} dF_Y(t) = F_X(x) F_Y(y) = F_{X,Y}(x, y),$$

onde a penúltima passagem segue da natureza da integral de Stieltjes (item 3, §3.1) e a última é conseqüência da independência.

Em resumo, nosso candidato satisfaz as duas relações e é a distribuição condicional.

Caso III. X e Y possuem densidade conjunta $f(x, y)$. Pensemos assim: dado que $Y = y$, os valores possíveis do vetor (X, Y) são os valores (x, y), $x \in \mathbb{R}$, y fixo. Ora, a densidade $f(x, y)$ representa a *chance relativa* de observar o par (x, y), no sentido de que *antes* de observar os valores das variáveis aleatórias X e Y, $f(x, y)$ fornece uma idéia da chance relativa de observar (x, y). Por exemplo, se $f(x_1, y) = 2f(x_2, y)$, então, informalmente, (x_1, y) tem o dobro da chance de (x_2, y) de ser observado. Neste caso, é razoável pensar que, dado que $Y = y$, x_1 mantenha duas vezes a chance de x_2, na distribuição condicional de X.

Portanto, nosso candidato para a distribuição condicional de X dado que $Y = y$ manterá as chances proporcionais a $f(x, y)$, com y fixo. Com efeito, vamos *normalizar* $f(x, y)$, $x \in \mathbb{R}$, para que seja uma nova densidade, a densidade da distribuição condicional. Para tanto, como ela já é uma função não-negativa, basta dividir por sua integral sobre a reta. Então, nosso candidato para a densidade da distribuição condicional será

$$f(x \mid y) = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dx} = \frac{f(x, y)}{f_Y(y)},\ x \in \mathbb{R}.$$

Se $f_Y(y) > 0$, então $f(x \mid y)$ é uma densidade (pois é não-negativa e $\int_{-\infty}^{\infty} f(x \mid y) dx = 1$), chamada *densidade condicional de X dado que* $Y = y$. Notemos que dado que $Y = y$, $f(x \mid y)$ realmente preserva chances relativas, pois

$$\frac{f(x_1 \mid y)}{f(x_2 \mid y)} = \frac{f(x_1, y)}{f(x_2, y)}.$$

Esse é um exemplo de aplicação do *princípio da preservação de chances relativas* ("odds" em inglês). Adiante voltaremos a falar deste princípio.

Vejamos agora se nosso candidato realmente satisfaz as duas relações que definem a distribuição condicional. Para a Relação 1, já que precisamos verificar se nosso candidato $f(x \mid y)$ é realmente a densidade da distribuição condicional, basta verificar que para cada boreliano B,

$$\int_B f(x \mid y) dx = P(X \in B \mid Y = y) = \lim_{\Delta y \to 0} P(X \in B \mid Y \in I).$$

Daremos uma prova parcial desta relação; a prova completa necessita de conceitos da Teoria da Medida.

Sejam y fixo e I um intervalo de comprimento Δy contendo y. Temos

$$P(X \in B \mid Y \in I) = \frac{P(X \in B,\ Y \in I)}{P(Y \in I)} = \frac{\int_I \int_B f(x,t)dxdt}{\int_I f_Y(t)dt}.$$

Se a densidade f_Y é contínua no ponto y, sabemos do Cálculo que

$$\frac{\int_I f_Y(t)dt}{\Delta y} \to f_Y(y) \text{ quando } \Delta y \to 0.$$

Se definimos $g(t) = \int_B f(x,t)dx$, então, supondo g contínua em y,

$$\frac{\int_I \int_B f(x,t)dxdt}{\Delta y} = \frac{\int_I g(t)dt}{\Delta y} \xrightarrow[\Delta y \to 0]{} g(y).$$

Tomando o quociente dos dois limites, temos

$$P(X \in B \mid Y \in I) = \frac{\int_I g(t)dt}{\int_I f_Y(t)dt} \xrightarrow[\Delta y \to 0]{} \frac{g(y)}{f_Y(y)} =$$

$$= \frac{\int_B f(x,y)dx}{f_Y(y)} = \int_B \frac{f(x,y)}{f_Y(y)}dx = \int_B f(x \mid y)dx,$$

como queríamos demonstrar.

Quanto à Relação 2, nosso candidato para a função de distribuição condicional é $F_X(x \mid Y = y) = \int_{-\infty}^{x} f(s \mid y)ds$. A relação está satisfeita pelo candidato, pois

$$\int_{-\infty}^{y} \left(\int_{-\infty}^{x} f(s \mid t)ds \right) dF_Y(t) = \int_{-\infty}^{y} \left(\int_{-\infty}^{x} \frac{f(s,t)}{f_Y(t)}ds \right) f_Y(t)\,dt =$$

$$= \int_{-\infty}^{y} \int_{-\infty}^{x} f(s,t)dsdt = \text{(pela definição de densidade)} =$$

$$= F_{X,Y}(x,y), \quad \forall x, y.$$

Observação. Se $f_Y(y) = 0$, $f(x \mid y)$ pode ser arbitrariamente definida, digamos $f(x \mid y) = f_X(x)$. Isso é intuitivamente evidente, pois os valores y tais que $f_Y(y) = 0$ não são valores possíveis de Y. Formalmente,

$$P(Y \in \{y : f_Y(y) = 0\}) = P_Y(\{y : f_Y(y) = 0\}) =$$

$$= \int_{\{y : f_Y(y) = 0\}} f_Y(y)dy = \int_{\{y : f_Y(y) = 0\}} 0 \cdot dy = 0.$$

Outras notações para a densidade condicional de X dada Y:

$$f(x \mid y) = f_{X \mid Y}(x \mid y) = f_X(x \mid Y = y).$$

EXEMPLO 5. Suponha que o vetor (X, Y) possua distribuição normal bivariada com densidade

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

$$\left.\left. - 2\rho \left(\frac{x-\mu_1}{\sigma_1} \right) \left(\frac{y-\mu_2}{\sigma_2} \right) + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \right] \right\}.$$

Neste caso, já sabemos qual a densidade marginal de Y, pois $Y \sim N(\mu_2, \sigma_2^2)$. Mas para obter a densidade condicional, não é necessário calcular a densidade marginal, nem utilizá-la diretamente na fórmula, pois para y fixo a densidade condicional é proporcional a $f(x,y)$. Por isso, podemos tratar y como constante na densidade conjunta, padronizar $f(x,y)$ como função de x e assim obter a densidade condicional.

Portanto, colocando em evidência todo fator que não dependa de x, podemos escrever

$$f(x \mid y) = c(\sigma_1^2, \sigma_2^2, \rho, y) \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_1}{\sigma_1} \right)^2 - \right.\right.$$

$$\left.\left. - 2\rho \left(\frac{x-\mu_1}{\sigma_1} \right) \left(\frac{y-\mu_2}{\sigma_2} \right) \right] \right\},$$

onde a constante $c(\sigma_1^2, \sigma_2^2, \rho, y)$ é determinada pela equação $\int f(x \mid y)dx = 1$. É claro que essa densidade condicional é normal. Com efeito, completando o quadrado obtemos

$$f(x \mid y) = c_1(\sigma_1^2, \sigma_2^2, \rho, y) \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_1}{\sigma_1} \right) - \right.\right.$$

$$\left.\left. - \rho \left(\frac{y-\mu_2}{\sigma_2} \right) \right]^2 \right\} = c_1(\sigma_1^2, \sigma_2^2, \rho, y) \exp\left\{ -\frac{1}{2\sigma_1^2(1-\rho^2)} \times \right.$$

$$\left. \times \left[x - \mu_1 - \frac{\rho\sigma_1}{\sigma_2}(y-\mu_2) \right]^2 \right\}.$$

Esta é a densidade da distribuição normal com média $\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(y-\mu_2)$ e variância $\sigma_1^2(1-\rho^2)$. Por isso, escrevemos

$$X \mid Y = y \sim N\left(\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(y-\mu_2), \sigma_1^2(1-\rho^2) \right),$$

ou seja, a distribuição condicional de X dado que $Y = y$ é a citada normal.

§4.3 Definições formais e teoremas de existência

Nesta seção daremos os teoremas que justificam a nossa utilização de Relação 1 e Relação 2 como definições de distribuição condicional no caso geral. O leitor, se quiser, poderá pular esta parte teórica e passar diretamente às *observações gerais* no final da seção.

DEFINIÇÃO 4.1. Sejam X e Y variáveis aleatórias definidas no mesmo espaço de probabilidade $(\Omega, \mathscr{A}, P)$. Uma função $P(X \in B \mid Y = y)$, definida para B boreliano e $y \in \mathbb{R}$, será chamada uma *distribuição condicional (regular) para X dada Y* se

(i) para todo $y \in \mathbb{R}$ fixo, $P(X \in B \mid Y = y)$ define uma probabilidade em $\mathscr{B}$, a σ-álgebra de Borel na reta, e

(ii) para todo $B \in \mathscr{B}$ fixo, $P(X \in B \mid Y = y)$ é função mensurável de y e, $\forall (x, y) \in \mathbb{R}^2$,

$$\int_{-\infty}^{y} P(X \leq x \mid Y = t) dF_Y(t) = P(X \leq x, Y \leq y).$$

Linguagem. $P(X \in B \mid Y = y)$ é chamada *probabilidade condicional* de X pertencer a B dado que $Y = y$. A probabilidade $P(X \in \cdot \mid Y = y)$ é a *distribuição condicional* de X dado que $Y = y$. (Observação: o ponto "$\cdot$" representa o argumento da função, de modo que y é fixo.) A função $F_X(\cdot \mid Y = y) \stackrel{def}{=} P(X \leq \cdot \mid Y = y)$ é a *função de distribuição condicional* de X dado que $Y = y$. Então, a condição (ii) da definição se escreve assim:

$$\int_{-\infty}^{y} F_X(x \mid Y = t) dF_Y(t) = F_{X,Y}(x, y) \text{ para todo } x, y.$$

Esta é a Relação 2.

TEOREMA 4.1. *Sejam X e Y variáveis aleatórias em $(\Omega, \mathscr{A}, P)$. 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 $P_1(X \in B \mid Y = y)$ e $P_2(X \in B \mid Y = y)$ são ambas distribuições condicionais para X dada Y, então existe um boreliano B_0 tal que $P(Y \in B_0) = 1$ e $P_1(X \in B \mid Y = y) = P_2(X \in B \mid Y = y)$ $\forall B \in \mathscr{B}$, $\forall y \in B_0$.*

Prova. Teórica. Poderá ser tirada do Breiman [5], Capítulo 4. □

DEFINIÇÃO 4.2. Sejam $y \in \mathbb{R}$ e $B \in \mathscr{B}$. A *probabilidade condicional* de que $X \in B$ dado que $Y = y$, é definida por

$$P(X \in B \mid Y = y) = \lim_{\Delta y \to 0} P(X \in B \mid Y \in I),$$

onde I representa um intervalo de comprimento Δy contendo y.

Esta definição corresponde, então, à Relação 1.

TEOREMA 4.2. *Para cada $B \in \mathscr{B}$ 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 $P(X \in B \mid Y = y)$ como definida na Definição 4.1, quase certamente.*

Prova. Teórica. (Uma modificação do Teorema 8.6. de Rudin [17]). □

COROLÁRIO. *Para se achar a distribuição condicional de X dada Y, basta obter $\lim_{\Delta y \to 0} P(X \in B \mid Y \in I)$ para todo B em uma apropriada classe enumerável $\mathscr{C}$ de borelianos que gera a σ-álgebra de Borel. Por exemplo, $\mathscr{C}_1 = \{(-\infty, r] : r \text{ racional}\}$ ou $\mathscr{C}_2 = \{(r_1, r_2) : -\infty < r_1 < r_2 < \infty, r_1 \text{ e } r_2 \text{ racionais}\}$.*

Prova. As classes $\mathscr{C}_1$ e $\mathscr{C}_2$ são ambas enumeráveis. Pelo teorema, para cada membro de $\mathscr{C}_1$ ou $\mathscr{C}_2$, o limite acerta no valor da distribuição condicional regular com probabilidade um. Como a interseção de um número enumerável de eventos de probabilidade um também é de probabilidade um, o limite acerta no valor da distribuição condicional regular simultaneamente para todos os membros de $\mathscr{C}_1$ ou $\mathscr{C}_2$, com probabilidade um. Mas qualquer probabilidade é determinada pelos seus valores em $\mathscr{C}_1$ ou $\mathscr{C}_2$, logo segue-se o corolário. □

Observações gerais. Suponhamos que desejemos achar a distribuição condicional de X dada Y. Se o problema não se enquadra em um dos casos simples I, II, III do §4.2, então nosso método é chutar: aplicando certos princípios (por exemplo, preservação de chances relativas, que conduziu à solução no caso III), obtemos para cada $y \in \mathbb{R}$ uma probabilidade $P(B \mid y)$, $B \in \mathscr{B}$, que é nosso candidato para a distribuição condicional de X dado que $Y = y$. Então verificamos se $P(B \mid y) = P(X \in B \mid Y = y)$, conferindo a Relação 1 ou a Relação 2, i. e., conferindo a Definição 4.2 ou 4.1. Isto significa uma verificação de que

$$P(B \mid y) = \lim_{\Delta y \to 0} P(X \in B \mid Y \in I)$$

ou de que

$$F_{X,Y}(x, y) = \int_{-\infty}^{y} P((-\infty, x] \mid t)dF_Y(t), \ \forall x, y.$$

Em geral, a Definição 4.2 é mais fácil de se verificar. Além disso, esta não precisa ser verificada para todo y e B; basta uma verificação para quase todo y e para todo B intervalo aberto com extremos racionais. O seguinte exemplo é talvez a aplicação mais simples deste método.

EXEMPLO 6. Qual a distribuição condicional de Y dada Y? É fácil obter o candidato, usando nossa intuição: dado que $Y = y$, então Y é igual a y(!). Por isso, nosso candidato será

$$P(Y = y \mid Y = y) = 1,$$

ou seja, a distribuição condicional de Y dado que $Y = y$ é *massa pontual em* y, a distribuição que atribui probabilidade 1 ao ponto y.

Verifiquemos a Definição 4.2. Seja y fixo, e seja $B = (r_1, r_2)$, onde r_1 e r_2 são racionais tais que $r_1 < y < r_2$. Então

$$P(Y \in B \mid Y = y) = \lim_{\Delta y \to 0} P(Y \in B \mid Y \in I) = \lim_{\Delta y \to 0} \frac{P(Y \in B, Y \in I)}{P(Y \in I)}.$$

Para Δy pequeno, temos $I \subset B$ e $[Y \in B, Y \in I] = [Y \in I]$, logo

$$P(Y \in B \mid Y = y) = \lim_{\Delta y \to 0} \frac{P(Y \in I)}{P(Y \in I)} = 1.$$

Não é preciso obter o limite para outros borelianos B, pois a distribuição condicional já está determinada: como a distribuição condicional atribui probabilidade 1 a todo intervalo de extremos racionais contendo y, atribui probabilidade 1 ao ponto y (basta fazer B diminuir para $\{y\}$), de modo que $P(Y = y \mid Y = y) = 1$. Assim, nosso candidato está confirmado.

O exemplo 6 mostra que, para y fixo, geralmente não precisamos obter o limite para todo intervalo de extremos racionais, mas apenas para um número suficiente à determinação da distribuição condicional. Neste exemplo, nosso candidato era massa pontual em y, por isso consideramos apenas intervalos contendo y.

Notemos que a teoria diz que o limite na Definição 4.2 dá o valor certo para *todos* os intervalos racionais, para "quase todos" os valores y. Portanto, se y for um desses "bons" valores, o limite acertará para todos os intervalos racionais e, em particular, para os intervalos que contêm y. Isto quer dizer que para os "bons" valores y, nosso candidato é igual à distribuição condicional regular. Como os outros valores y têm, em conjunto, probabilidade zero, nosso candidato é a distribuição condicional.

Advertência. Para B fixo, o limite na Definição 4.2 acerta com probabilidade 1. Ocorre que (i) ainda no caso do limite existir, poderá errar no alvo e (ii) para y fixo, os limites para os vários borelianos B poderão ser incompatíveis. O problema é que o número de borelianos é não-enumerável e os eventos de probabilidade zero poderão acumular, formando em conjunto um evento até de probabilidade 1. É por isso que *recomendamos a obtenção do limite apenas para intervalos racionais.*

Para ilustrar o problema, vejamos de novo o exemplo 6. Como nosso candidato para a distribuição condicional, dado que $Y = y$, era massa pontual em y, seria talvez natural tentar verificar isso diretamente, calculando $P(Y = y \mid Y = y)$ como um limite. Então para y_0 fixo, ponhamos $B = \{y_0\}$ e calculemos

$$P(Y = y_0 \mid Y = y) = \lim_{\Delta y \to 0} P(Y = y_0 \mid Y \in I).$$

Se a distribuição de Y fosse contínua, digamos $N(0, 1)$, então teríamos $P(Y = y_0) = 0$ e

$$P(Y = y_0 \mid Y = y) = \lim_{\Delta y \to 0} \frac{P(Y = y_0, Y \in I)}{P(Y \in I)} = 0.$$

A conclusão seria de que $P(Y = y_0 \mid Y = y_0) = 0$, para todo y_0, o que contradiria nossa solução anterior, a solução certa.

O problema com esta tentativa é que, com $B = \{y_0\}$, o limite acerta na probabilidade condicional, dado que $Y = y$, para todo y exceto para o ponto y_0. (Notemos que não teríamos este problema no caso discreto.) Já que a probabilidade de Y tomar este valor é zero, o resultado não contradiz a teoria. Contradiz, contudo, a nossa intuição, pois é justamente no ponto y_0 que nosso candidato assume o valor 1.

Em resumo, esta tentativa mostra a futilidade do cálculo de $P(Y = y \mid Y = y)$ como um limite no caso geral. Como o número de pontos é não-enumerável, tal cálculo poderá resultar em contradições que dificultam a obtenção da distribuição condicional.

§4.4 Exemplos

Nesta seção, vamos considerar alguns exemplos que não se enquadram nos casos I, II, III. Para obter as distribuições condicionais, usaremos dois princípios importantes que enunciamos agora.

O princípio da preservação de chances relativas. Este princípio diz que condicionalmente, dada a ocorrência de um evento A, os resultados possíveis (i. e., $\omega \in A$) mantêm as mesmas chances relativas que tinham anterior à realização do experimento.

Em termos de variáveis aleatórias, o princípio da preservação de chances relativas pode ser expresso da seguinte forma: dado que $Y = y$, os valores possíveis de X mantêm as mesmas chances relativas de antes do experimento. Aqui, um valor possível de X é um x tal que (x, y) era um valor possível de (X, Y) antes do experimento, e o princípio diz que estes pontos x mantêm, na distribuição condicional, as mesmas chances relativas que os pontos (x, y) tinham na distribuição conjunta de X e Y.

Já vimos uma aplicação direta deste princípio quando obtivemos a densidade condicional no caso contínuo (caso III, §4.2). A distribuição condicional nos outros dois casos de fácil solução também pode ser considerada como conseqüência deste princípio (exercício 21). Além disso, o exemplo 6 é conseqüência imediata, pois dado que Y tomou o valor y, este mesmo valor tornou-se o único valor condicionalmente possível de Y — notemos que os pares (y, y) eram os únicos valores possíveis do vetor aleatório (Y, Y).

O princípio da preservação de chances relativas poderá ser aplicado *sempre*. O seguinte princípio é aplicável, e indicado, nos casos em que queremos obter a distribuição condicional de uma *função* de Y ou, mais geralmente, de uma função de Y e uma variável aleatória X cuja distribuição condicional já seja conhecida.

O 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 se obter a distribuição condicional de $\varphi(X, Y)$ dado que $Y = y$, basta substituir Y pelo valor y e X pela variável condicional. Formalmente, temos a seguinte

PROPOSIÇÃO 4.1. (*Princípio da substituição para distribuição condicional.*) *Sejam X e Y variáveis aleatórias em $(\Omega, \mathscr{A}, P)$, $\varphi(x, y)$ uma função mensurável. Se a distribuição condicional de X dada Y é*

$$P(X \in B \mid Y = y),\ B \in \mathscr{B},\ y \in \mathbb{R},$$

então a distribuição condicional para $\varphi(X, Y)$ dada Y é

$$P(\varphi(X, Y) \in B \mid Y = y) = P(\varphi(X, y) \in B \mid Y = y) =$$
$$= P(X \in \{x : \varphi(x, y) \in B\} \mid Y = y),\ B \in \mathscr{B},\ y \in \mathbb{R}.$$

Observação. A segunda probabilidade condicional acima deveria ser interpretada como uma etapa transicional cujo objetivo é ajudar a memória. Não é desejável calcular para cada y a distribuição condicional da variável aleatória $\varphi(X, y)$ — já que no caso geral o número de y é não-enumerável, os erros podem se acumular e estragar o resultado final — mas sim utilizar a distribuição condicional conhecida de X. Podemos colocar o resultado da proposição em símbolos assim:

$$\varphi(X, Y) \mid Y = y \sim \varphi(X \mid Y = y, y),$$

ou seja, a distribuição condicional de $\varphi(X, Y)$ dado que $Y = y$ é a mesma que a distribuição de $\varphi(X, y)$, onde a variável aleatória X possui a sua distribuição condicional.

Prova. Teoria da Medida para o caso geral. Provaremos para o caso de Y *discreta*. Suponha que $P(Y = y_n) = p(y_n) \geq 0$, onde $\sum_n p(y_n) = 1$. Para $B \in \mathscr{B}$, temos

$$P(\varphi(X, Y) \in B \mid Y = y_n) = \frac{P(\varphi(X, Y) \in B,\ Y = y_n)}{p(y_n)} =$$
$$= \frac{P(\varphi(X, y_n) \in B,\ Y = y_n)}{p(y_n)},$$

pois os eventos $[\varphi(X, Y) \in B,\ Y = y_n]$ e $[\varphi(X, y_n) \in B,\ Y = y_n]$ são idênticos (você sabe explicar por quê?).

Assim, a primeira equação da proposição está provada, para "quase todo" y. (Novamente, os outros valores y são desprezíveis; neste caso, $y \notin \{y_n : n \geq 1\}$.) Para a segunda equação, temos

$$\frac{P(\varphi(X, y_n) \in B,\ Y = y_n)}{p(y_n)} = \frac{P(X \in \{x : \varphi(x, y_n) \in B\},\ Y = y_n)}{p(y_n)} =$$
$$= P(X \in \{x : \varphi(x, y_n) \in B\} \mid Y = y_n). \quad \square$$

EXEMPLO 7. Dado que $Y = y$, qual a distribuição condicional de $Z = g(Y)$? Já vimos que a distribuição condicional de Y é massa pontual em y, ou seja, a variável aleatória condicional é constante e assume o valor y. Portanto, a distribuição condicional de Z é também massa pontual, desta vez em $g(y)$. Em outras palavras,

$$P(g(Y) = g(y) \mid Y = y) = 1,$$

o que é intuitivamente óbvio.

Notemos que na aplicação da proposição, identificamos X e Y, pois Z é função somente de Y.

EXEMPLO 8. 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 dada a variável aleatória $|X|$?

Utilizemos o princípio da preservação de chances relativas para achar um candidato. Dado que $|X| = y > 0$, então os únicos valores possíveis de X são y e $-y$. De fato, $|X| = y$ se, e somente se, $X = y$ ou $X = -y$. Pela simetria de X, os dois valores y e $-y$ tinham, antes do experimento, a mesma chance de serem o valor de X. Portanto, nosso candidato é

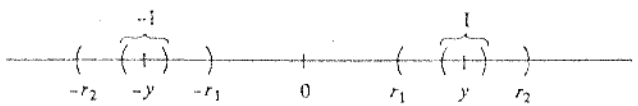
$$P(X = y \mid |X| = y) = \frac{1}{2} = PX = -y \mid |X| = y), \text{ se } y > 0.$$

Se $y = 0$, nosso candidato é

$$P(X = 0 \mid |X| = 0) = 1,$$

pois $X = 0 \Leftrightarrow |X| = 0$. Como $|X| \geq 0$, não precisamos considerar os valores $y < 0$. (Para satisfazer a definição formal, basta definirmos arbitrariamente se $y < 0$.)

Supondo $y > 0$, vamos verificar a Definição 4.2. Desta vez, a distribuição condicional estaria concentrada nos dois pontos y e $-y$, de modo que é conveniente considerar pequenos intervalos racionais contendo esses pontos:



Seja $B = (r_1, r_2)$, onde $0 < r_1 < y < r_2$, r_1 e r_2 racionais. Para I pequeno tal que $I \subset B$, temos

$$P(X \in B, \mid X \mid \in I) = P(X \in I) = \text{(pela simetria)} =$$

$$= \frac{1}{2}\{P(X \in I) + P(X \in -I)\} = \frac{1}{2}P(\mid X \mid \in I),$$

e

$$P(X \in -B, \mid X \mid \in I) = P(X \in -I) = \frac{1}{2}P(\mid X \mid \in I).$$

Logo $I \subset B$ implica

$$P(X \in B \mid \mid X \mid \in I) = \frac{1}{2} = P(X \in -B \mid \mid X \mid \in I).$$

Portanto,

$$P(X \in B \mid \mid X \mid = y) = \lim_{\Delta y \to 0} P(X \in B \mid \mid X \mid \in I) = \frac{1}{2}$$

e

$$P(X \in -B \mid \mid X \mid = y) = \lim_{\Delta y \to 0} P(X \in -B \mid \mid X \mid \in I) = \frac{1}{2}.$$

Fazendo B decrescer para $\{y\}$, vemos que a distribuição condicional atribui probabilidade $\frac{1}{2}$ a cada um dos pontos y e $-y$. A prova de que $P(X = 0 \mid |X| = 0)$ é deixada para o leitor (basta escolher $B = (r_1, r_2)$, onde $r_1 < 0 < r_2$).

Observação. Antes de prosseguirmos, notemos que as definições de distribuição condicional podem ser estendidas ao caso de $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ *vetores* aleatórios. (No exemplo 3, do processo de Poisson, $\underset{\sim}{X}$ foi o vetor $(T_1, T_1 + T_2, \ldots, T_1 + \ldots + T_n)$.) Em particular, se $\underset{\sim}{X} = (X_1, \ldots, X_k)$ e $\underset{\sim}{Y} = (Y_1, \ldots, Y_n)$, então $P(\underset{\sim}{X} \in B \mid \underset{\sim}{Y} = \underset{\sim}{y})$ é função de $B \in \mathcal{B}^k$, i. e., B boreliano no $\mathbb{R}^k$, e de $\underset{\sim}{y} \in \mathbb{R}^n$. O Teorema 4.1 ainda vale neste caso, no sentido de que a distribuição condicional (regular) existe e é única. O Teorema 4.2 vale se Y é variável aleatória; se $\underset{\sim}{Y}$ é vetor, vale caso I seja certo tipo de retângulo n-dimensional, e não vamos nos preocupar com este ponto. A Proposição 4.1 (princípio da substituição) também vale no caso geral.

EXEMPLO 9. No processo de Poisson, podemos obter a distribuição condicional de T_1 dado que $X_t = n$, onde $n \geq 1$, utilizando o princípio da substituição. Vimos no exemplo 3 que os n primeiros tempos de chegada, dado que $X_t = n$, tinham a distribuição das estatísticas de ordem de uma amostra aleatória de tamanho n da distribuição $U[0, t]$. Então T_1, sendo função desses n tempos, a saber, o primeiro deles, possui condicionalmente a distribuição do mínimo de uma amostra de n uniformes em $[0, t]$. Portanto, para $n \geq 1$,

$$P(T_1 \leq s \mid X_t = n) = \begin{cases} 0 & \text{se } s < 0 \\ 1 - \left(\dfrac{t-s}{t}\right)^n & \text{se } 0 \leq s \leq t \\ 1 & \text{se } s > t. \end{cases}$$

(*Exercício.* Verifique esse resultado através de um cálculo direto, que não é mais difícil neste caso. Ache, também, a distribuição condicional de T_1 dado que $X_t = 0$.)

EXEMPLO 10. Sejam X e Y variáveis aleatórias independentes, cada uma tendo distribuição $N(0, \sigma^2)$, onde $\sigma^2 > 0$. Qual a distribuição condicional de (X, Y) dada $\sqrt{X^2 + Y^2}$, a distância entre (X, Y) e a origem?

Para $z > 0$, $\sqrt{X^2 + Y^2} = z$ se, e somente se, (X, Y) pertence à circunferência do círculo de raio z. Logo a distribuição condicional, dado que $\sqrt{X^2 + Y^2} = z$, está concentrada em $\{(x, y): x^2 + y^2 = z^2\}$. Apliquemos, então, o princípio da preservação de chances relativas: como a densidade conjunta de X e Y é

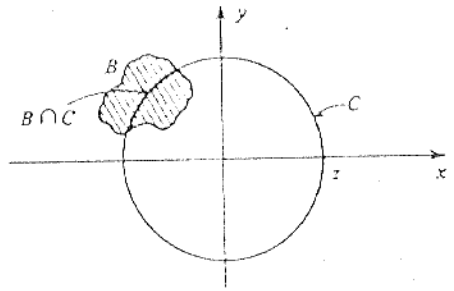
$$f(x, y) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{(x^2 + y^2)}{2\sigma^2}\right\} = \frac{1}{2\pi\sigma^2} \exp\left\{\frac{z^2}{2\sigma^2}\right\},$$

a densidade é constante para todos os (x, y) na circunferência do círculo de raio z. Por isso, antes do experimento todos esses (x, y) eram "equiprováveis", e nosso candidato será a distribuição uniforme na circunferência do círculo.

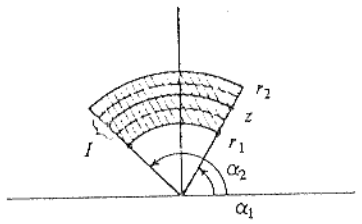
Logo nosso candidato, para $B \in \mathcal{B}^2$ e $z > 0$, é

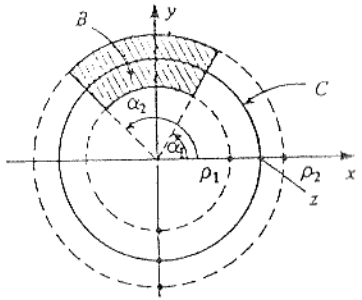
$$P((X, Y) \in B \mid \sqrt{X^2 + Y^2} = z) = \frac{\text{comprimento } (B \cap C)}{2\pi z},$$

onde C é a circunferência do círculo:



Verifiquemos a Definição 4.2, obtendo, como limite, a probabilidade condicional de certos retângulos. Por conveniência, usaremos coordenadas polares: sejam $\theta = \theta(x, y)$ e $\rho = \rho(x, y)$ as coordenadas polares do ponto (x, y), e seja B o "retângulo" polar $\{(x, y): \alpha_1 < \theta < \alpha_2, \rho_1 < \rho < \rho_2\}$, onde $\alpha_1, \alpha_2, \rho_1, \rho_2$ são racionais tais que $0 \leq \alpha_1 < \alpha_2 \leq 2\pi$ e $0 < r_1 < z < r_2$. Então B é a seguinte região aberta:





B = região hachurada aberta

Seja $Z = \sqrt{X^2 + Y^2}$ e consideremos $P((X, Y) \in B \mid Z \in I)$, onde I é um intervalo contendo z, de comprimento Δz. Quando Δz é suficientemente pequeno, de modo que $I \subset (\rho_1, \rho_2)$, temos

$$P((X, Y) \in B \mid Z \in I) = \frac{P((X, Y) \in B, Z \in I)}{P(Z \in I)} =$$

$$= \frac{\iint\limits_{B \cap [\rho \in I]} f(x, y)dxdy}{\iint\limits_{[\rho \in I]} f(x, y)dxdy},$$

onde $[\rho \in I] = \{(x, y): \sqrt{x^2 + y^2} \in I\}$.

Pela simetria circular da densidade conjunta (se quiser, poderá fazer uma mudança de variáveis, passando para coordenadas polares),

$$P((X, Y) \in B \mid Z \in I) = \frac{\alpha_2 - \alpha_1}{2\pi},$$

quando $I \subset (\rho_1, \rho_2)$. Portanto,

$$P((X, Y) \in B \mid Z = z) = \lim_{\Delta z \to 0} P((X, Y) \in B \mid Z \in I) = \frac{\alpha_2 - \alpha_1}{2\pi}.$$

Fazendo com que B diminua para o arco da circunferência, com α_1 e α_2 fixos, vemos que a distribuição condicional concentra-se na circunferência e é uniforme nela.

A distribuição condicional dado que $Z = 0$ é arbitrária, já que $P(\sqrt{X^2 + Y^2} = 0) = 0$. Costuma-se adotar a definição intuitiva

$$P((X, Y) = (0, 0) \mid \sqrt{X^2 + Y^2} = 0) = 1,$$

pois $\sqrt{X^2 + Y^2} = 0$ se, e somente se, $(X, Y) = (0, 0)$.

EXEMPLO 11. Consideremos agora a distribuição condicional de $\underset{\sim}{X} = (X_1, \ldots, X_n)$ dadas as estatísticas de ordem (veja o exemplo 22 do Capítulo 2). Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas, com $\underset{\sim}{X} = (X_1, \ldots, X_n)$ e $\underset{\sim}{Y} = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$, o vetor das estatísticas de ordem. Suponha que a função de distribuição F_{X_1} seja contínua. Queremos a distribuição condicional de $\underset{\sim}{X}$ dado $\underset{\sim}{Y}$. Para tanto, consideremos $\underset{\sim}{y} = (y_1, \ldots, y_n)$, onde $y_1 < y_2 < \ldots < y_n$.

É evidente que $\underset{\sim}{Y} = \underset{\sim}{y}$ se, e somente se, $\underset{\sim}{X}$ é uma permutação de $\underset{\sim}{y}$. Apliquemos o princípio da preservação de chances relativas: toda permutação de $\underset{\sim}{y}$ tinha, antes da observação das estatísticas de ordem, a mesma chance relativa de ser o valor de $\underset{\sim}{X}$, pois as X_i eram independentes e identicamente distribuídas. Por exemplo, suponha que $n = 2$ e que o valor observado de $\underset{\sim}{Y}$ foi $\underset{\sim}{Y} = (y_1, y_2)$, onde $y_1 < y_2$. Então ou $\underset{\sim}{X} = (y_1, y_2)$ ou $\underset{\sim}{X} = (y_2, y_1)$ e os dois valores tinham *a priori*, i. e., anterior à observação das Y_i, a mesma chance de serem escolhidos (pois $(X_1, X_2) \sim (X_2, X_1)$, ou seja, os dois vetores possuem a mesma distribuição). Logo é natural pensar que

$$P(\underset{\sim}{X} = (y_1, y_2) \mid \underset{\sim}{Y} = (y_1, y_2)) = P(\underset{\sim}{X} = (y_2, y_1) \mid \underset{\sim}{Y} = (y_1, y_2)) = \frac{1}{2}.$$

Então, nosso candidato para n geral será

$$P(\underset{\sim}{X} = (y_{\pi_1}, \ldots, y_{\pi_n}) \mid \underset{\sim}{Y} = (y_1, \ldots, y_n)) = \frac{1}{n!}, \; y_1 < y_2 < \ldots < y_n,$$

onde $(\pi_1, \ldots, \pi_n)$ é uma permutação de $(1, \ldots, n)$.

(*Observação.* Os outros valores de $\underset{\sim}{Y}$ têm probabilidade zero, pois a probabilidade de pelo menos um empate é zero:

$$P(\bigcup_{i<j} [X_i = X_j]) \leq \sum_{i<j} P(X_i = X_j) \leq \sum_{i<j} P(F(X_i) = F(X_j)) = 0,$$

onde a última igualdade decorre do fato de $F(X_1), \ldots, F(X_n)$ serem independentes e $U[0, 1]$ – veja o exercício 37 do Capítulo 2.)

Verifiquemos a Definição 4.2. Para tanto, seja $A_n = \{(x_1 \ldots, x_n) \in \mathbb{R}^n : x_1 < x_2 < \ldots < x_n\}$ e seja $B \subset A_n$, B retângulo aberto racional contendo $\underset{\sim}{y}$ ("racional" quer dizer que todos os lados têm extremos racionais). Suponha que I seja um pequeno retângulo contendo $\underset{\sim}{y}$, de modo que $\underset{\sim}{y} \in I \subset B \subset A_n$.

Para toda permutação $\pi = (\pi_1, \ldots, \pi_n)$, definamos $\underset{\sim}{X}_\pi = (X_{\pi_1}, \ldots, X_{\pi_n})$. É fácil de ver que $\underset{\sim}{X}$ e $\underset{\sim}{X}_\pi$ possuem a mesma distribuição, para toda permutação π. Então, os $n!$ eventos $[\underset{\sim}{X}_\pi \in I]$ são equiprováveis e disjuntos 2 a 2 em relação a π. (Por exemplo, se $\underset{\sim}{X}(\omega) \in I$, então $X_1(\omega) < X_2(\omega) < \ldots < X_n(\omega)$ e $\underset{\sim}{X}_\pi(\omega) \notin I$ para $\pi \neq (1, 2, \ldots, n)$.) Mas $\underset{\sim}{Y} \in I$ se, e somente se, $\underset{\sim}{X}_\pi \in I$ para *algum* π, logo os eventos $[\underset{\sim}{X}_\pi \in I]$ formam uma partição de $[\underset{\sim}{Y} \in I]$ e

$$P(\underset{\sim}{Y} \in I) = \sum_{\pi \text{ permutação}} P(\underset{\sim}{X}_\pi \in I) = n!\, P(\underset{\sim}{X}_\pi \in I), \; \forall \pi.$$

Portanto, para toda permutação π,

$$P(\underset{\sim}{X}_\pi \in B \mid \underset{\sim}{Y} \in I) = \frac{P(\underset{\sim}{X}_\pi \in B, \underset{\sim}{Y} \in I)}{P(\underset{\sim}{Y} \in I)} = \frac{P(\underset{\sim}{X}_\pi \in I)}{P(\underset{\sim}{Y} \in I)} = \frac{1}{n!},$$

onde usamos o fato de que $[\underset{\sim}{X}_\pi \in B, \underset{\sim}{Y} \in I] = [\underset{\sim}{X}_\pi \in I, \underset{\sim}{Y} \in I] = [\underset{\sim}{X}_\pi \in I]$. Mas quando $\underset{\sim}{X}_\pi \in B$, π é a permutação que coloca as coordenadas de $\underset{\sim}{X}$ em ordem. Neste caso, $\underset{\sim}{X} \in \pi^{-1}(B)$, onde π^{-1} é a permutação inversa de π, ou seja, a permutação que recoloca os valores $(\pi_1, \ldots, \pi_n)$ na ordem natural $(1, \ldots, n)$.

Logo $[\underset{\sim}{X}_\pi \in B] = [\underset{\sim}{X} \in \pi^{-1}(B)]$ e temos

$$P(\underset{\sim}{X} \in \pi^{-1}(B) \mid \underset{\sim}{Y} \in I) = P(\underset{\sim}{X}_\pi \in B \mid \underset{\sim}{Y} \in I) = \frac{1}{n!}, \; \forall \pi.$$

Fazendo B diminuir para $\underset{\sim}{y}$, de modo que $\pi^{-1}(B)$ diminui para $\pi^{-1}(\underset{\sim}{y})$, temos

$$P(\underset{\sim}{X} = \pi^{-1}(\underset{\sim}{y}) \mid \underset{\sim}{Y} = \underset{\sim}{y}) = \frac{1}{n!}, \; \forall \pi.$$

Como a classe das permutações inversas é igual à das permutações, o nosso candidato está verificado.

§4.5 Esperança condicional

DEFINIÇÃO 4.3. 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,

$$E(X \mid Y = y) = \int x \, dF_X(x \mid Y = y).$$

TEOREMA 4.3. *Se X é integrável, então $E(X \mid Y = y)$ existe e é finita quase certamente, i. e., existe um boreliano B_0 tal que $P(Y \in B_0) = 1$ e $E(X \mid Y = y)$ é finita para $y \in B_0$.*

Prova. Teoria da Medida (decorre do Teorema de Radon-Nikodym). □

Se definimos $\varphi(y) = E(X \mid Y = y)$, a variável aleatória $\varphi(Y) = E(X \mid Y)$ chama-se *esperança condicional de X dada Y*. (Sem perda de generalidade, φ é mensurável no caso de X integrável.)

A esperança condicional, sendo a esperança da distribuição condicional, possui (condicionalmente) todas as propriedades da esperança ordinária já enunciadas, mais a propriedade importante de que $E\{E(X \mid Y)\} = EX$, ou equivalentemente,

$$EX = \int E(X \mid Y = y) dF_Y(y).$$

Esta equação já foi verificada no caso de Y discreta (fórmulas (4.1)). Vamos verificá-la quando X e Y têm densidade conjunta $f(x, y)$:

A densidade condicional é $f(x|y) = f(x, y)/f_Y(y)$, e segue-se que

$$E(X \mid Y = y) = \int x dF_X(x \mid Y = y) = \int_{-\infty}^{\infty} x f(x|y) dx = \int_{-\infty}^{\infty} x \frac{f(x, y)}{f_Y(y)} dx,$$

se $f_Y(y) > 0$. Logo, quando X é integrável,

$$E\{E(X \mid Y)\} = \int E(X \mid Y = y) dF_Y(y) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x \frac{f(x, y)}{f_Y(y)} dx \right) f_Y(y) dy =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx dy = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(x, y) dy \right) x dx =$$

$$= \int_{-\infty}^{\infty} x f_X(x) dx = EX.$$

Enunciaremos a seguir, sem provas, algumas importantes *propriedades da esperança condicional*. Seja Y uma variável aleatória qualquer, e sejam X, X_1, X_2, etc., variáveis aleatórias *integráveis*, todas definidas no mesmo espaço de probabilidade.

EC1. *Propriedade básica.* $E\{E(X \mid Y)\} = EX$.

Propriedades de $E(X \mid Y)$ análogas às propriedades da esperança ordinária. *Convenção*: não distinguiremos variáveis aleatórias que sejam iguais quase certamente. Por exemplo, em EC2 basta que $X = c$ quase certamente, i.e., $P(X = c) = 1$. Além disso, os resultados enunciados são necessariamente válidos apenas quase certamente.

EC2. *Se $X = c$, para alguma constante c, então* $E(X \mid Y) = c$ (veja o exercício 26).

EC3. *Se $X_1 \leq X_2$, então* $E(X_1 \mid Y) \leq E(X_2 \mid Y)$.

Casos particulares: (i) Se $X \geq 0$, então $E(X \mid Y) \geq 0$.
(ii) Se $a \leq X \leq b$, então $a \leq E(X \mid Y) \leq b$.
(iii) $E(X \mid Y) \leq E(|X| \mid Y)$ e, por EC4, $|E(X \mid Y)| \leq E(|X| \mid Y)$.

EC4. *Linearidade.* $E(aX_1 + bX_2 \mid Y) = aE(X_1 \mid Y) + bE(X_2 \mid Y)$.

EC5. *Desigualdade de Jensen. Seja φ uma função convexa. Então* $\varphi\{E(X \mid Y)\} \leq E\{\varphi(X) \mid Y\}$.

EC6. *Teorema da Convergência Monótona. Se $X_n \geq 0$ e $X_n \uparrow X$, então* $E(X_n \mid Y) \uparrow E(X \mid Y)$.

EC7. *Teorema da Convergência Dominada. Se $X_n \to X$ e se existe X_0 integrável tal que $|X_n| \leq X_0$, então*

$$\lim_{n \to \infty} E(X_n \mid Y) = E(X \mid Y).$$

Antes de vermos exemplos, notemos que o princípio da substituição tem a seguinte forma para esperança condicional.

PROPOSIÇÃO 4.2. *(Princípio da substituição para esperança condicional). Se $\varphi(X, Y)$ é integrável, então*

$$E\{\varphi(X, Y) \mid Y = y\} = E\{\varphi(X, y) \mid Y = y\} = \int \varphi(x, y) dF_X(x \mid Y = y).$$

Prova. Pela Proposição 4.1, a distribuição condicional de $\varphi(X, Y)$ dado que $Y = y$, é a distribuição condicional de $\varphi(X, y)$ dado

que $Y = y$, que é interpretada como a distribuição de $\varphi(X, y)$, onde X tem a sua distribuição condicional dado que $Y = y$. □

Observação. O princípio da substituição vale também no caso de φ ser função de vetores aleatórios $\underset{\sim}{X}$ e $\underset{\sim}{Y}$.

EXEMPLOS (12). Como a distribuição condicional de $X = g(Y)$, dado que $Y = y$, é massa pontual em $g(y)$, temos $E\{g(Y) \mid Y = y\} = g(y)$, i. e., $E\{g(Y) \mid Y\} = g(Y)$.

(13). Se X e Y são independentes, então $P(X \in B \mid Y = y) = P(X \in B)$, de modo que $E(X \mid Y = y) = EX$ e $E(X \mid Y) = EX$.

(14). Se X é integrável, a desigualdade de Jensen implica que $E^2(X \mid Y) \overset{def}{=} \{E(X \mid Y)\}^2 \leq E(X^2 \mid Y)$. A diferença entre as duas, $E(X^2 \mid Y) - E^2(X \mid Y)$, é chamada *variância condicional* de X dada Y. (Veja o exercício 36(*b*) para o resultado: variância de X é igual à soma da esperança da variância condicional e a variância da esperança condicional.)

(15). Se X e Y são independentes e identicamente distribuídas, com $X \sim N(0, \sigma^2)$, então já vimos no exemplo 10 que a distribuição de (X, Y) dado que $\sqrt{X^2 + Y^2} = z$, era uniforme na circunferência do círculo de raio z. Pelo princípio da substituição, a distribuição condicional de X dado que $\sqrt{X^2 + Y^2} = z$, é simétrica em torno de zero, pois tem a distribuição da primeira coordenada de um vetor distribuído uniformemente na circunferência. Portanto, $E(X \mid \sqrt{X^2 + Y^2}) = 0$ e, analogamente, $E(Y \mid \sqrt{X^2 + Y^2}) = 0$.

(16). Sejam X e Y independentes e identicamente distribuídas, com $X \sim U[0, 1]$, e sejam $U = \min(X, Y)$ e $V = \max(X, Y)$. Qual a esperança condicional de U dada V?

Neste caso, U e V são as estatísticas de ordem da amostra de tamanho 2. Já sabemos que (U, V) possui distribuição uniforme em $A_2 = \{(u, v): 0 \leq u < v \leq 1\}$. Pelo princípio da preservação de chances relativas, ou simplesmente pelo fato da densidade condicional de U dado que $V = v$ ser constante em $[0, v]$, a distribuição condicional de U é uniforme em $[0, v]$ (veja o exercício 17). Portanto,

$$E(U \mid V = v) = \frac{v}{2} \quad \text{e} \quad E(U \mid V) = \frac{V}{2}.$$

(17). Sejam $X_1, \ldots, X_n$ independentes, identicamente distribuídas e integráveis, e seja $S = X_1 + \ldots + X_n$.

Problema: demonstre que para todo $i = 1, \ldots, n$, $E(X_i \mid S) = \dfrac{S}{n}$.

Prova. Os vetores $(X_1, \ldots, X_n)$ e $(X_i, X_2, \ldots, X_{i-1}, X_1, X_{i+1}, \ldots, X_n)$ têm a mesma distribuição (por quê?). Por isso, (X_1, S) e (X_i, S) possuem a mesma distribuição (use o seguinte fato: se $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ são vetores aleatórios identicamente distribuídos, então $f(\underset{\sim}{X})$ e $f(\underset{\sim}{Y})$ também têm a mesma distribuição, onde a função f pode ser um vetor de funções reais).

Como a distribuição condicional é determinada pela distribuição conjunta, segue-se que X_1 e X_i têm a mesma distribuição condicional dado que $S = s$, i. e.,

$$P(X_1 \in B \mid S = s) = P(X_i \in B \mid S = s), \quad i = 1, 2, \ldots, n.$$

Logo $E(X_1 \mid S = s)$ e $E(X_i \mid S = s)$, sendo esperanças da mesma distribuição, são iguais. Portanto,

$$E(X_1 \mid S = s) = E(X_2 \mid S = s) = \ldots = E(X_n \mid S = s),$$

e

$$nE(X_i \mid S = s) = \sum_{i=1}^{n} E(X_i \mid S = s) = (\text{por linearidade}) =$$

$$= E\left(\sum_{i=1}^{n} X_i \mid S = s\right) = E(S \mid S = s) = s.$$

Concluímos que $E(X_i \mid S = s) = \dfrac{s}{n}$, ou seja, $E(X_i \mid S) = \dfrac{S}{n}$. □

Decorre do princípio da substituição que se X é integrável e $Xg(Y)$ também o é, então

$$E(Xg(Y) \mid Y = y) = g(y)\, E(X \mid Y = y).$$

Isso vale porque dado que $Y = y$, $g(Y)$ é igual à constante $g(y)$ e, por EC4, uma constante pode ser levada para fora da esperança. Para uma prova formal, notemos que

$$E(Xg(Y) \mid Y = y) = E(Xg(y) \mid Y = y) = \int xg(y)\, dF_X(x) \mid Y = y) =$$

$$= g(y) \int x\, dF_X(x \mid Y = y) = g(y)\, E(X \mid Y = y).$$

Usaremos este resultado no seguinte exemplo.

(18) Seja (X, Y) de distribuição normal bivariada com parâmetros $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$. Vimos no §4.2 que

$$X \mid Y = y \sim N\left(\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(y - \mu_2), \sigma_1^2(1 - \rho^2)\right).$$

Portanto,

$$E(X \mid Y = y) = \mu_1 + \frac{\rho\sigma_1}{\sigma_2}(y - \mu_2)$$

ou, equivalentemente, $E(X \mid Y) = \mu_1 + \frac{\rho\sigma_1}{\sigma_2}(Y - \mu_2)$.

Vamos calcular $\text{Cov}(X, Y) = EXY - \mu_1\mu_2$ sem mexer com a densidade conjunta no cálculo de EXY. Supondo que XY seja integrável (veremos isto no final deste exemplo), temos

$$E(XY \mid Y = y) = yE(X \mid Y = y) = \mu_1 y + \frac{\rho\sigma_1}{\sigma_2}(y^2 - \mu_2 y),$$

logo

$$EXY = E\{E(XY \mid Y)\} = E\{\mu_1 Y + \frac{\rho\sigma_1}{\sigma_2}(Y^2 - \mu_2 Y)\} =$$

$$= \mu_1\mu_2 + \frac{\rho\sigma_1}{\sigma_2}(\sigma_2^2 + \mu_2^2 - \mu_2^2) = \mu_1\mu_2 + \rho\sigma_1\sigma_2.$$

Portanto, $\text{Cov}(X, Y) = \rho\sigma_1\sigma_2$ e $\rho(X, Y) = \frac{1}{\sigma_1\sigma_2}\text{Cov}(X, Y) = \rho$, demonstrando assim que o parâmetro ρ é o coeficiente de correlação entre X e Y.

A integrabilidade de XY decorre da *desigualdade de Cauchy-Schwarz*, cuja formulação para variáveis aleatórias diz que $E|XY| \leq \sqrt{EX^2 \cdot EY^2}$. Em outras palavras, se X e Y têm variâncias finitas, então XY é integrável. Neste exemplo, X e Y são normais, logo XY é integrável. Notemos que X e Y têm momentos finitos de toda ordem (exercício 20(*c*) do Cap. 3), uma implicação disso sendo que todos os momentos "mistos", ou momentos-produto, são finitos: $EX^k Y^m$ é finito para todo $k, m \geq 1$ (pois $E|X^k Y^m| \leq \sqrt{EX^{2k} \cdot EY^{2m}} < \infty$).

Observação. A prova da desigualdade de Cauchy-Schwarz é quase idêntica à da Proposição 3.7, sobre o coeficiente de correlação, que é realmente conseqüência de Cauchy-Schwarz. *Provemos Cauchy-Schwarz*, i. e., provemos que $E|XY| \leq \sqrt{EX^2 \cdot EY^2}$ para qualquer par de variáveis aleatórias definidas no mesmo espaço de probabilidade:

Sejam $EX^2 = a$, $EY^2 = b$. Se $0 < a < \infty$ e $0 < b < \infty$, então

$$0 \leq E\left(\frac{|X|}{\sqrt{a}} - \frac{|Y|}{\sqrt{b}}\right)^2 = E\left(\frac{X^2}{a} - \frac{2|XY|}{\sqrt{ab}} + \frac{Y^2}{b}\right) = 2 - \frac{2}{\sqrt{ab}}E|XY|.$$

Logo

$$E|XY| \leq \sqrt{ab} = \sqrt{EX^2 \cdot EY^2}.$$

Se $a = 0$, então $P(X^2 = 0) = 1$, $P(X = 0) = 1$ e $P(XY = 0) = 1$, logo $E|XY| = 0$ e $E|XY| \leq \sqrt{ab} = 0$ (convenção: $0 \cdot \infty = 0$). Se $b = 0$, vale a mesma coisa. Se $a = +\infty$ e $b > 0$, a desigualdade é trivial: $E|XY| \leq +\infty$. Idem se $a > 0$ e $b = +\infty$. □

Sejam, agora, Y uma variável aleatória e A um evento aleatório. Como $A = [I_A = 1]$ e $P(A|B) = P(I_A = 1 \mid B)$ para B evento aleatório, é bem natural definir

$$P(A \mid Y = y) = P(I_A = 1 \mid Y = y) = \lim_{\Delta y \to 0} P(I_A = 1 \mid Y \in I) =$$

$$\lim_{\Delta y \to 0} P(A \mid Y \in I).$$

Como a variável aleatória I_A assume somente os valores 0 e 1, temos, pelo princípio da preservação de chances relativas,

$$P(I_A = 1 \mid Y = y) + P(I_A = 0 \mid Y = y) = 1.$$

(*Exercício.* Verifique, através da Definição 4.2 (Relação 1), que se $P(X \in B) = 1$, então $P(X \in B \mid Y = y) = 1$.)

Portanto,

$$E(I_A \mid Y = y) = 1 \cdot P(I_A = 1 \mid Y = y) + 0 \cdot P(I_A = 0 \mid Y = y)$$
$$= P(I_A = 1 \mid Y = y) = P(A \mid Y = y).$$

Por nossa definição, então, *a probabilidade condicional de um evento é a esperança condicional de seu indicador*: $P(A \mid Y) = E(I_A \mid Y)$. É analogia direta com o fato de que $P(A) = EI_A$. De fato, como I_A é integrável temos

$$P(A) = EI_A = E\{E(I_A \mid Y)\} = E\{P(A \mid Y)\},$$

ou seja, *a probabilidade de um evento é a esperança de sua probabilidade condicional dada Y*, para qualquer Y.

EXEMPLO 19. Sejam X e Y variáveis aleatórias. Calculemos a distribuição de $Z = X + Y$. Temos

$$P(X + Y \leq z) = E\{P(X + Y \leq z \mid Y)\} = \int P(X + Y \leq z \mid Y = y)\, dF_Y(y) =$$
$$= \text{(pelo princípio da substituição)} = \int P(X \leq z - y \mid Y = y)\, dF_Y(y) =$$
$$= \int F_X(z - y \mid Y = y)\, dF_Y(y).$$

Se X e Y são independentes, então $F_Y(z-y \mid Y=y) = F_X(z-y)$ e temos

$$F_Z(z) = P(X+Y \leq z) = \int F_X(z-y)\, dF_Y(y) = \text{(trocando } X \text{ por } Y) = \int F_Y(z-x)\, dF_X(x).$$

Esta distribuição é a *convolução* das distribuições de X e Y e, como no caso da convolução de densidades, escrevemos $F_Z = F_Y * F_X = F_X * F_Y$.

EXERCÍCIOS DO CAPÍTULO 4

§4.1

1. Sejam X_1 e X_2 variáveis aleatórias independentes, cada uma com distribuição geométrica definida por

   $$P(X_i = n) = p(1-p)^n, \quad n = 0, 1, 2, \ldots; \quad i = 1, 2;$$

   onde $0 < p < 1$. (Observação: esta versão da distribuição geométrica corresponde à distribuição do número de fracassos antes do primeiro sucesso em uma seqüência de ensaios de Bernoulli.)
   (a) Calcule $P(X_1 = X_2)$ e $P(X_1 < X_2)$.
   (b) Determine a distribuição condicional de X_1 dada $X_1 + X_2$.

2. Uma certa lâmpada tem uma vida, em *horas*, tendo distribuição exponencial de parâmetro 1. Um jogador acende a lâmpada e, enquanto a lâmpada ainda estiver acesa, lança um dado equilibrado de quinze em quinze *segundos*. Qual o número esperado de 3's lançados pelo jogador até a lâmpada se apagar?

3. Partículas chegam em um contador segundo um processo de Poisson com parâmetro λ. Em um determinado tempo t, produz-se uma voltagem, multiplicando o número de partículas que já entraram no contador por um fator que é independente desse número e que tem densidade:

   $$f(x) = \begin{cases} \dfrac{1}{(1+x)^2}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

   Ache a probabilidade da voltagem produzida ser menor que 1.

4. Mostre que se X é uma variável aleatória tendo distribuição simétrica em torno de zero, e se $P(X=0) = 0$, então a distribuição condicional de X^2 dado que $X > 0$ é igual à própria distribuição de X^2.

5. Partículas radioativas chegam a um contador segundo um processo de Poisson com uma taxa média de três por segundo, mas o contador registra somente cada segunda partícula (i.e., são registradas somente as partículas n.º 2, 4, 6, ...).
   (a) Seja X_t o número de partículas registradas até o tempo t. É $\{X_t : t \geq 0\}$ um processo de Poisson? Se for, qual o parâmetro? Se não for, explique o porquê.
   (b) Supondo que o contador registrou exatamente uma partícula durante o primeiro segundo, qual a probabilidade de que ele não registre mais partícula alguma antes do tempo 2?

6. Um contador recebe impulsos de duas fontes independentes, A e B. Fonte A gera impulsos conforme um processo de Poisson com parâmetro $\lambda > 0$, enquanto a fonte B gera impulsos segundo um processo de Poisson com parâmetro $\xi > 0$. Suponha que o contador registre todo impulso gerado pelas duas fontes.
   (a) Seja X_t o número de impulsos registrados pelo contador até o tempo t, $t > 0$ ($X_0 = 0$). Explique porque $\{X_t : t \geq 0\}$ é um processo de Poisson (basta uma explicação intuitiva). Qual o parâmetro?
   (b) Qual a probabilidade de que o primeiro impulso registrado seja da fonte A?
   (c) Dado que exatamente 100 impulsos foram contados durante a primeira unidade de tempo, qual a distribuição que você atribuiria ao número emitido pela fonte A?

7. Diz-se que $\underset{\sim}{X} = (X_1, \ldots, X_k)$ tem *distribuição multinomial* com parâmetros $p_1, \ldots, p_k$ e n, onde $p_i \geq 0$ e $\sum_{i=1}^{k} p_i = 1$, se

   $$P(\underset{\sim}{X} = (j_1, \ldots, j_k)) = \frac{n!}{j_1! j_2! \ldots j_k!}\, p_1^{j_1} p_2^{j_2} \ldots p_k^{j_k},$$

   para toda escolha de $j_1, \ldots, j_k$ inteiros não-negativos tais que $\sum_{\ell=1}^{k} j_\ell = n$.

(a) Prove que $X_i \sim b(n, p_i)$, $i = 1, \ldots, k$.

(b) Sejam $0 < s_1 < s_2 < \ldots < s_k = t$. Mostre que no processo de Poisson, dado que $X_t = n$, a distribuição condicional de $(X_{s_1}, X_{s_2} - X_{s_1}, \ldots, X_t - X_{s_{k-1}})$ é multinomial com parâmetros $\frac{s_1}{t}, \frac{s_2 - s_1}{t}, \ldots, 1 - \frac{s_{k-1}}{t}$ e n. (Note que essa distribuição não depende do parâmetro λ do processo.)

8. Uma exposição funciona pelo periodo de T horas. Visitantes chegam à exposição segundo um processo de Poisson com taxa λ visitantes/hora. Os visitantes permanecem na exposição até o fim do periodo. Calcule o tempo médio total gasto pelos visitantes na exposição. (*Sugestão.* Dado que chegou um só visitante durante as T horas, qual a média do tempo que ele permanece na exposição?)

9. Suponha que o número de passas num bolo inglês tenha distribuição de Poisson de parâmetro 60. Um jogador compra um bolo, tira todas as passas uma por uma e reparte as passas entre ele e você da seguinte maneira: depois da extração de cada passa ele joga uma moeda equilibrada, dando a passa para você se der cara, comendo ele mesmo a passa se der coroa. Qual a distribuição de número de passas que você recebe? A esperança?

10. Sejam X e Y independentes tais que $X \sim b(m, p)$ e $Y \sim b(n, p)$. Obtenha a distribuição condicional de X dada $X + Y$. Como se chama essa distribuição?

11. Duas fontes radioativas, I e II, emitem partículas (independentemente) segundo processos de Poisson com, respectivamente, parâmetros λ e ξ. Seja Z_t o número total de partículas emitidas até o instante t, para $t > 0$. Dado que $Z_t = k$, onde $k > 0$, qual a probabilidade condicional da última partícula emitida antes do instante t ter sido da fonte I? (A resposta é igual à do exercício 6(b). Um possível método de verificação: use o exercício 16(e) do Capítulo 1, com A_n o evento "n partículas emitidas até o instante t pela fonte I".)

12. Considere um processo de Poisson com parâmetro $\lambda > 0$.

(a) Para $t > 0$ fixo, seja Z_t o tempo transcorrido até o instante t desde a ocorrência ("chegada") imediatamente anterior. ($Z_t = t$ se não houve nenhuma chegada em $(0, t]$.) Calcule a distribuição de Z_t. (Note que essa distribuição é aproximadamente exponencial quando t é grande).

(b) Se T_{n+1} é o tempo transcorrido entre a n-ésima chegada e a chegada número $n + 1$, qual a distribuição de T_{n+1}? Determine a distribuição de W_t = tempo que transcorre entre o instante t e a próxima chegada.

(c) Mostre que $Z_t + W_t$ o tempo entre as chegadas que "cercam" o instante t, é estocasticamente estritamente maior que T_{n+1}, i.e., $P(Z_t + W_t \leq x) < P(T_{n+1} \leq x)$ $\forall x > 0$. (Esse é o "paradoxo do tempo de espera".)

13. Seja $X_1, X_2, \ldots$ uma seqüência de variáveis aleatórias independentes e identicamente distribuídas tendo distribuição exponencial com média $\frac{1}{\lambda}$, onde $\lambda > 0$. Para $t > 0$ fixo, seja $N = \max\{n \geq 0 : S_n \leq t\}$, onde $S_0 = 0$ e $S_n = \sum_{i=1}^{n} X_i$, de modo que N é o índice da última soma parcial menor ou igual a t. Mostre que N tem distribuição de Poisson com média λt. (*Sugestão.* "Mergulhe" a seqüência no processo de Poisson.)

§4.2

14. Suponha que Y possua densidade $f_Y(y)$ e que a distribuição condicional de X, dado que $Y = y$, possua densidade $f(x \mid y)$, para todo y (ou pelo menos para "quase todo" valor possível de Y). Demonstre que

$$f(x, y) = f_Y(y)\, f(x \mid y)$$

é a densidade conjunta de X e Y. (Observação: foi provado no caso III que $f(x \mid y) = \frac{f(x, y)}{f_Y(y)}$ *se* existe uma densidade conjunta, de modo que a distribuição conjunta determina $f_Y(y)$ e $f(x \mid y)$. Este exercício diz que $f_Y(y)$ e $f(x \mid y)$ determinam $f(x, y)$.)

15. Considere o seguinte experimento de duas etapas: primeiro, escolhe-se um ponto x de acordo com a distribuição uniforme em $(0, 1)$; depois, escolhe-se um ponto y de acordo com a distribuição uniforme em $(-x, x)$. Se o vetor aleatório (X, Y) representar o resultado do experimento, qual será a densidade conjunta de X e Y? A densidade marginal de Y? A densidade condicional de X dada Y? (*Sugestão.* Exercício 14.)

16. Observam-se duas lâmpadas durante suas vidas úteis. Suponha as vidas independentes e exponenciais de parâmetro λ. Sejam X o tempo de queima da primeira lâmpada a queimar e Y o tempo de queima da segunda a queimar ($X \le Y$).
    (a) Qual a distribuição condicional de X dada Y?
    (b) Qual a distribuição de Y dada X?

17. Suponha que (X, Y) possua distribuição uniforme em A, onde A é uma região de área positiva. Mostre que a distribuição condicional de X dado que $Y = y$ é uniforme em A_y, a secção de A por y, onde definimos $A_y = \{x : (x, y) \in A\}$.

18. Demonstre que se $P(X \in B \mid Y = y) = P(X \in B)$ para todo $B \in \mathcal{B}$ e $y \in \mathbb{R}$, então X e Y são independentes, de modo que X e Y são independentes se, e somente se, a distribuição condicional não depende do valor de Y. (*Sugestão.* Mostre que a função de distribuição conjunta fatora.)

§4.4

19. Seja X uma variável aleatória com densidade $f(x)$, onde f é contínua. Qual a distribuição condicional de X dada $|X|$? Verifique sua resposta. (Observação: a hipótese de continuidade de f não é necessária. Tente verificar a Relação 4.1 sem essa hipótese.)

20. Sejam X e Y independentes, cada uma com distribuição $N(0, \sigma^2)$. Qual a distribuição condicional de X dada $\sqrt{X^2 + Y^2}$? (*Sugestão.* Exemplo 10 e o princípio da substituição.) Determine $E(X \mid \sqrt{X^2 + Y^2})$.

21. Explique como as soluções nos casos I e II do §4.2 podem ser consideradas conseqüências do princípio da preservação de chances relativas.

22. Sejam X e Y o mínimo e o máximo de duas variáveis aleatórias independentes com distribuição comum $\exp(\lambda)$, onde $\lambda > 0$. Mostre de duas maneiras que $Y - X \mid X \sim \exp(\lambda)$:
    (a) A partir da densidade conjunta de X e $Y - X$.
    (b) Utilizando o princípio da substituição e o resultado do exercício 16(b).

23. Sejam X e Z variáveis aleatórias em $(\Omega, \mathcal{A}, P)$.
    (a) Mostre que para A e B borelianos,

$$P(Z \in B,\ X \in A \mid X = x) = \begin{cases} P(Z \in B \mid X = x) & \text{se } x \in A \\ 0 & \text{se } x \notin A. \end{cases}$$

    (b) Mostre que se $P(Z \in B \mid X = x) \ge \frac{1}{2}$ para todo $x \in A$, então $P(Z \in B \mid X \in A) \ge \frac{1}{2}$.

    (c) Sejam X e Y independentes tais que $P(Y \ge 0) \ge \frac{1}{2}$. Demonstre que $P(X + Y \ge a \mid X \ge a) \ge \frac{1}{2}$.

24. (a) Seja $\underset{\sim}{X} = (X_1, \ldots, X_n)$ um vetor aleatório com densidade $f(x_1, \ldots, x_n)$. Use o princípio da preservação de chances relativas para obter $f(x_1, \ldots, x_k \mid x_{k+1}, \ldots, x_n)$, a densidade condicional de $(X_1, \ldots, X_k)$ dado que $(X_{k+1}, \ldots, X_n) = (x_{k+1}, \ldots, x_n)$, onde $1 \le k \le n - 1$. (Não é preciso demonstrar formalmente.)
    (b) Sejam X_1, X_2 e X_3 independentes com distribuição comum $U[0, 1]$, com $X_{(1)}$, $X_{(2)}$ e $X_{(3)}$ as estatísticas de ordem. Determine a distribuição condicional de $X_{(2)}$ dadas $X_{(1)}$ e $X_{(3)}$.

25. Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas, com distribuição contínua F. Seja $X = \max_{1 \le i \le n} X_i$.
    (a) Mostre que para todo $k = 1, 2, \ldots, n$,

$$P(X_k \le x \mid X = t) = \begin{cases} \dfrac{(n-1)F(x)}{nF(t)}, & \text{se } x < t \\ 1, & \text{se } x \ge t. \end{cases}$$

    (*Sugestão.* $x^n - y^n = (x - y)(x^{n-1} + x^{n-2}y + \ldots + xy^{n-2} + y^{n-1})$.)
    (b) Suponha F diferenciável. Existe densidade condicional no item (a)?

§4.5

26. Prove: se X é constante quase certamente, i.e., $P(X = c) = 1$, então $P(X = c \mid Y = y) = 1$. (*Sugestão.* Use a Relação 4.1.) Deduza a propriedade EC2.

27. Sejam X e Y variáveis aleatórias tais que $EX^2 < \infty$ e $EY^2 < \infty$. Demonstre que $\mathrm{Cov}(X, Y) = \mathrm{Cov}(X, E(Y \mid X))$.

28. Suponha que X e Y possuam densidade conjunta

$$f(x, y) = \begin{cases} \frac{1}{\pi}, & \text{se } x^2 + y^2 \leq 1 \\ 0, & \text{caso contrário.} \end{cases}$$

(a) Ache a distribuição condicional de Y dada X. Calcule $E(Y \mid X)$.
(b) X e Y são independentes? Por quê?
(c) Prove que X e Y são não-correlacionadas. (*Sugestão.* Use o exercício anterior.)

29. Seja X uma variável aleatória. Cauchy-padrão.
(a) Conforme a Definição 4.3, qual é $E(X \mid\mid X \mid = y)$?
(b) Deduza que $E[E(X \mid\mid X|)] = 0$.
(c) A propriedade basica $E[E(X \mid Y)] = EX$ não vale nesse caso. Existe uma contradição nisso?

30. Se $X \geq 0$, ocorre que $EX = E[E(X \mid Y)]$ sem a hipótese de integrabilidade de X (se $EX = +\infty$, então $E[E(X \mid Y)] = +\infty$ também). Prove esse resultado quando X e Y são discretas.

31. O número de acidentes que ocorrem em certa fábrica em uma semana é uma variável aleatória com média μ e variância σ^2. Os números de indivíduos feridos nos diferentes acidentes são independentes e identicamente distribuidos com média ξ e variância τ^2, e são independentes do número de acidentes. Determine a média e a variância do número de indivíduos feridos em uma semana. (*Sugestão.* Use o exercício 30.)

32. Calcule $E(X_{(2)} \mid X_{(1)}, X_{(3)})$, onde $X_{(1)}, X_{(2)}$ e $X_{(3)}$ são as estatísticas de ordem de uma amostra aleatória da $U[0, 1]$. (Veja o exercício 24(b).)

33. Seja X uma variável aleatória com distribuição exponencial de parâmetro 1. Seja $t > 0$ fixo. Ache $E(X \mid \max(X, t))$ e $E(X \mid \min(X, t))$.

34. Mostre que se X e Y são variáveis aleatórias, então

$$F_{X,Y}(x, y) \leq \sqrt{F_X(x)\, F_Y(y)}.$$

(*Sugestão.* Use indicadores e Cauchy-Schwarz.)

35. Seja (X, Y) um vetor aleatório bidimensional. Suponha que (*i*) X tem distribuição exponencial com parâmetro $\frac{1}{2}$ e (ii) para cada $x > 0$, a distribuição condicional de Y dado que $X = x$ é $U[0, x^2]$. Em outras palavras, $X \sim \exp\left(\frac{1}{2}\right)$ e $Y \mid X = x \sim U[0, x^2]$.
(a) Qual é a distribuição de $Z = Y/X^2$?
(b) Calcule EX, EY e $\mathrm{Cov}(X, Y)$.

36. Sejam X e Y variáveis aleatórias tais que Y tem esperança finita.
(a) Mostre que a variância condicional de Y dada X, como definida no exemplo 14, é a variância da distribuição condicional.
(b) Demonstre que $\mathrm{Var}\, Y = E[\mathrm{Var}(Y \mid X)] + \mathrm{Var}\,[E(Y \mid X)]$, i.e., a variância de Y é a soma da esperança da variância condicional e a variância da esperança condicional. (Para simplificar a demonstração, pode supor que a variância de Y seja finita.)

37. Sejam X e Y variáveis aleatórias com segundos momentos finitos, e seja Z uma outra variável aleatória. Demonstre a seguinte fórmula:

$$\mathrm{Cov}(X, Y) = E\{\mathrm{Cov}((X, Y) \mid Z)\} + \mathrm{Cov}(E(X \mid Z), E(Y \mid Z)),$$

onde $\mathrm{Cov}((X, Y) \mid Z) \stackrel{def}{=} E(XY \mid Z) - E(X \mid Z)\, E(Y \mid Z)$.

38. Suponha que em um temporal, o número X de gotas de chuva que caem no IMPA durante um segundo tenha distribuição de Poisson com parâmetro $\lambda > 0$, onde λ representa a intensidade de chuva. Suponha que o parâmetro λ seja uma variável aleatória que tenha distribuição gama com parâmetros $\alpha > 0$ e $\beta = 1$, i.e., que sua densidade seja dada por

$$f(\lambda) = \begin{cases} \frac{1}{\Gamma(\alpha)} \lambda^{\alpha - 1} e^{-\lambda}, & \lambda \geq 0 \\ 0, & \lambda < 0. \end{cases}$$

(a) Mostre que

$$P(X = k) = \frac{\Gamma(k + \alpha)}{\Gamma(\alpha)\, \Gamma(k + 1)} \left(\frac{1}{2}\right)^{k + \alpha}, \quad k = 0, 1, 2, \ldots.$$

(b) Usando métodos probabilísticos, demonstre que

$$\sum_{k=1}^{\infty} \binom{k + n - 1}{n} \frac{1}{2^k} = 2^n, \text{ para } n = 1, 2, \ldots.$$

(*Sugestão.* Calcule a esperança de X de duas maneiras diferentes).

39. Calcule $E(X_k \mid X = t)$ no exercício 25.

40. Determine $E(X \mid Y)$ e $E(Y \mid X)$ no exercício 15. (Você pode determiná-las só olhando para a densidade conjunta?) Calcule $\mathrm{Cov}(X, Y)$. X e Y são independentes?

41. Seleciona-se ao acaso (i.e., conforme a distribuição uniforme) um número entre 0 e 1. Se x é o número selecionado, lança-se n vezes (independentemente) uma moeda com probabilidade x de dar "cara". Seja Y a variável aleatória que representa o número de caras lançadas.
    (a) Y é variável aleatória discreta ou contínua?
    (b) Calcule a esperança e a variância de Y.

42. (a) Demonstre que se X e Y são independentes, então

$$P(X < Y) = \int_{-\infty}^{\infty} (1 - F_Y(x))\, dF_X(x).$$

    (b) Sejam X e Y independentes, X exponencial de parâmetro λ e Y uniforme em $[0, \lambda]$, onde $\lambda > 0$. Calcule $P(X < Y)$, $P(X > Y)$ e $P(X = Y)$.

43. Fregueses entram em um supermercado conforme um processo de Poisson com parâmetro λ. Há luz no supermercado enquanto não queima um fusível instalado no tempo $t = 0$, e cuja vida útil T tem distribuição exponencial com parâmetro ξ. Qual o número esperado de fregueses que entram no supermercado enquanto há luz? (Suponha T independente do processo de Poisson.)

CAPÍTULO V

A LEI DOS GRANDES NÚMEROS

§ 5.1 Introdução às Leis Fraca e Forte dos Grandes Números

A idéia em que se baseia a Lei dos Grandes Números já foi abordada no Capítulo 3, para o caso discreto (mas a idéia é a mesma para o caso geral): consideremos certo experimento básico, com a variável aleatória X representando o valor de um característico numérico do resultado. Pensemos na realização deste experimento n vezes (n grande), de tal maneira que as realizações sejam independentes. Suponhamos que depois de cada "ensaio" do experimento registre-se o valor do característico numérico do resultado; chamemos este valor um *observado*. Em linguagem informal, usada às vezes na Estatística Aplicada, diz-se que os n observado formam "uma amostra aleatória da variável aleatória X" ou "n observações sobre a variável aleatória X". A Lei dos Grandes Números afirma que a média aritmética dos n valores observados é aproximadamente igual a EX quando n é grande; de fato, ela afirma que esta média aritmética das observações *converge*, em certo sentido, para a média EX, quando $n \to \infty$.

Tratemos agora de construir um modelo para o experimento repetido que contemplamos acima. É claro que o experimento que consideramos é um exemplo de *experimento composto*, pois consiste em realizar n experimentos sucessivos, a saber, n ensaios independentes do experimento básico. Para experimentos desta natureza, um resultado possível é uma seqüência de n resultados possíveis do experimento básico. Portanto, se Ω_0 é o espaço amostral do experimento básico, então o espaço amostral para o experimento composto é o conjunto de seqüências de extensão n de elementos de Ω_0, i.e.,

$$\Omega_n = \{(\omega_1, \ldots, \omega_n) : \omega_i \in \Omega_0,\ i = 1, \ldots, n\},$$

ou seja, Ω_n é o *espaço produto* $\Omega_0 \times \ldots \times \Omega_0 = \Omega_0^n$, o produto cartesiano de Ω_0 consigo mesmo n vezes.

Mas não é exatamente Ω_n em que estamos interessados para fazer nosso modelo, pois na realidade estamos considerando um experimento global que consiste em realizar n ensaios do experimento, para n grande, e depois *passar ao limite* para aplicar a convergência afirmada pela Lei dos Grandes Números. Por isso, o espaço amostral do experimento global consiste nas seqüências infinitas de elementos de Ω_0, i.e.,

$$\Omega = \{(\omega_1, \omega_2, \ldots): \omega_i \in \Omega_0, i = 1, 2, \ldots\}$$
$$= \Omega_0 \times \Omega_0 \times \ldots \overset{def}{=} \Omega_0^{\infty}.$$

Aqui, ω_n é o resultado do n-ésimo ensaio do experimento básico.

Não vamos completar o modelo probabilístico para o experimento global, pois utilizaria conceitos da Teoria da Medida e fugiria dos propósitos deste livro. (Poderíamos completar o modelo utilizando para $\mathscr{A}$ a σ-álgebra produto e para P, pela hipótese de independência dos ensaios, a probabilidade produto. O leitor interessado pode consultar qualquer livro mais avançado de probabilidade, tal como Feller [9], §IV.6. Para uma discussão mais elementar, no caso discreto, veja Feller [8], §V.4.

Já que vamos registrar um certo característico numérico do resultado do n-ésimo ensaio, para todo n, estaremos registrando os valores de uma seqüência de variáveis aleatórias. Com efeito, como $X(\omega_0)$ representa o valor do característico numérico do resultado do experimento básico ($\omega_0 \in \Omega_0$), então, quando o resultado da seqüência de ensaios for $\omega = (\omega_1, \omega_2, \ldots)$, os valores observados serão $X(\omega_1)$, $X(\omega_2), \ldots$. É conveniente representar por X_n o observado do n-ésimo ensaio. Assim, X_n é função do resultado ω do experimento global, com

$$X_n(\omega) = X(\omega_n),$$

e no decorrer do experimento serão registrados os valores das variáveis aleatórias $X_1, X_2, \ldots$. 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 ensaios independentes, elas são independentes. Como têm todas a mesma distribuição, são *identicamente distribuídas*. Na literatura de Probabilidade, é comum dizer que X_1, $X_2, \ldots$ são *i.i.d.* (independentes e identicamente distribuídas. Recordemos que são independentes, por definição, se $X_1, \ldots, X_n$ são independentes para todo $n \geq 2$.). Se X_1 é integrável, então todas elas o são, pois possuem a mesma distribuição, e $EX_n = EX_1 \; \forall n$. Neste caso, como a Lei dos Grandes Números diz que o valor médio observado em n ensaios converge para $EX(=EX_1)$ quando $n \to \infty$, temos a seguinte versão da Lei dos Grandes Números em termos de variáveis aleatórias:

Se $X_1, X_2, \ldots$ são independentes, identicamente distribuídas e integráveis, então

$$\frac{X_1 + \ldots + X_n}{n} \to EX_1.$$

(O tipo de convergência será visto adiante.)

EXEMPLO 1. Se você jogar uma moeda honesta n vezes, independentemente, e contar o número S_n de caras obtidas, então $\frac{S_n}{n}$, a freqüência relativa de caras, convergirá para 1/2 quando $n \to \infty$. Esta conseqüência da Lei dos Grandes Números é bastante intuitiva e todo mundo a aceita sem maiores problemas. É claro que é conseqüência imediata da definição de probabilidade como limite de freqüências relativas, mas essa definição não é a adotada por nós. Portanto, vejamos agora que este resultado realmente decorre da Lei dos Grandes Números enunciada acima.

O espaço amostral do experimento global é o conjunto de seqüências infinitas de caras e coroas:

$$\Omega = \{(\omega_1, \omega_2, \ldots): \omega_n = c \text{ ou } \omega_n = \tilde{c}, \; n = 1, 2, \ldots\}.$$

O característico numérico do resultado básico (c *ou* $\tilde{c}$) em que estamos interessados é o indicador da propriedade de ser cara, i.e., $X(c) = 1$, $X(\tilde{c}) = 0$. Portanto, com $X_n(\omega) = X(\omega_n)$, temos $X_n = 1$ se, e somente se, o n-ésimo lançamento dá cara (com $X_n = 0$ se dá coroa). Em outras palavras, a variável aleatória X_n é o indicador do evento "cara no n-ésimo lançamento".

Como a moeda é honesta, X_n tem distribuição binomial com parâmetros 1 e 1/2; $X \sim b(1, 1/2)$. Recordemos que os lançamentos da moeda formam uma seqüência de "ensaios binomiais" ou "ensaios de Bernoulli"; a variável aleatória X_n é, portanto, o indicador de "sucesso" no n-ésimo ensaio, e S_n é o número de sucessos nos primeiros n ensaios. (Foi Bernoulli quem provou a primeira Lei dos Grandes Números, justamente para o caso de ensaios binomiais. Veja o Corolário do Teorema 5.1.)

Então, $X_1, X_2, \ldots$ compõem uma seqüência de variáveis aleatórias independentes e identicamente distribuídas, com esperança comum $EX_n = 1/2$, e a Lei dos Grandes Números diz que

$$\frac{S_n}{n} = \frac{X_1 + \ldots + X_n}{n} \to EX_1 = \frac{1}{2}.$$

Observação. De um modo geral, na Teoria da Probabilidade o espaço amostral em si é de menor importância que as relações entre as variáveis aleatórias consideradas. Isto vale porque as propriedades de variáveis aleatórias são determinadas por suas distribuições, inclusive distribuições conjuntas, independentemente do espaço amostral no qual as variáveis são definidas. No exemplo 1, o importante é que $X_1, X_2, \ldots$ são independentes e identicamente distribuídas com distribuição comum $b(1, 1/2)$. Poderíamos ter chegado imediatamente a essa seqüência sem ter passado pelo espaço Ω, através do seguinte argumento:

Se X_n é o indicador de sucesso (cara) no n-ésimo ensaio, então decorre diretamente das hipóteses (a moeda é honesta e os lançamentos independentes) que as X_n são independentes e identicamente distribuídas, com $X_n \sim b(1, 1/2)$.

A questão óbvia agora é a seguinte: de que tipo é a convergência afirmada pela Lei dos Grandes Números? A resposta que será dada neste capítulo: convergência em probabilidade (a Lei Fraca de Khintchin) e convergência quase certa (a Lei Forte de Kolmogorov). Consideremos, portanto, as definições dos dois tipos de convergência.

Sejam $Y, Y_1, Y_2, \ldots$ variáveis aleatórias definidas em um mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$.

DEFINIÇÃO 5.1. Y_n converge para Y *em probabilidade* se para todo $\varepsilon > 0$,

$$P(|Y_n - Y| \geq \varepsilon) \to 0 \text{ quando } n \to \infty.$$

Notação. $Y_n \xrightarrow{P} Y$.

DEFINIÇÃO 5.2. Y_n converge para Y *quase certamente* se $P(Y_n \to Y \text{ quando } n \to \infty) = 1$, i.e., se o evento $A_0 = \{\omega: Y_n(\omega) \to Y(\omega)\}$ é de probabilidade 1.

Notemos que convergência quase certa é convergência pontual, com probabilidade 1 – costuma-se dizer que $Y_n(\omega)$ converge para $Y(\omega)$ para "quase todo" ω. Interpretando $\omega \in \Omega$ como um resultado possível de um experimento, a seqüência $Y_n(\omega)$ de característicos numéricos de ω converge para $Y(\omega)$ para quase todo resultado ω do experimento, quando $Y_n \to Y$ quase certamente.

No exemplo 1 (lançamentos de uma moeda honesta), as variáveis aleatórias $Y_n = \frac{S_n}{n}$ formam uma seqüência de característicos numéricos do resultado do experimento, pois se $\omega = (\omega_1, \omega_2, \ldots)$ é uma seqüência de caras e coroas (um resultado possível), então $Y_n(\omega) = 1/n \times$ (número de c's entre $\omega_1, \ldots, \omega_n$). Será conseqüência da Lei Forte dos Grandes Números que $Y_n \to 1/2$ quase certamente.

Por outro lado, convergência em probabilidade não diz respeito à convergência pontual – apenas afirma que para valores grandes de n as variáveis Y_n e Y são aproximadamente iguais com probabilidade bem alta. Convergência em probabilidade é mais fraca que convergência quase certa, já que

convergência quase certa $\Rightarrow$ convergência em probabilidade,
convergência em probabilidade $\not\Rightarrow$ convergência quase certa.

PROPOSIÇÃO 5.1. *Se $Y_n \to Y$ quase certamente, então $Y_n \xrightarrow{P} Y$.*

Prova. Suponha que $Y_n \to Y$ quase certamente e seja $\varepsilon > 0$ fixo. Precisamos provar que

$$P(|Y_n - Y| \geq \varepsilon) \to 0.$$

Seja $A_0 = \{\omega: Y_n(\omega) \to Y(\omega)\}$. Por hipótese, $P(A_0) = 1$. Para todo $\omega \in A_0$, $|Y_n(\omega) - Y(\omega)| < \varepsilon$ para todo n suficientemente grande. Seja A_n o evento "para todo $k \geq n$, $|Y_k - Y| < \varepsilon$", i.e.,

$$A_n = \bigcap_{k=n}^{\infty} [|Y_k - Y| < \varepsilon].$$

Se $\omega \in A_0$, então $\omega \in A_n$ para algum n. Mas $A_n \subset A_{n+1}$, logo

$$A_0 \subset \bigcup_{n \geq 1} A_n = \lim_{n \to \infty} A_n.$$

Portanto, $1 = P(A_0) \leq P(\bigcup_{n \geq 1} A_n)$ e, por continuidade de probabilidade, $P(A_n) \uparrow 1$.

Mas $A_n \subset [|Y_n - Y| < \varepsilon]$, logo $P(|Y_n - Y| < \varepsilon) \to 1$ e $P(|Y_n - Y| \geq \varepsilon) = 1 - P(|Y_n - Y| < \varepsilon) \to 0$. □

EXEMPLO 2. Uma seqüência de variáveis aleatórias que converge em probabilidade e não quase certamente.

Seja X uma variável aleatória com distribuição uniforme no intervalo $[0, 1]$, e coloquemos os intervalos binários básicos de $[0, 1]$ na seguinte ordem:

$$I_1 = [0, 1], I_2 = [0, 1/2], I_3 = [1/2, 1], I_4 = [0, 1/4], \ldots, I_7 = [3/4, 1], I_8 = [0, 1/8], \ldots,$$

de maneira que para $m = 0, 1, 2, \ldots$ e $i = 0, 1, \ldots, 2^m - 1$, temos

$$I_{2^m+i} = \left[\frac{i}{2^m}, \frac{i+1}{2^m}\right].$$

Então os 2^m intervalos de comprimento $1/2^m$ cobrem o intervalo $[0, 1]$, ao mesmo tempo que seu comprimento fica cada vez menor.

Definamos Y_n igual ao indicador do evento $[X \in I_n]$, ou seja,

$$Y_n = \begin{cases} 1 \text{ se } X \in I_n \\ 0 \text{ se } X \notin I_n. \end{cases}$$

A seqüência $Y_1, Y_2, \ldots$ converge em probabilidade para a variável aleatória constante $Y = 0$, pois, se $0 < \varepsilon \leq 1$,

$$P(|Y_n - 0| \geq \varepsilon) = P(Y_n = 1) = P(X \in I_n),$$

e esta probabilidade, que é igual ao comprimento de I_n, converge para zero quando $n \to \infty$. (Se $\varepsilon > 1$, é impossível que $|Y_n - 0| \geq \varepsilon$, logo $P(|Y_n - 0| \geq \varepsilon) = P(\varnothing) = 0$.)

Por outro lado, Y_n não converge quase certamente para zero. De fato, não converge em ponto algum, pois qualquer que seja o valor de X, este valor pertence a um ou dois dos 2^n intervalos de comprimento $1/2^n$, para todo n. Logo, se $\omega \in \Omega$, $Y_n(\omega)$ assume o valor 1 para um número infinito de n's, assim como assume o valor 0 também para um número infinito de n's. Portanto, $Y_n(\omega)$ não converge, para cada $\omega \in \Omega$.

Agora formularemos a Lei dos Grandes Números de uma maneira mais geral do que foi feito no início deste capítulo:

Sejam $X_1, X_2, \ldots$ variáveis aleatórias *integráveis* em $(\Omega, \mathcal{A}, P)$, e sejam $S_1, S_2, \ldots$ as *somas parciais*, definidas por $S_n = X_1 + \ldots + X_n$. (Notemos que $S_1, S_2, \ldots$ também são variáveis aleatórias em $(\Omega, \mathcal{A}, P)$.)

Dizemos que $X_1, X_2, \ldots$ satisfazem a *Lei Fraca dos Grandes Números* se

$$\frac{S_n - ES_n}{n} \to 0 \text{ em probabilidade,}$$

ou, equivalentemente, se

$$P\left(\left|\frac{X_1 + \ldots + X_n - (EX_1 + \ldots + EX_n)}{n}\right| \geq \varepsilon\right) \to 0, \forall \varepsilon > 0.$$

Dizemos que $X_1, X_2, \ldots$ satisfazem a *Lei Forte dos Grandes Números* se

$$\frac{S_n - ES_n}{n} \to 0 \text{ quase certamente,}$$

ou, equivalentemente, se

$$\frac{(X_1 - EX_n) + (X_2 - EX_2) + \ldots + (X_n - EX_n)}{n} \to 0 \text{ quase certamente.}$$

Pela Proposição 5.1, se a seqüência $X_1, X_2, \ldots$ satisfaz a Lei Forte, então satisfaz a Lei Fraca (logo a Lei Forte é realmente mais "forte").

Em termos intuitivos, este conceito mais geral da Lei dos Grandes Números pode ser expresso assim: uma seqüência de variáveis aleatórias satisfaz a Lei dos Grandes Números se, quando n é grande, a média aritmética dos primeiros n observados é aproximadamente igual à média aritmética das suas médias (esperanças), ou seja, S_n/n é aproximadamente igual a $\frac{ES_n}{n} = \frac{EX_1 + \ldots EX_n}{n}$.

Observação. Se as variáveis aleatórias X_n têm a mesma média finita μ, então elas satisfazem a Lei Fraca (ou Forte) se, e somente se, $\frac{S_n}{n} \to \mu$ em probabilidade (ou quase certamente). Isto vale pois neste caso $\frac{ES_n}{n} = \mu$, e $\frac{S_n}{n} - \mu \to 0$ em probabilidade (ou quase certamente) se, e somente se, $\frac{S_n}{n} \to \mu$ em probabilidade (ou quase certamente.)

TEOREMA 5.1. (*Lei Fraca de Tchebychev*). *Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes 2 a 2 com variâncias finitas e uniformemente limitadas (i.e., existe c finito tal que para todo n, $\operatorname{Var} X_n \leq c$). Então $X_1, X_2, \ldots$ satisfazem a Lei Fraca dos Grandes Números*:

$$\frac{S_n - ES_n}{n} \xrightarrow{P} 0.$$

Prova. Precisamos mostrar que para $\varepsilon > 0$,

$$P\left(\frac{|S_n - ES_n|}{n} \geq \varepsilon\right) \to 0 \text{ quando } n \to \infty.$$

Como $\operatorname{Var} S_n = \operatorname{Var}(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \operatorname{Var} X_i \leq nc$, a desigualdade de Tchebychev implica

$$P(|S_n - ES_n| \geq \varepsilon n) \leq \frac{\operatorname{Var} S_n}{\varepsilon^2 n^2} \leq \frac{c}{\varepsilon^2 n} \to 0. \quad \square$$

Observação. Pela Proposição 3.6, basta que as X_n sejam não-correlacionadas em vez de independentes 2 a 2.

COROLÁRIO. (*Lei dos Grandes Números de Bernoulli, publicada em* Ars Conjectandi, 1713). *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*

$$\frac{S_n}{n} \to p \text{ em probabilidade.}$$

Prova. Seja

$$X_n = \begin{cases} 1 \text{ se o } n\text{-ésimo ensaio é sucesso} \\ 0 \text{ se o } n\text{-ésimo ensaio é fracasso.} \end{cases}$$

Então $X_1, X_2, \ldots$ são independentes, identicamente distribuídas e integráveis, com média $\mu = p$. Como $\operatorname{Var} X_n = p(1-p)$, a Lei Fraca de Tchebychev implica que

$$\frac{S_n - np}{n} \xrightarrow{P} 0,$$

ou, equivalentemente,

$$\frac{S_n}{n} \xrightarrow{P} p. \quad \square$$

A hipótese de variâncias finitas foi eliminada por Khintchin, que conseguiu provar a Lei dos Grandes Números no caso de variáveis independentes e identicamente distribuídas, supondo apenas integrabilidade:

TEOREMA 5.2. (*Lei Fraca de Khintchin*). *Se $X_1, X_2, \ldots$ são independentes, identicamente distribuídas e integráveis, com média comum μ, então*

$$\frac{S_n}{n} \to \mu \text{ em probabilidade.}$$

Prova. Omitida, pois o teorema já não tem muita importância (é conseqüência da Lei Forte de Kolmogorov). Porém, o método de prova (truncamento) utilizado por Khintchin, aparentemente introduzido por Markov, será usado na prova da Lei Forte. $\square$

§ 5.2 Seqüências de eventos e o Lema de Borel-Cantelli

Neste parágrafo, consideraremos o lema de Borel-Cantelli, uma ferramenta das mais úteis na Teoria da Probabilidade e uma peça importante na prova da Lei Forte.

Se $A_1, A_2, \ldots$ é uma seqüência de eventos, i.e., se $A_n \subset \Omega$ para $n = 1, 2, \ldots$, o *limite superior* da seqüência é definido por

$$\limsup_{n \to \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k;$$

o *limite inferior* é, por definição,

$$\liminf_{n \to \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k.$$

Por conveniência, costuma-se indicar esses limites por $\limsup A_n$ e $\liminf A_n$, deixando implícito o qualificador "$n \to \infty$".

O evento $\limsup A_n$ é o evento "ocorrência de um número infinito dos A_n", pelo segundo raciocínio:

Se $\omega \in \limsup A_n$, então $\omega \in \bigcup_{k=n}^{\infty} A_k, \forall n$. Como $\omega \in \bigcup_{k=1}^{\infty} A_k$, $\omega \in A_{k_1}$ para algum k_1. Mas $\omega \in \bigcup_{k=k_1+1}^{\infty} A_k$, logo $\omega \in A_{k_2}$ para algum $k_2 > k_1$. Continuando, temos $\omega \in \bigcup_{k=k_2+1}^{\infty} A_k$, logo $\omega \in A_{k_3}$ para algum $k_3 > k_2$, etc. Desta maneira obtemos uma seqüência crescente de inteiros positivos $k_1 < k_2 < k_3 < \ldots$, que dependem de ω, tais que $\omega \in A_{k_n}, \forall n$. Portanto, ω pertence a um número infinito dos A_n.

Reciprocamente, se ω pertence a um número infinito dos A_n, então $\omega \in \bigcup_{k=n}^{\infty} A_k, \forall n$, de modo que $\omega \in \limsup A_n$. Concluímos que $\omega \in \limsup A_n$ se, e somente se, ω pertence a um número infinito dos A_n.

Notação. $\limsup A_n = [A_n \text{ infinitas vezes}]$. (Salientamos aqui: o evento "A_n infinitas vezes" é o evento "ocorrência de um número infinito dos A_n". Cada A_n ocorre ou não, portanto é importante não cair no erro de pensar em infinitas ocorrências de, por exemplo, A_1.)

O evento $\liminf A_n$ também tem uma interpretação intuitiva: é o evento "ocorrência de A_n para todo n suficientemente grande". Para ver isso, note que $\omega \in \liminf A_n$ se, e somente se, $\omega \in \bigcap_{k=n_0}^{\infty} A_k$ para algum $n_0 = n_0(\omega)$, ou seja, $\omega \in A_k$ para todo k suficientemente grande ($k \geq n_0$).

Caso $\limsup A_n = \liminf A_n \stackrel{def}{=} A$, este evento é chamado *limite* de A_n e escrevemos $A = \lim_{n\to\infty} A_n$ ou $A_n \to A$. Neste caso, $P(A_n)$ converge para $P(A)$, se os eventos são aleatórios (exercício 7 do Cap. 1), de modo que probabilidade é contínua não somente para seqüências monótonas (a propriedade P6, §1.1), como também para seqüências convergentes neste sentido mais geral. (*Exercício.* Mostre que se $A_n \uparrow A$ ou $A_n \downarrow A$, então $A = \lim_{n\to\infty} A_n$.)

PROPOSIÇÃO 5.2. (*Lema de Borel-Cantelli*). *Sejam $A_1, A_2, \ldots$ eventos aleatórios em $(\Omega, \mathcal{A}, P)$, i.e., $A_n \in \mathcal{A}\ \forall n$.*

(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* A_n *são* independentes, *então* $P(A_n \text{ infinitas vezes}) = 1$.

Observação. O item (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. (a) Se $\sum P(A_n) < \infty$, então $\sum_{k=n}^{\infty} P(A_k) \to 0$ quando $n \to \infty$. Mas $[A_n \text{ infinitas vezes}] \subset \bigcup_{k=n}^{\infty} A_k\ \forall n$, logo

$$P(A_n \text{ infinitas vezes}) \leq P\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \sum_{k=n}^{\infty} P(A_k) \to 0.$$

Portanto, $P(A_n \text{ infinitas vezes}) = 0$.

(b) Basta provar que $P\left(\bigcup_{k=n}^{\infty} A_k\right) = 1\ \forall n$ (pois $[A_n \text{ infinitas vezes}] = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$ e a interseção de um número enumerável de eventos de probabilidade um, é também de probabilidade um). Para tanto, seja $B_n = \bigcup_{k=n}^{\infty} A_k$. Então B_n contém $\bigcup_{k=n}^{n+m} A_k$ para todo m, e

$$B_n^c \subset \left(\bigcup_{k=n}^{n+m} A_k\right)^c = \bigcap_{k=n}^{n+m} A_k^c.$$

Logo para todo m,

$$1 - P(B_n) = P(B_n^c) \leq P\left(\bigcap_{k=n}^{n+m} A_k^c\right) = \text{(pela independência)} =$$

$$= \prod_{k=n}^{n+m} P(A_k^c) = \prod_{k=n}^{n+m} (1 - P(A_k)).$$

Como $1 - p \leq e^{-p}$ para $0 \leq p \leq 1$, temos

$$1 - P(B_n) \leq \prod_{k=n}^{n+m} e^{-P(A_k)} = \exp\left\{-\sum_{k=n}^{n+m} P(A_k)\right\} \to 0$$

quando $m \to \infty$, pois $\sum_{k=n}^{n+m} P(A_k) \to +\infty$ quando $m \to \infty$. Logo $P(B_n) = 1$ $\forall n$, como queríamos demonstrar. □

COROLÁRIO. *Consideremos uma sequência de ensaios binomiais independentes com probabilidade p_n de sucesso no n-ésimo ensaio. Seja $X_n = 1$ se o n-ésimo ensaio é sucesso, $= 0$ se é fracaso. Então,*

vale o seguinte: se $\sum_{n=1}^{\infty} p_n = +\infty$, *então* $P\left(\sum_{n=1}^{\infty} X_n = \infty\right) = 1$; *por outro lado, se* $\sum_{n=1}^{\infty} p_n < \infty$, *então* $P\left(\sum_{n=1}^{\infty} X_n < \infty\right) = 1$.

Em outras palavras,

$\Sigma p_n < \infty \Rightarrow$ *um número finito de sucessos, quase certamente,*
$\Sigma p_n = \infty \Rightarrow$ *um número infinito de sucessos, quase certamente.*

Prova. Seja A_n o evento "sucesso no n-ésimo ensaio" $= [X_n = 1]$. Então $P(A_n) = p_n$ e os A_n são independentes. Por Borel-Cantelli,

$\Sigma p_n < \infty \Rightarrow P(A_n$ infinitas vezes$) = 0$ (item (a))
$\Sigma p_n = \infty \Rightarrow P(A_n$ infinitas vezes$) = 1$ (item (b)).

Mas $[A_n$ infinitas vezes$]$ = "um número infinito de sucessos entre todos os ensaios" e $[A_n$ infinitas vezes$]^c$ = "apenas um número finito de sucessos entre todos os ensaios". Portanto,

$$[A_n \text{ infinitas vezes}] = [\Sigma X_n = \infty],$$
$$[A_n \text{ infinitas vezes}]^c = [\Sigma X_n < \infty]. \quad \square$$

EXEMPLO 3. (Este exemplo é clássico.) Colocar um macaco diante de uma máquina de escrever, e é razoável supor que haja uma probabilidade positiva, embora reduzidíssima, dele bater as obras completas de Shakespeare sem erro. Chamar o primeiro ensaio de *sucesso* se ele realiza essa façanha, *fracasso* quando ele faz o primeiro erro. No final do primeiro ensaio, que provavelmente chegará logo, dar-lhe comida (para garantir a independência dos ensaios) e começar o segundo ensaio. Continuar assim até o infinito.

Pelo corolário, com $p_n = p > 0\ \forall n$, há probabilidade 1 dele escrever as obras completas de Shakespeare um número infinito de vezes.

EXEMPLO 4. Para complementar o exemplo 2, exibiremos outra seqüência de variáveis aleatórias Y_n que converge em probabilidade mas não quase certamente.

Sejam $X_1, X_2, \ldots$ independentes e identicamente distribuídas, com $X_n \sim \exp(1)$, i.e.,

$$F_{X_n}(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - e^{-x} & \text{se } x \geq 0, \end{cases}$$

e seja $Y_n = \dfrac{X_n}{\log n}$ para $n > 1$.

Então $Y_n \xrightarrow{P} 0$, pois $\forall \varepsilon > 0$,

$$P(|Y_n| \geq \varepsilon) = P(X_n \geq \varepsilon \log n) = e^{-\varepsilon \log n} = \frac{1}{n^\varepsilon} \to 0.$$

Para provar que $Y_n \not\to 0$ quase certamente, basta verificar que $P(Y_n \geq \varepsilon$ infinitas vezes$) = 1$ para algum $\varepsilon > 0$, pois, nesse caso, teremos $Y_n \geq \varepsilon$ infinitas vezes, com probabilidade 1, e este evento implica que Y_n não converge para zero. (Formalmente, seja $A_n = [Y_n \geq \varepsilon]$. Se $\omega \in [A_n$ infinitas vezes$]$, então $Y_n(\omega) \geq \varepsilon$) para um número infinito de n's, logo $Y_n(\omega)$ não converge para zero. Se provarmos $P(A_n$ infinitas vezes$) = 1$, teremos $P(Y_n \not\to 0) = 1$.)

Os eventos $[Y_n \geq \varepsilon]$ são independentes, pois as Y_n o são, e

$$\Sigma P(Y_n \geq \varepsilon) = \Sigma P(X_n \geq \varepsilon \log n) = \Sigma e^{-\varepsilon \log n} = \Sigma \frac{1}{n^\varepsilon} = \infty,$$

se $0 < \varepsilon \leq 1$. Pelo item (b) de Borel-Cantelli, $P(Y_n \geq \varepsilon$ infinitas vezes$) = 1$, se $0 < \varepsilon \leq 1$.

Observação. Para completar este exemplo, precisamos verificar a *existência* de variáveis aleatórias independentes com distribuição comum exp(1). A justificativa dessa existência já apareceu na introdução deste capítulo, quando falamos de um experimento básico em que podíamos observar o valor de uma variável aleatória X e consideramos uma seqüência de repetições independentes do experimento, com X_n representando a observação da variável aleatória no n-ésimo ensaio. De fato, podemos afirmar a existência de variáveis aleatórias independentes $X_1, X_2, \ldots$ tendo qualquer seqüência fixa de distribuições. A prova disso não será dada, mas o resultado é intuitivamente razoável: basta pensar em realizar uma seqüência de experimentos independentes, de maneira que uma variável aleatória X_n com a distribuição desejada seja observada no n-ésimo experimento. Enunciamos esta afirmação na forma de uma proposição:

PROPOSIÇÃO 5.3. *Sejam $F_1, F_2, \ldots$ funções de distribuição quaisquer. Então existem um espaço de probabilidade $(\Omega, \mathcal{A}, P)$ e uma sequência de variáveis aleatórias independentes $X_1, X_2, \ldots$, definidas neste espaço de probabilidade, tais que F_n é a função de distribuição de X_n.*

§5.3 A Lei Forte

As provas desta seção são de natureza mais ou menos técnica e poderão ser omitidas em uma primeira leitura. É recomendável, contudo, que mais cedo ou mais tarde o leitor tente entender a prova do Teorema 5.5 (a Lei Forte de Kolmogorov).

Como mais uma aplicação de Borel-Cantelli, temos:

TEOREMA 5.3. (*Recíproca para a Lei Forte de Kolmogorov*). *Sejam* $X_1, X_2, \ldots$ *variáveis aleatórias independentes e identicamente distribuídas. Se* $E|X_1| = +\infty$, *então, com probabilidade* 1, *a sequência*

$$\frac{|S_n|}{n} = \frac{|X_1 + \ldots + X_n|}{n}, \quad n = 1, 2, \ldots,$$

não é limitada.

Observação. A Lei Forte afirma que se as 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.

Prova. Se $E|X_1| = +\infty$, então

$$E\left(\frac{|X_1|}{k}\right) = +\infty, \ \forall k = 1, 2, \ldots.$$

Pelo critério para integrabilidade (§3.3),

$$\sum_{n=1}^{\infty} P\left(\frac{|X_1|}{k} \geq n\right) = \infty, \ \forall k.$$

As variáveis X_n são identicamente distribuídas, logo

$$\sum_{n=1}^{\infty} P\left(\frac{|X_1|}{k} \geq n\right) = \sum_{n=1}^{\infty} P\left(\frac{|X_n|}{k} \geq n\right) = \sum_{n=1}^{\infty} P\left(\frac{|X_n|}{n} \geq k\right), \ \forall k.$$

Por independência das X_n, os eventos $A_n = \left[\frac{|X_n|}{n} \geq k\right]$ são independentes, e Borel-Cantelli implica

$$P\left(\frac{|X_n|}{n} \geq k \text{ infinitas vezes}\right) = 1, \ \forall k.$$

Fazendo $B_k = \left[\frac{|X_n|}{n} \geq k \text{ infinitas vezes}\right]$, temos

$$P\left(\bigcap_{k=1}^{\infty} B_k\right) = 1,$$

pois a interseção de um número enumerável de eventos de probabilidade 1 também tem probabilidade 1. Mas o evento $\bigcap_{k=1}^{\infty} B_k$ é o evento "$\frac{|X_n|}{n} \geq k$ para um número infinito de n, para todo k", ou seja, é o evento "a seqüência $\frac{|X_n|}{n}$ é ilimitada." (Isto significa: $\omega \in \cap B_k \Leftrightarrow$ $\Leftrightarrow \frac{|X_n(\omega)|}{n}$ é seqüência 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 $n = 1, 2, \ldots$. Portanto, se $\frac{|X_n|}{n}$ é ilimitada, então $\frac{|S_n|}{n}$ é ilimitada ou $\frac{|S_{n-1}|}{n}$ é ilimitada (se as duas seqüências fossem limitadas, a sua soma também seria). Mas se $n \geq 2$,

$$\frac{|S_{n-1}|}{n} = \frac{|S_{n-1}|}{(n-1)} \cdot \frac{(n-1)}{n}$$

e $\frac{1}{2} \leq \frac{n-1}{n} < 1$, de modo que $\frac{|S_{n-1}|}{n}$ é ilimitada se, e somente se, $\frac{|S_n|}{n}$ também o é (notemos que $\frac{|S_{n-1}|}{n-1}$, para $n \geq 2$, forma a mesma seqüência que $\frac{|S_n|}{n}$, $n \geq 1$). □

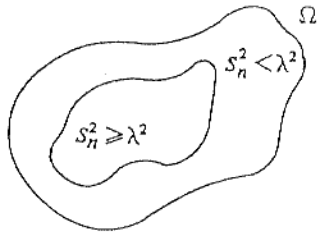
Começaremos agora a provar a Lei Forte. Primeiro, uma extensão da desigualdade de Tchebychev.

PROPOSIÇÃO 5.4. (*Desigualdade de Kolmogorov*). *Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes tais que $EX_k = 0$ e $Var\, X_k < \infty$, $k = 1, \ldots, n$. Então, para todo $\lambda > 0$,*

$$P(\max_{1 \le k \le n} |S_k| \ge \lambda) \le \frac{1}{\lambda^2} Var\, S_n = \frac{1}{\lambda^2} \sum_{k=1}^{n} Var\, X_k,$$

onde $S_k = X_1 + \ldots + X_k$.

Prova. Recordemos a técnica usada na prova alternativa da desigualdade generalizada de Tchebychev, que pode ser assim esquematizada:

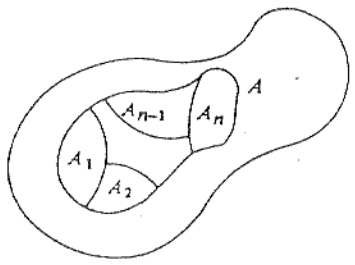


$$S_n^2 \ge S_n^2 I_{[S_n^2 \ge \lambda^2]} \ge \lambda^2 I_{[S_n^2 \ge \lambda^2]} \Rightarrow$$
$$ES_n^2 \ge \lambda^2 E I_{[S_n^2 \ge \lambda^2]} = \lambda^2 P(S_n^2 \ge \lambda^2) \Rightarrow$$
$$P(|S_n| \ge \lambda) \le \frac{1}{\lambda^2} ES_n^2 = \frac{1}{\lambda^2} \operatorname{Var} S_n.$$

Agora queremos uma cota superior para $P(\max_{1 \le k \le n} S_k^2 \ge \lambda^2)$. Para tanto, seja $A = [\max_{1 \le k \le n} S_k^2 \ge \lambda^2]$. Vamos decompor A conforme a primeira vez que $S_k^2 \ge \lambda^2$: definamos

$$A_1 = [S_1^2 \ge \lambda^2],$$
$$A_2 = [S_1^2 < \lambda^2, S_2^2 \ge \lambda^2],$$
$$A_k = [S_1^2 < \lambda^2, \ldots, S_{k-1}^2 < \lambda^2, S_k^2 \ge \lambda^2], \text{ para } 2 \le k \le n.$$

Então os A_k são disjuntos 2 a 2 e $A = \bigcup_{k=1}^{n} A_k$:



Logo $I_A = \sum_{k=1}^{n} I_{A_k}$ e

$$S_n^2 \ge S_n^2 I_A = \sum_{k=1}^{n} S_n^2 I_{A_k} \Rightarrow$$

$$ES_n^2 \ge \sum_{k=1}^{n} ES_n^2 I_{A_k}. \tag{5.1}$$

Queremos substituir S_n^2 por S_k^2 no somatório (pois $S_k^2 \ge \lambda^2$ em A_k, e não vale necessariamente $S_n^2 \ge \lambda^2$); o truque é escrever

$$S_n^2 = (S_n - S_k + S_k)^2 = (S_n - S_k)^2 + S_k^2 + 2(S_n - S_k)S_k$$
$$\ge S_k^2 + 2(S_n - S_k)S_k.$$

Portanto,

$$ES_n^2 I_{A_k} \ge ES_k^2 I_{A_k} + 2E[(S_n - S_k) S_k I_{A_k}].$$

Como $S_n - S_k = X_{k+1} + \ldots + X_n$ e $S_k I_{A_k}$ depende só de $X_1, \ldots, X_k$, 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

$$ES_n^2 I_{A_k} \ge ES_k^2 I_{A_k} \ge (\text{pois } S_k^2 \ge \lambda^2 \text{ em } A_k)$$
$$\ge E\lambda^2 I_{A_k} = \lambda^2 P(A_k).$$

Por (5.1),

$$ES_n^2 \ge \sum_{k=1}^{n} \lambda^2 P(A_k) = \lambda^2 P(A),$$

logo

$$P(A) \le \frac{1}{\lambda^2} ES_n^2 = \frac{1}{\lambda^2} \operatorname{Var} S_n. \quad \square$$

TEOREMA 5.4. (*Primeira Lei Forte de Kolmogorov*). *Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e integráveis, e suponha que*

$$\sum_{n=1}^{\infty} \frac{Var\, X_n}{n^2} < +\infty.$$

Então as X_n satisfazem a Lei Forte dos Grandes Números, i.e.,

$$\frac{X_1 + \ldots + X_n}{n} - \frac{(EX_1 + \ldots + EX_n)}{n} \to 0 \text{ quase certamente.}$$

Prova. Basta *supor* $EX_n = 0\ \forall n$. (No caso geral, seja $Y_n = X_n - EX_n$. Então $EY_n = 0$, Var $Y_n =$ Var X_n e as Y_n satisfazem as condições do teorema. Se o teorema vale para o caso $EY_n = 0$, então

$$\frac{Y_1 + \ldots + Y_n}{n} \to 0 \text{ quase certamente,}$$

i.e.,

$$\frac{(X_1 - EX_1) + \ldots + (X_n - EX_n)}{n} \to 0 \text{ quase certamente.)}$$

Queremos mostrar que $\frac{S_n}{n} \to 0$ quase certamente, onde $S_n = X_1 + \ldots + X_n$. Para tanto, basta mostrar que

$$M_n \overset{def}{=} \max_{2^n < k \le 2^{n+1}} \frac{|S_k|}{k} \to 0 \text{ quase certamente (quando } n \to \infty).$$

Provaremos isto em duas etapas:

(i) $\sum_{n=1}^{\infty} P\left(M_n \ge \frac{1}{m}\right) < \infty$, $\forall m = 1, 2, \ldots$ (usaremos a desigualdade de Kolmogorov), e

(ii) $M_n \to 0$ quase certamente (por Borel-Cantelli).

Prova de (i). Seja m fixo. Então, para todo n,

$$P\left(\max_{2^n < k \le 2^{n+1}} \frac{|S_k|}{k} \ge \frac{1}{m}\right) \le P\left(\max_{2^n < k \le 2^{n+1}} |S_k| \ge \frac{2^n}{m}\right)$$

$$\le P\left(\max_{1 \le k \le 2^{n+1}} |S_k| \ge \frac{2^n}{m}\right) \le \frac{m^2}{4^n} \sum_{k=1}^{2^{n+1}} \text{Var } X_k,$$

onde vale a última passagem pela desigualdade de Kolmogorov. Se definimos

$$A_n = \left[\max_{2^n < k \le 2^{n+1}} \frac{|S_k|}{k} \ge \frac{1}{m}\right] = \left[M_n \ge \frac{1}{m}\right],$$

temos

$$\sum_{n=1}^{\infty} P(A_n) \le m^2 \sum_{n=1}^{\infty} \left(\frac{1}{4^n} \sum_{k=1}^{2^{n+1}} \text{Var } X_k\right)$$

$$= m^2 \sum_{k=1}^{\infty} \sum_{n:\, 2^{n+1} \ge k} \left(\frac{1}{4^n} \text{Var } X_k\right)$$

$$= m^2 \sum_{k=1}^{\infty} \left(\text{Var } X_k \sum_{n:\, 2^{n+1} \ge k} \frac{1}{4^n}\right)$$

Como

$$\sum_{n=j}^{\infty} \frac{1}{4_n} = \frac{1}{4^j}\left(1 + \frac{1}{4} + \left(\frac{1}{4}\right)^2 + \ldots\right) = \frac{1}{4^j} \cdot \frac{4}{3}$$

e

$$\sum_{n:\, 2^{n+1} \ge k} \frac{1}{4^n} = \sum_{n=j(k)}^{\infty} \frac{1}{4^n},$$

onde $j(k)$ satisfaz $2^{j(k)+1} \ge k > 2^{j(k)}$, temos

$$\sum_{n:\, 2^{n+1} \ge k} \frac{1}{4^n} = \frac{4}{3} \cdot \frac{1}{4^{j(k)}} \le (\text{pois } 2^{j(k)} \ge k/2) \le \frac{4}{3} \cdot \frac{4}{k^2} = \frac{16}{3k^2}.$$

Portanto,

$$\sum_{n=1}^{\infty} P(A_n) \le \frac{16m^2}{3} \sum_{k=1}^{\infty} \frac{\text{Var } X_k}{k^2} < \infty \text{ (por hipótese).}$$

Prova de (ii). Com a mesma notação de (i), temos $P(A_n$ infinitas vezes$) = 0$, por Borel-Cantelli. Em outras palavras, para todo m fixo a probabilidade é 0 de que M_n assuma um valor $\ge 1/m$ infinitas vezes. Isto significa que para todo m, a probabilidade é 1 de que M_n assuma um valor $\ge 1/m$ para somente um número finito de n's. Fazendo

$B_m =$ "M_n assume um valor $\ge \frac{1}{m}$ para somente um número finito de n's",

temos, então, $P(B_m) = 1\ \forall m$ e $P\left(\bigcap_{m=1}^{\infty} B_m\right) = 1$.

Para terminar a prova, basta observar que o evento $\bigcap_{m=1}^{\infty} B_m$ é equivalente ao evento $[M_n \to 0]$. Para ver essa equivalência, notemos que $M_n \geq 0$ e portanto: $\omega \in \bigcap_{m=1}^{\infty} B_m \Leftrightarrow$

$\Leftrightarrow \forall m,\ M_n(\omega) \geq 1/m$ para somente um número finito de n's
$\Leftrightarrow \forall m,\ 0 \leq M_n(\omega) < 1/m$ para todo n suficientemente grande
$\Leftrightarrow M_n(\omega) \to 0$ quando $n \to \infty$. □

COROLÁRIO. *A Lei Forte é satisfeita por toda sequência de variáveis aleatórias independentes e uniformemente limitadas.*

Prova. Se $X_1, X_2, \ldots$ são uniformemente limitadas, então existe c finito tal que $|X_n| \leq c\ \forall n$. Neste caso, $\operatorname{Var} X_n \leq EX_n^2 \leq c^2$ e, como as variâncias estão limitadas, a condição do teorema está satisfeita. □

EXEMPLO 5. Consideremos uma seqüência de ensaios binomiais independentes, com probabilidade p_n de sucesso no n-ésimo ensaio. Se X_n é o indicador de sucesso no n-ésimo ensaio, então $X_1, X_2, \ldots$ são independentes e uniformemente limitadas ($|X_n| \leq 1\ \forall n$). Portanto, decorre da Lei Forte que

$$\frac{S_n}{n} - \frac{p_1 + \ldots + p_n}{n} \to 0 \text{ quase certamente,}$$

onde $S_n = X_1 + \ldots + X_n$. Em outras palavras, a freqüência relativa de sucessos em n ensaios independentes, menos a freqüência relativa esperada, converge para zero quase certamente quando $n \to \infty$.

Por exemplo, se você jogar uma moeda honesta no 1.º ensaio, com sucesso igual a cara e $p_1 = 1/2$; lançar um dado equilibrado no 2.º ensaio, com sucesso igual a 4 e $p_2 = 1/6$; colocar um macaco diante de uma máquina de escrever no 3.º ensaio, com as obras completas de Shakespeare como sucesso e $p_3 = \Delta$ (pequena mas positiva); etc., então para n grande a sua freqüência relativa de sucessos será aproximadamente igual a

$$\frac{1/2 + 1/6 + \Delta + \ldots + p_n}{n}.$$

(Com probabilidade 1, a diferença entre estes dois valores convergirá para zero quando $n \to \infty$.)

LEMA 5.1. (Este lema será usado na prova do teorema a seguir.) *Seja X uma variável aleatória integrável com função de distribuição F. Então,*

$$\sum_{n=1}^{\infty} \left\{ \frac{1}{n^2} \int_{-n}^{n} x^2\, dF(x) \right\} < \infty.$$

Observação. Lembre-se de que o extremo direito está incluído no intervalo de integração e o extremo esquerdo não está, ou seja, $\int_{-n}^{n} x^2\, dF(x) = \int_{(-n,n]} x^2\, dF(x)$ (veja os itens 9 e 10 do §3.1). O lema ainda vale quando a integral é substituída por $\int_{[-n,n)} x^2\, dF(x)$, e a prova é quase a mesma.

Prova. Vamos utilizar o seguinte *fato*: para $j = 1, 2, \ldots$,

$$\sum_{n=j}^{\infty} \frac{1}{n^2} \leq \frac{2}{j}.$$

(Prova do fato: para $n = 2, 3, \ldots$,

$$\frac{1}{n^2} \leq \frac{1}{n(n-1)} = \frac{1}{n-1} - \frac{1}{n},$$

logo

$$\sum_{n=j}^{\infty} \frac{1}{n^2} = \frac{1}{j^2} + \sum_{n=j+1}^{\infty} \frac{1}{n^2} \leq \frac{1}{j^2} + \sum_{n=j+1}^{\infty} \left(\frac{1}{n-1} - \frac{1}{n} \right) =$$
$$= \frac{1}{j^2} + \frac{1}{j} \leq \frac{2}{j}.\ )$$

Como

$$\int_{-n}^{n} x^2\, dF(x) = \sum_{j=-n+1}^{n} \int_{j-1}^{j} x^2\, dF(x),$$

temos

$$\sum_{n=1}^{\infty} \left\{ \frac{1}{n^2} \int_{-n}^{n} x^2\, dF(x) \right\} = \sum_{n=1}^{\infty} \sum_{j=-n+1}^{n} \left\{ \frac{1}{n^2} \int_{j-1}^{j} x^2\, dF(x) \right\} =$$
$$= \sum_{j=1}^{\infty} \sum_{n=j}^{\infty} \left\{ \frac{1}{n^2} \int_{j-1}^{j} x^2\, dF(x) \right\} + \sum_{j=-\infty}^{0} \sum_{n=|j|+1}^{\infty} \left\{ \frac{1}{n^2} \int_{j-1}^{j} x^2\, dF(x) \right\}$$
$$\leq \text{(pelo fato)} \leq 2 \sum_{j=1}^{\infty} \int_{j-1}^{j} \frac{x^2}{j}\, dF(x) + 2 \sum_{j=-\infty}^{0} \int_{j-1}^{j} \frac{x^2}{|j|+1}\, dF(x).$$

Como $\frac{x^2}{j} \le x$ em $(j-1, j]$, para $j \ge 1$, e $\frac{x^2}{|j|+1} \le |x|$ em $(j-1, j]$, para $j \le 0$, temos

$$\sum_{n=1}^{\infty} \left\{ \frac{1}{n^2} \int_{-n}^{n} x^2 \, dF(x) \right\} \le 2 \sum_{j=1}^{\infty} \int_{j-1}^{j} x \, dF(x) + 2 \sum_{j=-\infty}^{0} \int_{j-1}^{j} |x| \, dF(x) =$$

$$= 2 \sum_{j=-\infty}^{\infty} \int_{j-1}^{j} |x| \, dF(x) = 2 \int_{-\infty}^{\infty} |x| \, dF(x) = 2E|X| < \infty. \quad \square$$

TEOREMA 5.5. (*A Lei Forte de Kolmogorov*). *Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes, identicamente distribuídas e integráveis, com $EX_n = \mu$. Então*

$$\frac{X_1 + \ldots + X_n}{n} \to \mu \text{ quase certamente.}$$

(Já vimos no Teorema 5.3 a recíproca deste teorema.)

Prova. Basta *supor* $\mu = 0$. (No caso geral, seja $U_n = X_n - \mu$. Então as variáveis aleatórias U_n são independentes e identicamente distribuídas e $EU_n = 0$. Se

$$\frac{U_1 + \ldots + U_n}{n} \to 0 \text{ quase certamente,}$$

então

$$\frac{X_1 + \ldots + X_n}{n} \to \mu \text{ quase certamente.)}$$

Vamos *truncar* as variáveis X_n: definamos

$$Y_n = \begin{cases} X_n \text{ se } -n < X_n \le n \\ 0 \quad \text{caso contrário,} \end{cases}$$

i.e., $Y_n = X_n I_{[-n < X_n \le n]}$.

Seja $Z_n = X_n - Y_n$, de modo que $X_n = Y_n + Z_n$ e

$$\frac{X_1 + \ldots + X_n}{n} = \frac{Y_1 + \ldots + Y_n}{n} + \frac{Z_1 + \ldots + Z_n}{n}$$

A prova terá três partes:

(a) $\frac{Z_1 + \ldots + Z_n}{n} \to 0$ quase certamente (usaremos Borel-Cantelli),

(b) $\frac{Y_1 + \ldots + Y_n}{n} - \frac{EY_1 + \ldots + EY_n}{n} \to 0$ quase certamente (Lema 5.1 e Teorema 5.4),

(c) $\frac{EY_1 + \ldots + EY_n}{n} \to 0$ (pelo Teorema da Convergência Dominada).

(a), (b) e (c) implicam o teorema, pois basta somar os três termos na interseção dos eventos quase certos de (a) e (b): sejam

$$A = \left[\frac{Z_1 + \ldots + Z_n}{n} \to 0 \right] \text{ e } B = \left[\frac{Y_1 + \ldots + Y_n}{n} - \frac{EY_1 + \ldots + EY_n}{n} \to 0 \right],$$

então

$$P(A) = 1, P(B) = 1, P(A \cap B) = 1 \text{ e } \frac{X_1 + \ldots + X_n}{n} \to 0 \text{ em } A \cap B.$$

Prova de (a). Por definição, $Z_n \ne 0 \Leftrightarrow Y_n \ne X_n \Leftrightarrow X_n \in (-n, n]$. Logo,

$$P(Z_n \ne 0) = P(X_n \notin (-n, n]) \le P(|X_n| \ge n).$$

Mas os eventos $A_n = [Z_n \ne 0]$ satisfazem

$$\sum_{n=1}^{\infty} P(A_n) \le \sum_{n=1}^{\infty} (P|X_n| \ge n) = \text{(são identicamente distribuídas)} =$$

$$= \sum_{n=1}^{\infty} P(|X_1| \ge n) < \infty,$$

onde a última passagem é conseqüência da integrabilidade de X_1.

Portanto decorre de Borel-Cantelli que $P(A_n \text{ infinitas vezes}) = 0$, ou seja

$$P(Z_n \ne 0 \text{ infinitas vezes}) = 0.$$

Isso significa

$$P(Z_n \ne 0 \text{ para apenas um número finito de } n) = 1,$$

i.e.,

$$P(Z_n = 0 \text{ para todo } n \text{ suficientemente grande}) = 1.$$

Mas se $Z_n = 0$ para n suficientemente grande, então $Z_n \to 0$ e

$$\frac{Z_1 + \ldots + Z_n}{n} \to 0, \text{ logo } P\left(\frac{Z_1 + \ldots + Z_n}{n} \to 0 \right) = 1.$$

Prova de (*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 $Y_n = X_n I_{[-n<X_n\le n]}$, temos

$$\text{Var } Y_n \le EY_n^2 = E(X_n^2 I_{[-n<X_n\le n]}) = \text{(pelo Teorema 3.1)} =$$
$$= \int x^2 I_{(-n,n]}(x)\, dF(x) = \int_{-n}^{n} x^2\, dF(x),$$

valendo a última igualdade pelo item 10, §3.1. Logo

$$\sum_{n=1}^{\infty} \frac{\text{Var } Y_n}{n^2} \le \sum_{n=1}^{\infty} \frac{1}{n^2} \int_{-n}^{n} x^2\, dF(x) < \infty,$$

pelo Lema 5.1. Segue-se da primeira Lei Forte de Kolmogorov que as Y_n satisfazem a Lei Forte, e (b) está provado.

Prova de (*c*). É suficiente demonstrar que $EY_n \to 0$. Mas

$$EY_n = E(X_n I_{[-n<X_n\le n]}) = \text{(são identicamente distribuídas)} =$$
$$= E(X_1 I_{[-n<X_1\le n]}) \to EX_1 = 0,$$

pelo Teorema da Convergência Dominada. (O teorema é aplicável, pois $|X_1 I_{[-n<X_1\le n]}| \le |X_1|$ integrável e $X_1 I_{[-n<X_1\le n]} \to X_1$ em toda parte. Notemos que $[-n < X_n \le n] = \{\omega: -n < X_1(\omega) \le n\} \uparrow \Omega$, de modo que $I_{[-n<X_1\le n]}(\omega) \to 1\ \forall \omega$.) □

Observação. Consideremos um evento aleatório A associado a um experimento E, tal que a probabilidade de ocorrência de A, quando se realiza o experimento E, é p. Se o experimento é realizado independentemente n vezes, e se S_n é o número de ocorrências do evento A nessas n realizações, então $S_n/n \to p$ quase certamente quando $n \to \infty$. S_n/n é a freqüência relativa de ocorrência de A nos n ensaios. Esse resultado, um caso particular da Lei Forte de Kolmogorov e também da primeira Lei Forte de Kolmogorov (veja o exemplo 5), é devido a Borel, e pode ser formalmente enunciado da seguinte maneira:

COROLÁRIO. (*Lei Forte de Borel*, 1909). *Sejam $X_1, X_2, \ldots$ independentes e identicamente distribuídas tais que $P(X_n = 1) = p$, $P(X_n = 0) = 1 - p$. Então $S_n/n \to p$ quase certamente, onde $S_n = X_1 + \ldots + X_n$.*

EXEMPLO 6. *Números normais* (Borel, 1909). Seja $x \in [0, 1]$ e seja $0, x_1 x_2 x_3 \ldots$ a expansão binária de x, de modo que $x = \sum_{n=1}^{\infty} \frac{x_n}{2^n}$ (quando houver duas expansões, tome a que termina, por exemplo $1/4 = 0{,}01 = 0{,}0100\ldots$). Então $\frac{x_1 + \ldots + x_n}{n}$ é a freqüência relativa de 1's entre os primeiros n algarismos da expansão. É evidente que se x for racional binário, então $\frac{x_1 + \ldots + x_n}{n} \to 0$, pois a expansão "termina" e $x_n = 0$ para todo n suficientemente grande (a escolha da outra expansão implicaria convergência para 1). Mas se x não for racional binário?

Ocorre que o caso dos racionais é um tanto patológico, pois $\frac{x_1 + \ldots + x_n}{n} \to \frac{1}{2}$ para quase todo x ("quase todo" em relação à medida de Lebesgue, i.e., o conjunto dos x tais que $\frac{x_1 + \ldots + x_n}{n} \not\to \frac{1}{2}$ tem comprimento zero). Notemos que $\frac{x_1 + \ldots + x_n}{n} \to \frac{1}{2}$ se, e somente se, $\frac{(1 - x_1) + \ldots + (1 - x_n)}{n} \to \frac{1}{2}$, de modo que a freqüência relativa de 0's também converge a 1/2 para quase todo x. Um x que tem essa propriedade é chamado de *número (simplesmente) normal com relação à base* 2 (x seria normal em relação à base k, para k inteiro ≥ 2, se $1/k$ fosse o limite da freqüência relativa de j na expansão de x na base k, para $j = 0, 1, \ldots, k - 1$). Provemos agora que quase todo número pertencente a $[0, 1]$ é normal com relação à base 2.

Seja $\Omega = [0, 1]$, $\mathscr{A} = \mathscr{B}_{[0,1]}$ = borelianos de $[0, 1]$, P = probabilidade uniforme em $[0, 1]$ (= medida de Lebesgue). Definamos uma seqüência de variáveis aleatórias:

$$X_1 = \begin{cases} 0 \text{ em } [0, 1/2) \\ 1 \text{ em } [1/2, 1], \end{cases}$$
$$X_2 = \begin{cases} 0 \text{ em } [0, 1/4) \cup [1/2, 3/4) \\ 1 \text{ em } [1/4, 1/2) \cup [3/4, 1], \end{cases}$$

etc. Então $X_n(x) = x_n$ e é fácil ver que as X_n tem mesma distribuição, com $P(X_n = 0) = 1/2 = P(X_n = 1)$.

Além disso, elas são independentes, pois, por exemplo, $P(X_1 = 1, X_2 = 1) = P([3/4, 1]) = 1/4 = P(X_1 = 1)\, P(X_2 = 1)$. Como $EX_n = 1/2$, de-

corre da Lei Forte que $\frac{X_1 + \ldots + X_n}{n} \to \frac{1}{2}$ quase certamente, ou seja, $\frac{x_1 + \ldots + x_n}{n} \to \frac{1}{2}$ para quase todo x.

Agora perguntamos: qual o limite da freqüência relativa de "11" na expansão binária de x? Queremos dizer: se $y_n = 1$ quando $x_n = 1 = x_{n+1}$ e $y_n = 0$ caso contrário, qual o limite de $\frac{y_1 + \ldots + y_n}{n}$? (Observação: se a expansão de x começa com 0,0111011 então $y_1 = 0$, $y_2 = 1$, $y_3 = 1$, $y_4 = 0$, $y_5 = 0$, $y_6 = 1$.) Resposta: 1/4 para quase todo x!

A prova disto é só um pouco mais complicada que a anterior: definimos $Y_n = X_n \cdot X_{n+1}$, $n = 1, 2, \ldots$. Então as variáveis aleatórias Y_n são identicamente distribuidas, com $P(Y_n = 1) = P(X_n = 1, X_{n+1} = 1) = 1/4$ e $P(Y_n = 0) = 3/4$, mas não são independentes (por quê?). Porém, Y_1, Y_3, Y_5, ... são independentes, como também o são Y_2, Y_4, Y_6, ... Logo

$$\frac{Y_1 + Y_3 + Y_5 + \ldots + Y_{2n-1}}{n} \to \frac{1}{4} \text{ quase certamente}$$

e

$$\frac{Y_2 + Y_4 + \ldots + Y_{2n}}{n} \to \frac{1}{4} \text{ quase certamente.}$$

Segue-se que

$$\frac{Y_1 + \ldots + Y_{2n}}{2n} = \frac{1}{2}\left(\frac{Y_1 + Y_3 + \ldots + Y_{2n-1}}{n} + \frac{Y_2 + Y_4 + \ldots + Y_{2n}}{n}\right) \to \frac{1}{4} \text{ quase certamente.}$$

Como convergência da subseqüência correspondente aos números pares já determina convergência de toda a seqüência (pois

$$\left|\frac{Y_1 + \ldots + Y_{2n+1}}{2n+1} - \frac{Y_1 + \ldots + Y_{2n}}{2n}\right| \leq \frac{|2n\, Y_{2n+1} - (Y_1 + \ldots + Y_{2n})|}{2n(2n+1)} \leq \leq \frac{1}{2n+1} \to 0),$$

o resultado está provado.

Analogamente, o limite da freqüência relativa de "01" é 1/4, de "010" é 1/8, de "1101011" é $1/2^7$, etc., onde para cada tal seqüência

finita o resultado vale para quase todo x. Como existe um número enumerável de seqüências finitas de 0's e 1's, segue-se que quase todo x é *inteiramente normal com relação à base* 2. (Definição: se $1/2^m$ é o limite da freqüência relativa de $\delta_1 \ldots \delta_m$ na expansão binaria de x, para toda seqüência $\delta_1 \ldots \delta_m$ de 0's e 1's, e para todo $m \geq 1$, então dizemos que x é inteiramente normal com relação à base 2.) Para ver isso, basta fazer $A_{\delta_1 \ldots \delta_m} = \{x \in [0, 1]$: $1/2^m$ é limite da freqüência relativa de $\delta_1 \ldots \delta_m$ na expansão binária de $x\}$. Como $P(A_{\delta_1 \ldots \delta_m}) = 1$ $\forall \delta_1 \ldots \delta_m$ e $\forall m$, temos $P(\cap A_{\delta_1 \ldots \delta_m}) = 1$, logo quase todo x é inteiramente normal com relação à base 2.

O mesmo resultado vale para qualquer base $k \geq 2$: quase todo x é inteiramente normal com relação à base k (i.e., a freqüência relativa de $\delta_1 \ldots \delta_m$ tende a $1/k^m$ $\forall \delta_j = 0, 1, \ldots, k-1$, $\forall j = 1, \ldots m$, $\forall m \geq 1$). Se x é inteiramente normal à base k para todo inteiro $k \geq 2$, x é chamado *absolutamente normal*. Então o resultado de Borel é que quase todo x é absolutamente normal. (Não é fácil encontrar um tal x. Qualquer racional não serve, pois a expansão se repete. Exemplo de número inteiramente normal com relação à base 2: $x = 0{,}0|1|00|01|10|11|000|001|\ldots.$)



EXERCÍCIOS DO CAPÍTULO 5

§5.1

1. Seja $A_1, A_2, \ldots$ uma seqüência de eventos aleatórios em $(\Omega, \mathcal{A}, P)$, com indicadores $I_{A_1}, I_{A_2}, \ldots$. Mostre que $P(A_n) \to 0$ se, e somente se, $I_{A_n} \xrightarrow{P} 0$.

2. Sejam $X_1, X_2, \ldots$ variáveis independentes com distribuição comum Poisson (λ). Qual o limite em probabilidade da seqüência $(Y_n)_{n \geq 1}$, onde $Y_n = \frac{X_1^2 + \ldots + X_n^2}{n}$?

3. Seja $(X_n)_{n \geq 1}$ uma seqüência de variáveis aleatórias. Prove que se $EX_n \to \alpha$ e $\mathrm{Var}(X_n) \to 0$, então $X_n \xrightarrow{P} \alpha$.

4. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes tais que $X_1 = 0$ e para $j \geq 2$, X_j é variável aleatória discreta satisfazendo

$$P(X_j = k) = \begin{cases} \frac{1}{j^3} \text{ se } k = \pm 1, \pm 2, \ldots, \pm j \\ 1 - \frac{2}{j^2} \text{ se } k = 0. \end{cases}$$

Prove que

$$\frac{\sum_{j=1}^{n} X_j}{n^\alpha} \xrightarrow{P} 0$$

quando $n \to \infty$, se $\alpha > \frac{1}{2}$. $\left(\text{Dado: } \sum_{k=1}^{j} k^2 = \frac{j(j+1)(2j+1)}{6}.\right)$

§5.2.

5. Seja S uma seqüência finita de caras e coroas. Demonstre que se uma moeda não necessariamente honesta (com probabilidade de cara igual a p, $0 < p < 1$) for jogada independentemente um número infinito de vezes, então S sairá infinitas vezes na seqüência obtida, com probabilidade 1.

6. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes tais que X_n tem distribuição $U[0, a_n]$ onde $a_n > 0$. Mostre:
   (a) Se $a_n = n^2$, então com probabilidade 1, somente um número finito das X_n toma valores menores que 1.
   (b) Se $a_n = n$, então com probabilidade 1, um número infinito das X_n toma valores menores que 1.

7. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes tais que $P(X_n = 1) = 1/n$, $P(X_n = 0) = 1 - 1/n$. Mostre que $X_n \xrightarrow{P} 0$ mas $P(X_n \to 0) = 0$.

8. Observa-se uma seqüência infinita de lançamentos independentes de moedas, onde o n-ésimo lançamento é duma moeda com probabilidade p_n de cair "cara". Determine a probabilidade de cara sair infinitas vezes na seqüência observada, se
   (a) $\Sigma p_n = +\infty$;
   (b) $\Sigma p_n < \infty$.

9. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas, com distribuição exponencial de parâmetro 1.

Mostre que $P\left(\frac{X_n}{\log n} > 1 \text{ infinitas vezes}\right) = 1$ mas $P\left(\frac{X_n}{\log n} > 2 \text{ infinitas vezes}\right) = 0$.

10. (Exemplo de uma seqüência de variáveis aleatórias que converge quase certamente sem a convergência de momento algum.) Sejam $X_1, X_2, \ldots$ variáveis aleatórias tais que

$$P(X_n = 0) = 1 - \frac{1}{n^2}, \quad P(X_n = n^2) = \frac{1}{n^2},$$

para $n = 1, 2, \ldots$. Demonstre que X_n converge quase certamente (ache o limite X), mas $EX_n^m \not\to EX^m$ quando $n \to \infty$, para todo $m = 1, 2, \ldots$.

11. Sejam $X_1, X_2, \ldots$ variáveis aleatórias.
   (a) Demonstre: se $\sum_{n=1}^{\infty} P(|X_n| > n) < \infty$, então $\limsup_{n\to\infty} \frac{|X_n|}{n} \leq 1$ quase certamente.
   (b) Se as X_n são identicamente distribuídas e integráveis, demonstre que

$$\limsup_{n\to\infty} \frac{|X_n|}{n} \leq 1 \text{ quase certamente.}$$

12. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas tais que $X_1 \sim U[0, 1]$. Prove que $n^{-X_n} \to 0$ em probabilidade, mas n^{-X_n} não converge quase certamente para 0. (Sugestão para a parte quase certa: prove que $P(n^{-X_n} \to 0) = 0$.)

13. Prove que para cada seqüência $(X_n)_{n\geq 1}$ de variáveis aleatórias em $(\Omega, \mathcal{A}, P)$, existe uma seqüência $(b_n)_{n\geq 1}$ de números reais positivos tal que $\frac{X_n}{b_n} \to 0$ quase certamente. (*Sugestão.* Mostre que para cada n existe b_n tal que $P\left(|X_n| > \frac{b_n}{n}\right) < \frac{1}{n^2}$.)

§5.3.

14. Sejam $X_1, X_2, \ldots$ independentes e identicamente distribuídas, com $X_1 \sim U[0, 1]$. Ache o limite quase certo da média geométrica $\left(\prod_{k=1}^{n} X_k\right)^{1/n}$. (*Sugestão.* Tome logaritmos.)

15. Demonstre: se $X_1, X_2, \ldots$ são independentes e identicamente distribuídas, com $EX_1 = 1 = \text{Var}\, X_1$, então

$$\frac{\sum_{i=1}^{n} X_i}{\sqrt{n \sum_{i=1}^{n} X_i^2}} \to \frac{1}{\sqrt{2}} \text{ quase certamente.}$$

16. Seja $0 < \theta < 1/2$. Prove que se $X_1, X_2, \ldots$ são independentes tais que $P(X_n = n^\theta) = 1/2 = P(X_n = -n^\theta)$, então

$$\frac{X_1 + \ldots + X_n}{n} \to 0 \text{ quase certamente.}$$

17. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes com densidade comum

$$f(x) = \begin{cases} e^{-(x+1/2)}, & x \geq -1/2 \\ 0, & x < 1/2. \end{cases}$$

Demonstre que $S_n \to +\infty$ quase certamente, onde $S_n = X_1 + \ldots + X_n$.

18. Sejam $X_1, X_2, \ldots$ independentes e identicamente distribuídas com média μ_1 e variância σ_1^2, e sejam $Y_1, Y_2, \ldots$ independentes e identicamente distribuídas com média μ_2 e variância σ_2^2, onde $0 < \sigma_1^2 < \infty$ e $0 < \sigma_2^2 < \infty$. Defina-se uma seqüência de variáveis aleatórias $Z_1, Z_2, \ldots$ da seguinte maneira: joga-se uma moeda honesta e define-se $Z_1 = X_1$ se dá cara e $Z_1 = Y_1$ se dá coroa. Depois joga-se a moeda de novo, definindo-se $Z_2 = X_2$ se dá cara e $Z_2 = Y_2$ se dá coroa, etc. (ad infinitum). Suponha que todas as X's e Y's são independentes e que os lançamentos da moeda não dependem das X's e Y's. Explique se a seqüência $Z_1, Z_2, \ldots$ obedece à Lei Forte dos Grandes Números. Se obedece, qual o limite de

$$\bar{Z}_n = \frac{Z_1 + \ldots + Z_n}{n}?$$

19. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes tais que $X_k \sim b(n_k, p)$, onde $0 < p < 1$ (p fixo).

(a) Qual a distribuição de $S_n = \sum_{k=1}^{n} X_k$?

(b) Se $n_k \leq \sqrt{k}$, mostre que a seqüência satisfaz a Lei Forte.

20. Uma massa radioativa emite partículas segundo um processo de Poisson com parâmetro $\lambda > 0$. Sejam $T_1, T_2, \ldots$ os tempos transcorridos entre emissões sucessivas. Ache o $\lim_{n\to\infty} \frac{T_1^2 + \ldots + T_n^2}{n}$. É limite quase certo ou em probabilidade?

21. Sejam $X_1, X_2, \ldots$ independentes com distribuição comum $N(0, 1)$. Qual o limite quase certo de

$$\frac{X_1^2 + \ldots + X_n^2}{(X_1 - 1)^2 + \ldots + (X_n - 1)^2}?$$

22. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes tais que $X_n \sim U[0, n]$, $n = 1, 2, \ldots$. Chame o n-ésimo ensaio de *sucesso* se $X_{2n} > X_{2n-1}$, *fracasso* se $X_{2n} \leq X_{2n-1}$, para $n = 1, 2, \ldots$. Determine a probabilidade de haver sucesso no n-ésimo ensaio e ache o limite (se existir) de S_n/n, onde S_n = número de sucessos nos primeiros n ensaios. Esse limite é limite em probabilidade e/ou quase certo?

23. A Lei Forte para variáveis aleatórias independentes, identicamente distribuídas e integráveis pode ser estendida ao caso de esperanças infinitas, se admitirmos limites infinitos. Em particular, se $X_1, X_2, \ldots$ são independentes e identicamente distribuídas tais que $EX_n = +\infty$, então $S_n/n \to +\infty$ quase certamente. (Compare com o Teorema 5.3. Qual a diferença?) Prove esse resultado em 3 etapas:
(a) Para m inteiro positivo fixo, seja Y_n o truncamento de X_n em m:

$$Y_n = \begin{cases} X_n \text{ se } X_n \leq m \\ 0 \quad \text{se } X_n > m. \end{cases}$$

Então $\frac{Y_1 + \ldots + Y_n}{n} \to EY_1$ quase certamente, onde

$$EY_1 = \int_{-\infty}^{m} x\, dF_{X_1}(x).$$

(b) $\liminf_{n\to\infty} \frac{S_n}{n} \geq \int_{-\infty}^{m} x\, dF_{X_1}(x)$ quase certamente. (*Sugestão*: $X_n \geq Y_n$.)

(c) $\frac{S_n}{n} \to +\infty$ quase certamente. (Faça $m \to +\infty$ em (b)).

24. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes, identicamente distribuídas e integráveis. Determine $\lim_{n\to\infty} E(X_1 \mid X_1 + \ldots + X_n)$.

Qual o tipo de convergência?

25. Seja $(X_n)_{n\geq 1}$ uma seqüência de variáveis aleatórias, cada qual tomando valores 0 ou 1. Suponha $P(X_1 = 1) = 1/2$ e $P(X_{n+1} = X_n \mid X_1 = x_1, \ldots, X_n = x_n) = 1 - \alpha_n$ para todo $(x_1, \ldots, x_n)$, $n = 1, 2, \ldots$. Faça-se $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Discuta se $Y_n \to 1/2$ em probabilidade ou quase certamente quando
(a) $\alpha_n = 1/2\ \forall n$, e
(b) $\Sigma\alpha_n$ converge.

26. Sejam $X_1, X_2, \ldots$ independentes tais que $EX_n = 0\ \forall n$. Demonstre que se $\sum_{n=1}^{\infty} \operatorname{Var} X_n < \infty$, então $E(\sup_{n\geq 1} |S_n|) < \infty$, onde $S_n = X_1 + \ldots + X_n$. (*Sugestão.* Use o critério para integrabilidade do §3.3 e a desigualdade de Kolmogorov.)

FUNÇÕES CARACTERÍSTICAS E CONVERGÊNCIA EM DISTRIBUIÇÃO

§6.1 Funções características

Neste capítulo estudaremos o conceito de convergência em distribuição de seqüências de variáveis e vetores aleatórias. Uma ferramenta de grande utilidade para este estudo é a função característica. O resultado mais importante deste capítulo é que uma seqüência de variáveis aleatórias converge em distribuição se, e somente se, a seqüência de suas funções características converge pontualmente para a função característica do limite. A definição de convergência em distribuição será dada no parágrafo 6.2; nesta seção, veremos a definição e algumas propriedades básicas de funções características.

Embora funções características assumam valores complexos, não é preciso ter muita familiaridade com números complexos para poder trabalhar com elas. Isto ficará claro durante o decorrer da discussão desta seção. Neste capítulo, o *símbolo i representará sempre o número imaginário* $\sqrt{-1}$.

Se X e Y são variáveis aleatórias em $(\Omega, \mathcal{A}, P)$, então $Z = X + iY$ é chamada uma *variável aleatória complexa*. Notemos que Z é uma função definida em Ω e que assume valores complexos, com $Z(\omega) = X(\omega) + iY(\omega)$ para $\omega \in \Omega$. A *esperança* EZ é definida por linearidade, $EZ = EX + iEY$, se EX e EY são finitas.

Pela *fórmula de Euler* $e^{ix} = \cos x + i \operatorname{sen} x$, $x \in \mathbb{R}$, vemos que a variável aleatória complexa $e^{iX} = \cos X + i \operatorname{sen} X$ sempre possui esperança finita, para toda variável aleatória X, pois as variáveis aleatórias $\cos X$ e $\operatorname{sen} X$ são limitadas. Assim, a esperança na definição seguinte é finita, e garantimos que a função característica está bem definida.

DEFINIÇÃO 6.1. Seja X uma variável aleatória. A *função característica* de X é a função $\varphi\colon \mathbb{R} \to \mathbb{C}$ definida por

$$\varphi(t) = \varphi_X(t) = Ee^{itX},$$

onde definimos

$$Ee^{itX} = E\cos(tX) + iE\,\mathrm{sen}(tX),\ t \in \mathbb{R}.$$

Observação. Pelo Teorema 3.1 (esperança de uma função de X), temos

$$\varphi_X(t) = \int \cos(tx)\,dF_X(x) + i\int \mathrm{sen}(tx)\,dF_X(x)$$
$$= \int e^{itx}\,dF_X(x),\ t \in \mathbb{R},$$

onde a última igualdade decorre da linearidade da integral de Stieltjes para o caso de integrandos complexos. A última integral acima é chamada, em Análise, a transformada de Fourier-Stieltjes de F_X, e fornece uma *definição alternativa* de função característica. Notemos que a função característica é determinada pela função de distribuição; se X e Y são identicamente distribuídas, então $\varphi_X = \varphi_Y$. Mais adiante veremos que a distribuição é determinada pela função característica (propriedade FC6).

Propriedades da função característica.

FC1. *A função característica é limitada por* 1: $|\varphi_X(t)| \le 1,\ \forall t \in \mathbb{R}$.

Prova. O valor absoluto de um número complexo $z = x + iy$ é $|z| = \sqrt{x^2 + y^2}$, de modo que

$$|\varphi_X(t)| = |Ee^{itX}| = \sqrt{E^2\cos(tX) + E^2\,\mathrm{sen}(tX)} \le \text{(por Jensen)}$$
$$\le \sqrt{E\cos^2(tX) + E\,\mathrm{sen}^2(tX)} = \sqrt{E\{\cos^2(tX) + \mathrm{sen}^2(tX)\}} = 1,$$

onde a última passagem usa o fato de que $\cos^2 x + \mathrm{sen}^2 x = 1$. □

Observação. É propriedade da integral de Stieltjes para integrandos complexos que $\left|\int g(x)\,dF_X(x)\right| \le \int |g(x)|\,dF_X(x)$. Usaremos esta propriedade sem prová-la. Ela fornece uma prova alternativa para a propriedade FC1, pois implica que

$$|\varphi_X(t)| = \left|\int e^{itx}\,dF_X(x)\right| \le \int |e^{itx}|\,dF_X(x) =$$
$$= \int \{\cos^2(tx) + \mathrm{sen}^2(tx)\}^{1/2}\,dF_X(x) = 1.$$

FC2. *A função característica assume o valor* 1 *no ponto* 0: $\varphi_X(0) = 1$.

Prova. $\varphi_X(0) = Ee^{i\cdot 0\cdot X} = E1 = 1$. □

FC3. $\overline{\varphi_X(t)} = \varphi_X(-t)$, *onde* $\bar{c}$ *é o complexo conjugado de* c. (Se $c = x + iy$, o seu *complexo conjugado* é $\bar{c} = x - iy$.)

Prova. $\cos(-tX) = \cos(tX)$ e $\mathrm{sen}(-tX) = -\mathrm{sen}(tX)$, logo

$$\varphi_X(-t) = E\cos(-tX) + iE\,\mathrm{sen}(-tX) = E\cos(tX) - iE\,\mathrm{sen}(tX)$$
$$= \overline{Ee^{itX}}.\ \square$$

FC4. φ_X *é uniformemente contínua na reta.*

Prova. Uma função φ é *uniformemente contínua*, por definição, se para todo $\varepsilon > 0$ existe $\delta = \delta(\varepsilon) > 0$ tal que $|\varphi(t) - \varphi(s)| < \varepsilon$ quando $|t - s| < \delta$. Notemos que δ precisa depender apenas de ε. Se uma função é uniformemente contínua, então é contínua em todo ponto. A recíproca não vale: a função $f(x) = x^2$ é contínua na reta mas não é uniformemente contínua. Ora,

$$|\varphi(t) - \varphi(s)| = \left|\int (e^{itx} - e^{isx})\,dF_X(x)\right| \le \int |e^{itx} - e^{isx}|\,dF_X(x) =$$
$$= \int |e^{isx}|\cdot|e^{i(t-s)x} - 1|\,dF_X(x) =$$
$$= \int |e^{i(t-s)x} - 1|\,dF_X(x) \overset{def}{=} h(t-s).$$

Basta provar que $h(u) \to 0$ quando $u \to 0$, pois, nesse caso, dado $\varepsilon > 0$, $\exists\,\delta > 0$ tal que $|u| < \delta \Rightarrow |h(u)| < \varepsilon$, i.e., $|t - s| < \delta \Rightarrow |\varphi(t) - \varphi(s)| \le h(t-s) < \varepsilon$.

Mas $h(u) \to 0$ pelo Teorema da Convergência Dominada: $h(u) = E|e^{iuX} - 1|$, $0 \le |e^{iuX} - 1| \le 2$, 2 é integrável, e $\lim_{u\to 0} |e^{iuX(\omega)} - 1| = 0$ para todo $\omega \in \Omega$. (Tecnicamente, decorre do Teorema da Convergência Dominada que $h(u_n) \to 0$ para toda seqüência $(u_n)_{n\ge 1}$ tal que $u_n \to 0$. Logo $\lim_{u\to 0} h(u) = 0$.) □

FC5. *Se* X *e* Y *são independentes, então* $\varphi_{X+Y}(t) = \varphi_X(t)\cdot\varphi_Y(t)\ \forall t \in \mathbb{R}$.

Prova. Temos

$$\varphi_{X+Y}(t) = Ee^{it(X+Y)} = E(e^{itX}e^{itY}) = \text{(pela independência)} =$$
$$= Ee^{itX}\cdot Ee^{itY} = \varphi_X(t)\cdot\varphi_Y(t),$$

pois a Proposição 3.5 vale também para variáveis aleatórias complexas tais como e^{itX} e e^{itY}. Se você quiser, verifique o resultado usando seno e co-seno, como na prova da Proposição 6.2 adiante (§6.3). □

Observações. (1) Essa propriedade implica que o produto de duas funções características também é função característica. De fato, se X e Y são variáveis aleatórias, então $\varphi_X\cdot\varphi_Y$ é a função característica de uma variável aleatória Z cuja função de distribuição é a convolução $F_X * F_Y$. Como a função característica determina a

distribuição (veja a propriedade a seguir), podemos afirmar que

$$\varphi_Z = \varphi_X \cdot \varphi_Y \text{ se, e somente se, } F_Z = F_X * F_Y.$$

(2) Por indução, $\varphi_{X_1+\cdots+X_n} \models \prod_{k=1}^{n} \varphi_{X_k}$ se as X_k são independentes.

A equação significa $\varphi_{X_1+\cdots+X_n}(t) = \prod_{k=1}^{n} \varphi_{X_k}(t),\ \forall\, t \in \mathbb{R}$.

FC6. *A função característica de uma variável aleatória X determina a função de distribuição de X.* Já vimos a recíproca: a função característica é determinada pela função de distribuição, pois $\varphi_X(t) = \int e^{itx}\, dF_X(x)$. Como conseqüência, temos $F_X = F_Y \Leftrightarrow \varphi_X = \varphi_Y$, de modo que a função característica é uma *representação* da distribuição. (Duas funções são iguais se, e só se, assumem valores iguais em cada ponto; então $\varphi_X = \varphi_Y$ significa $\varphi_X(t) = \varphi_Y(t)\ \forall t \in \mathbb{R}$.)

Esta propriedade decorre da *fórmula da inversão: seja X uma variável aleatória, F sua função de distribuição, φ sua função característica. Se x e y são pontos de continuidade de F tais que $x < y$, então*

$$F(y) - F(x) = \frac{1}{2\pi} \lim_{u\to\infty} \int_{-u}^{u} \frac{e^{-itx} - e^{-ity}}{it}\, \varphi(t)\, dt.$$

(Para $x < y$ em geral, este limite é igual a $\dfrac{F(y)+F(y-)}{2} - \dfrac{(F(x)+F(x-))}{2}$.)

Prova da fórmula da inversão. Sejam x e y pontos de continuidade de F, $x < y$. Para $u > 0$, a integral do termo à direita é uma integral iterada:

$$I(u) \overset{def}{=} \int_{-u}^{u} \frac{e^{-itx} - e^{-ity}}{it}\, \varphi(t)\, dt = \int_{-u}^{u} \left\{ \int_{-\infty}^{\infty} \frac{e^{-itx} - e^{-ity}}{it}\, e^{itz}\, dF(z) \right\} dt.$$

Como

$$\lim_{t\to 0} \frac{e^{-itx} - e^{-ity}}{it} = y - x,$$

define-se o integrando como $y - x$ quando $t = 0$. Neste caso, o integrando da integral dupla é limitado e contínuo na região de integração ($z \in \mathbb{R}$, $t \in [-u, u]$). Já que o integrando é integrável nessa região (pois

$|f(z,t)| \le c \Rightarrow \int_{-u}^{u} \{\int_{-\infty}^{\infty} |f(z,t)|\, dF(z)\}\, dt \le c \int_{-u}^{u} \int_{-\infty}^{\infty} 1\, dF(z)\, dt = 2uc$), é permitido trocar a ordem de integração:

$$I(u) = \int_{-\infty}^{\infty} \left\{ \int_{-u}^{u} \frac{e^{it(z-x)} - e^{it(z-y)}}{it}\, dt \right\} dF(z).$$

Para todo $a \in \mathbb{R}$, $\dfrac{\cos(at)}{t}$ é função ímpar e $\dfrac{\operatorname{sen}(at)}{t}$ é par, logo

$$\int_{-u}^{u} \frac{\cos(at)}{t}\, dt = 0 \ \text{ e } \ \int_{-u}^{u} \frac{\operatorname{sen}(at)}{t}\, dt = 2\int_{0}^{u} \frac{\operatorname{sen}(at)}{t}\, dt,\ \forall a \in \mathbb{R}.$$

Por isso,

$$I(u) = \int_{-\infty}^{\infty} \left\{ 2\int_{0}^{u} \frac{\operatorname{sen} t\,(z-x)}{t}\, dt - 2\int_{0}^{u} \frac{\operatorname{sen} t\,(z-y)}{t}\, dt \right\} dF(z) =$$
$$= (\text{pelo teorema 3.1}) = Eg_u(X),$$

onde

$$g_u(z) = 2\int_{0}^{u} \frac{\operatorname{sen} t(z-x)}{t}\, dt - 2\int_{0}^{u} \frac{\operatorname{sen} t(z-y)}{t}\, dt.$$

Agora queremos aplicar o Teorema da Convergência Dominada quando $u \to \infty$. Para isto, basta provar que existe uma variável aleatória X_0 tal que $g_u(X) \to X_0$ em toda parte, e que as variáveis aleatórias $g_u(X)$ são dominadas por uma variável aleatória Y integrável. Primeiro, recordemos a integral clássica de Dirichlet:

$$\lim_{u\to\infty} \int_{0}^{u} \frac{\operatorname{sen} t}{t}\, dt = \frac{\pi}{2}.$$

Daí podemos calcular $\lim_{u\to\infty} \int_{0}^{u} \frac{\operatorname{sen} at}{t}\, dt$ para qualquer $a \in \mathbb{R}$:

$$(6.1)\begin{cases} a > 0 \Rightarrow \displaystyle\int_{0}^{u} \frac{\operatorname{sen}(at)}{t}\, dt = \int_{0}^{au} \frac{\operatorname{sen} t}{t}\, dt \Rightarrow \lim_{u\to\infty} \int_{0}^{u} \frac{\operatorname{sen}(at)}{t}\, dt = \frac{\pi}{2}, \\ a < 0 \Rightarrow \displaystyle\int_{0}^{u} \frac{\operatorname{sen}(at)}{t}\, dt = -\int_{0}^{u} \frac{\operatorname{sen}(-at)}{t}\, dt = -\int_{0}^{-au} \frac{\operatorname{sen} t}{t}\, dt \Rightarrow \\ \qquad \Rightarrow \displaystyle\lim_{u\to\infty} \int_{0}^{u} \frac{\operatorname{sen}(at)}{t}\, dt = -\frac{\pi}{2}, \\ a = 0 \Rightarrow \displaystyle\lim_{u\to\infty} \int_{0}^{u} \frac{\operatorname{sen}(at)}{t}\, dt = 0. \end{cases}$$

Em outras palavras (recorde que $x < y$),

$$\lim_{u\to\infty} g_u(z) = \begin{cases} 0 & \text{se } z < x \\ \pi & \text{se } z = x \\ 2\pi & \text{se } x < z < y \\ \pi & \text{se } z = y \\ 0 & \text{se } z > y. \end{cases}$$

Em termos de variáveis aleatórias, temos

$$g_u(X) \xrightarrow[u\to\infty]{} \pi I_{[X=x]} + 2\pi I_{[x<X<y]} + \pi I_{[X=y]} \stackrel{def}{=} X_0.$$

Temos também que as integrais $\int_0^u (\text{sen}(at))/t\, dt$ são uniformemente limitadas em u e a: para todo $a \in \mathbb{R}$ (veja as fórmulas (6.1)),

$$\left| \int_0^u \frac{\text{sen}(at)}{t}\, dt \right| \le \sup_{u>0} \left| \int_0^u \frac{\text{sen } t}{t}\, dt \right| \stackrel{def}{=} M < \infty,$$

pois a função $f(u) = \int_0^u (\text{sen } t)/t\, dt$ é contínua em $[0, \infty)$ ($f(0) = 0$) e tende a um número finito quando $u \to \infty$. Logo, as funções g_u são limitadas por $4M$ e as variáveis aleatórias $g_u(X)$ são dominadas pela variável aleatória $Y \stackrel{def}{=} 4M$. Aplicando o Teorema da Convergência Dominada, temos

$$\lim_{u\to\infty} Eg_u(X) = EX_0 = \pi P(X = x) + 2\pi P(x < X < y) + \pi P(X = y).$$

Dividindo por 2π, concluímos que

$$\lim_{u\to\infty} \frac{1}{2\pi} I(u) = \frac{1}{2} P(X = x) + P(x < X < y) + \frac{1}{2} P(X = y),$$

como queríamos demonstrar. □

Observação. "φ_X determina F_X" é o *Teorema da Unicidade*. É corolário da fórmula da inversão, pois esta implica que para todo $z \in \mathbb{R}$,

$$F_X(z) = \lim_{y \downarrow z} \lim_{x\to-\infty} \lim_{u\to\infty} \frac{1}{2\pi} \int_{-u}^{u} \frac{e^{-itx} - e^{-ity}}{it} \varphi_X(t)\, dt.$$

Portanto, a função de distribuição pode ser calculada a partir da função característica, e se X e Y têm a mesma função característica, então possuem a mesma função de distribuição.

(Ocorre que não costuma ser prático obter-se a função de distribuição através da fórmula da inversão.)

FC7. *A variável aleatória X tem distribuição simétrica em torno de zero se, e somente se, $\varphi_X(t)$ é real para todo t.* (Por definição, X tem *distribuição simétrica em torno de zero* se $P(X \le x) = P(X \ge -x)$, $\forall x \in \mathbb{R}$. As vezes, dizemos neste caso que X é simétrica em torno de zero.)

Prova. X é simétrica em torno de zero se, e somente se, $P(X \le x) = P(-X \le x)$ $\forall x$, i.e., $F_X = F_{-X}$ e X e $-X$ são identicamente distribuídas. Mas $F_X = F_{-X} \Leftrightarrow \varphi_X = \varphi_{-X}$, de modo que X é simétrica em torno de zero se, e somente se, para todo $t \in \mathbb{R}$,

$$\varphi_X(t) = \varphi_{-X}(t) = (\text{por FC3}) = \overline{\varphi_{-X}(-t)} = \overline{Ee^{i(-t)(-X)}} = \overline{Ee^{itX}} = \overline{\varphi_X(t)}.$$

Como $c = \bar{c}$ se, e só se, c é real, X é simétrica em torno de zero $\Leftrightarrow$ $\Leftrightarrow \forall t$, $\varphi_X(t)$ é real. □

FC8. *Se $Y = aX + b$, então $\varphi_Y(t) = e^{itb}\, \varphi_X(at)$.*

Prova. $\varphi_Y(t) = Ee^{it(aX+b)} = e^{itb}\, Ee^{iatX} = e^{itb}\, \varphi_X(at)$. □

FC9. *Se $E|X|^n < \infty$, então φ_X possui n derivadas contínuas e*

$$\varphi_X^{(k)}(t) = \int (ix)^k e^{itx}\, dF_X(x), \quad k = 1, 2, \ldots, n.$$

Em particular, $\varphi_X^{(k)}(0) = i^k EX^k$, de modo que a função característica é uma espécie de função geradora de momentos.

(*Observação.* $\varphi_X^{(k)}(0) = \left.\dfrac{d^k\varphi(t)}{dt^k}\right|_{t=0}$, a k-ésima derivada avaliada no ponto $t = 0$.)

Prova. Como $\varphi_X(t) = \int e^{itx}\, dF_X(x)$, a diferenciação formal resulta em $\varphi_X^{(k)}(t) = \int (ix)^k e^{itx}\, dF_X(x)$ e $\varphi_X^{(k)}(0) = \int (ix)^k\, dF_X(x) = i^k EX^k$. Resta justificar a diferenciação dentro da integral.

Suponhamos primeiro que X seja integrável; queremos provar que $\varphi'(t) = \int ixe^{itx}\, dF_X(x)$. Como para $h \ne 0$,

$$\frac{\varphi(t+h) - \varphi(t)}{h} = \int \frac{e^{i(t+h)x} - e^{itx}}{h}\, dF(x) = \int e^{itx} \cdot \frac{e^{ihx} - 1}{h}\, dF(x) =$$

$$= E\left\{ e^{itX} \cdot \frac{(e^{ihX} - 1)}{h} \right\},$$

vemos que o resultado decorre do Teorema da Convergência Dominada:

Como $(e^{ihx}-1)/h \to ix$ quando $h \to 0$, $\forall x \in \mathbb{R}$, temos

$$e^{itX}.\frac{(e^{ihX}-1)}{h} \xrightarrow[h\to 0]{} iX\,e^{itX}.$$

Mostremos que estas variáveis aleatórias estão uniformemente dominadas. Já que para todo x,

$$\left|\frac{e^{ihx}-1}{h}\right| = \left|\frac{\int_0^h ix\,e^{isx}\,ds}{h}\right| = |x|\cdot\left|\frac{\int_0^h e^{isx}\,ds}{h}\right| \le |x|,$$

pois $|e^{isx}|=1$, temos $\left|e^{itX}.\frac{(e^{ihX}-1)}{h}\right| \le |X|$. Como $|X|$ é integrável, o Teorema da Convergência Dominada implica que

$$\varphi'(t) = \lim_{h\to 0}\frac{\varphi(t+h)-\varphi(t)}{h} = \lim_{h\to 0} E\left\{e^{itX}.\frac{(e^{ihX}-1)}{h}\right\} = E(iX\,e^{itX}) =$$
$$= \int ix\,e^{itx}\,dF(x).$$

Decorre também desse teorema que $\varphi'(t)$ é contínua em t, pois $ix\,e^{itx} = \lim_{s\to t} ix\,e^{isx}$ e $|ix\,e^{isx}| = |x|$.

O restante da prova vem por indução em n, e é deixado para o leitor. □

Exemplos de funções características

EXEMPLO 1. Suponhamos $X \sim$ Poisson (λ). Então

$$\varphi_X(t) = Ee^{itX} = \sum_{n=0}^{\infty} e^{itn}e^{-\lambda}.\frac{\lambda^n}{n!} = e^{-\lambda}\sum_{n=0}^{\infty}\frac{(\lambda e^{it})^n}{n!} =$$
$$= e^{-\lambda}e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)}.$$

Vamos obter a média e variância, utilizando FC9:

$$iEX = \varphi'(0) = e^{\lambda(e^{it}-1)}\lambda e^{it}\cdot i\big|_{t=0} = i\lambda$$
$$i^2EX^2 = \varphi''(0) = d(i\lambda e^{-\lambda}e^{it+\lambda e^{it}})/dt\,\big|_{t=0} = i\lambda e^{-\lambda}e^{it+\lambda e^{it}}(i+\lambda e^{it}\cdot i)\big|_{t=0} =$$
$$= i\lambda e^{-\lambda}e^{\lambda}(i+\lambda i) = i^2\lambda(1+\lambda),$$

portanto $EX=\lambda$, $EX^2 = \lambda+\lambda^2$ e $\operatorname{Var} X = EX^2-(EX)^2 = \lambda$.

Outro resultado conhecido que podemos verificar facilmente por meio de funções características é o seguinte: se X e Y são independentes

com $X \sim$ Poisson (λ) e $Y =$ Poisson (ξ), então $X+Y \sim$ Poisson $(\lambda+\xi)$. Basta aplicar a propriedade FC5: $\varphi_{X+Y}(t) = \varphi_X(t)\,\varphi_Y(t) = e^{(\lambda+\xi)(e^{it}-1)}$, i.e., a função característica de $X+Y$ é a da distribuição Poisson $(\lambda+\xi)$. Pelo Teorema da Unicidade, $X+Y \sim$ Poisson $(\lambda+\xi)$.

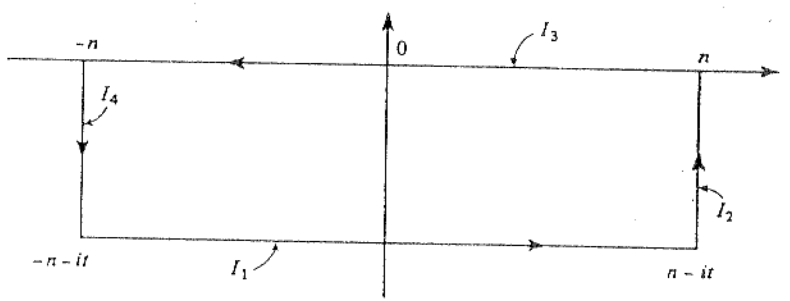
EXEMPLO 2. Seja $X \sim N(0,1)$. Então

$$\varphi_X(t) = \int e^{itx}\,dF_X(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{itx-x^2/2}\,dx = \text{(completando o quadrado)} = e^{-t^2/2}.\frac{1}{\sqrt{2\pi}}\int e^{-(x-it)^2/2}\,dx = e^{-t^2/2},$$

onde verificamos a última equação da seguinte maneira:

$$\int_{-\infty}^{\infty} e^{-(x-it)^2/2}\,dx = \lim_{n\to\infty}\int_{-n}^{n} e^{-(x-it)^2/2}\,dx = \text{(fazendo } z = x-it) =$$
$$= \lim_{n\to\infty}\int_{-n-it}^{n-it} e^{-z^2/2}\,dz,$$

onde o intervalo de integração nesta última integral (para n fixo e $t>0$) é I_1:



Como a função $f(z) = e^{-z^2/2}$ é analítica no plano complexo $\mathbb{C}$, o teorema de Cauchy diz que a integral de f sobre qualquer curva fechada dá zero (veja, por exemplo, Ahlfors [1], Teorema 2 do Cap. 4). Logo para $t>0$,

$$\int_{I_1} e^{-z^2/2}\,dz + \int_{I_2} e^{-z^2/2}\,dz + \int_{I_3} e^{-z^2/2}\,dz + \int_{I_4} e^{-z^2/2}\,dz = 0.$$

i.e., $$\int_{-n-it}^{n-it} e^{-z^2/2}\,dz = -\int_{I_2} e^{-z^2/2}\,dz + \int_{-n}^{n} e^{-t^2/2}\,dt - \int_{I_4} e^{-z^2/2}\,dz.$$

Como

$$\int_{-n}^{n} e^{-t^2/2}\,dt \xrightarrow[n\to\infty]{} \int_{-\infty}^{\infty} e^{-t^2/2}\,dt = \sqrt{2\pi},$$

basta provar que $\int_{I_2} e^{-z^2/2}\,dz$ e $\int_{I_4} e^{-z^2/2}\,dz$ tendem a zero quando $n\to\infty$ (assim estará provado que $\varphi_X(t) = e^{-t^2/2}$, $t>0$. Para $t<0$, o método é análogo, a única diferença sendo que os intervalos I_2 e I_4 invertem de sentido. Aliás, é suficiente notar que X simétrica $\Rightarrow \varphi_X(t) = \varphi_X(-t)$. Para $t=0$, o resultado é óbvio.).

Provaremos que $\int_{I_2} e^{-z^2/2}\,dz \to 0$; a prova de que $\int_{I_4} e^{-z^2/2}\,dz \to 0$ é análoga. Como o comprimento do intervalo I_2 é t, basta provar que $\max_{z\in I_2} |e^{-z^2/2}| \to 0$ quando $n\to\infty$.

Para $z = n - is$, onde $0 \le s \le t$, temos $z^2 = n^2 - s^2 - 2nsi$ e $e^{-z^2/2} = e^{(s^2-n^2)/2}\,e^{nsi}$. Como $|e^{it}| = 1\ \forall t$, temos $|e^{-z^2/2}| = e^{(s^2-n^2)/2}$. Portanto,

$$\max_{z\in I_2} |e^{-z^2/2}| = e^{(t^2-n^2)/2} \xrightarrow[n\to\infty]{} 0. \ \square$$

Segue-se da propriedade FC8 que se $Y = \sigma X + \mu$, onde $X \sim N(0,1)$, então $\varphi_Y(t) = e^{it\mu}\varphi_X(\sigma t) = e^{it\mu - \sigma^2 t^2/2}$. Em outras palavras, se $Y \sim N(\mu, \sigma^2)$, então $\varphi_Y(t) = e^{it\mu - \sigma^2 t^2/2}$. Novamente podemos obter os momentos da normal derivando a função característica, e ainda verificar que a soma de duas normais independentes também é normal.

EXEMPLO 3. Seja $\varphi(t) = \cos(at)$, onde $a>0$. Mostremos que φ é função característica, achando a distribuição correspondente. Já que assume valores reais, se φ fosse função característica de alguma variável aleatória X, então 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 à função característica φ. Portanto, φ é a função característica de X se, e somente se, $P(X=a) = 1/2 = P(X=-a)$.

6.2 Convergência em distribuição

Consideremos agora nosso terceiro tipo de convergência de variáveis aleatórias. Veremos adiante que é o mais fraco, no sentido de que é conseqüência tanto da convergência em probabilidade, como da convergência quase certa.

DEFINIÇÃO 6.2. Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias com, respectivamente, funções de distribuição $F, F_1, F_2, \ldots$. X_n *converge em distribuição* para X, quando $n\to\infty$, se $F_n(x) \to F(x)$ para todo x ponto de continuidade de F.

Notação: $X_n \xrightarrow{D} X$ ou $X_n \xrightarrow{D} F$. Também dizemos que X_n *converge em lei* para X e escrevemos $\mathscr{L}(X_n) \to \mathscr{L}(X)$.

O seguinte exemplo ilustra por que a definição requer convergência apenas nos pontos de continuidade de F.

EXEMPLO 4. Seja $(\Omega, \mathscr{A}, P)$ um espaço de probabilidade qualquer. Para todo $n = 1, 2, \ldots$, seja X_n a variável aleatória constante $1/n$ (i.e., $X_n(\omega) = 1/n\ \forall \omega \in \Omega$), e seja X igual à constante 0. Intuitivamente, X_n teria que convergir para X segundo qualquer critério razoável de convergência. De fato, é fácil ver que $X_n \to X$ quase certamente e $X_n \xrightarrow{P} X$.

A função de distribuição de X_n é

$$F_n(x) = \begin{cases} 1 \text{ se } x \ge \dfrac{1}{n} \\ 0 \text{ se } x < \dfrac{1}{n}; \end{cases}$$

e a de X é

$$F(x) = \begin{cases} 1 \text{ se } x \ge 0 \\ 0 \text{ se } x < 0. \end{cases}$$

Portanto, $x>0 \Rightarrow F_n(x) \to 1 = F(x)$ e $x<0 \Rightarrow F_n(x) \to 0 = F(x)$, de modo que $F_n(x)$ converge para $F(x)$ nos pontos de continuidade de F. Porém, no ponto $x=0$ temos $F_n(x) \to 0 \ne 1 = F(0)$.

Observação. Notemos que convergência em distribuição foi definida em termos das funções de distribuição; formalmente, não é preciso que as variáveis aleatórias estejam definidas no mesmo espaço de probabilidade.

Se $X_n \xrightarrow{D} X$, dizemos que F_{X_n} *converge fracamente* para F_X. Podemos considerar o conceito de convergência fraca de funções de distribuição sem levar em consideração qualquer variável aleatória: sejam $F, F_1, F_2, \ldots$ funções de distribuição. Então dizemos que F_n converge fracamente para F se $F_n(x) \to F(x)$ para todo x ponto de continuidade de F.

Nosso objetivo nesta seção é provar que $X_n \xrightarrow{D} X$ se, e somente se, $\varphi_{X_n}(t) \to \varphi_X(t)\ \forall\, t \in \mathbb{R}$, ou seja, que F_{X_n} converge fracamente para F_X se, e somente se, $\varphi_{X_n}(t) \to \varphi_X(t)\ \forall\, t \in \mathbb{R}$. Primeiro, provaremos a necessidade. (*Nota.* As provas dos dois teoremas seguintes são técnicas e poderão ser omitidas em uma primeira leitura.)

TEOREMA 6.1. (*Teorema de Helly-Bray*). *Sejam $F, F_1, F_2, \ldots$ funções de distribuição. Se F_n converge fracamente para F, então*

$$\int g(x)\, dF_n(x) \xrightarrow[n\to\infty]{} \int g(x)\, dF(x)$$

para toda função g contínua e limitada ($g\colon \mathbb{R} \to \mathbb{R}$).

Observação. Se $X_n \xrightarrow{D} X$, então decorre do teorema que $\int g(x)\, dF_{X_n}(x) \to \int g(x)\, dF_X(x)$ para toda g contínua e limitada, i.e., $Eg(X_n) \to Eg(X)$. Em particular, como as funções $\cos(tx)$ e $\operatorname{sen}(tx)$ são contínuas e limitadas para t fixo, temos $E\cos(tX_n) \xrightarrow[n\to\infty]{} E\cos(tX)$ e $E\operatorname{sen}(tX_n) \to E\operatorname{sen}(tX)$, de modo que $\varphi_{X_n}(t) \to \varphi_X(t)\ \forall\, t \in \mathbb{R}$.

Prova. Para $-\infty < a < b < \infty$,

$$\left|\int g dF_n - \int g dF\right| \le \left|\int g dF_n - \int_a^b g dF_n\right| + \left|\int_a^b g dF_n - \int_a^b g dF\right| + \left|\int_a^b g dF - \int g dF\right| \overset{def}{=} I + II + III.$$

Seja $c = \sup_{x\in\mathbb{R}} |g(x)| < \infty$ (por hipótese), e seja $\varepsilon > 0$.

Como $III = \left|\int_{-\infty}^a g dF + \int_b^\infty g dF\right| \le \int_{-\infty}^a c dF + \int_b^\infty c dF = c(F(a) + 1 - F(b))$, podemos escolher a e b pontos de continuidade de F tais que

$$III \le c(F(a) + 1 - F(b)) < \varepsilon,$$

pois $F(a) + 1 - F(b) \to 0$ quando $a \to -\infty$ e $b \to +\infty$, e basta escolher a suficientemente pequeno e b suficientemente grande (os pontos de continuidade são densos). Para esses valores de a e b, temos

$$I \le \int_{-\infty}^a c dF_n + \int_b^\infty c dF_n = c(F_n(a) + 1 - F_n(b)) \xrightarrow[n\to\infty]{} c(F(a) + 1 - F(b)),$$

logo $I + III < 2\varepsilon$ para n suficientemente grande.

Para terminar a prova, basta provar $II < 3\varepsilon\ \forall n$ suficientemente grande, pois neste caso temos $\left|\int g dF_n - \int g dF\right| < 5\varepsilon$ para todo n suficientemente grande, para todo ε, i.e., $\int g dF_n \to \int g dF$. Para tanto, sejam a e b os pontos já escolhidos, e consideremos a integral II.

Já que g é uniformemente contínua em $[a, b]$, podemos escolher $x_0, x_1, \ldots, x_N$ tais que (i) $a = x_0 < x_1 < \ldots < x_N = b$, (ii) os x_i sejam pontos de continuidade de F, e (iii) $|g(x) - g(x_i)| < \varepsilon$ para todo $x \in \in [x_i, x_{i+1}]$, $i = 0, 1, \ldots, N-1$. Então:

$$m_{ni} \overset{def}{=} (g(x_i) - \varepsilon)\{F_n(x_{i+1}) - F_n(x_i)\} \le \int_{x_i}^{x_{i+1}} g(x)\, dF_n(x) \le$$

$$\le (g(x_i) + \varepsilon)\{F_n(x_{i+1}) - F_n(x_i)\} \overset{def}{=} M_{ni},$$

$$m_i \overset{def}{=} (g(x_i) - \varepsilon)\{F(x_{i+1}) - F(x_i)\} \le \int_{x_i}^{x_{i+1}} g(x)\, dF(x) \le$$

$$\le (g(x_i) + \varepsilon)\{F(x_{i+1}) - F(x_i)\} \overset{def}{=} M_i.$$

Portanto,

$$m_{ni} - M_i \le \int_{x_i}^{x_{i+1}} g dF_n - \int_{x_i}^{x_{i+1}} g dF \le M_{ni} - m_i,$$

para $i = 0, 1, \ldots, N-1$. Somando, temos

$$\sum_{i=0}^{N-1} (m_{ni} - M_i) \le \int_a^b g dF_n - \int_a^b g dF \le \sum_{i=0}^{N-1} (M_{ni} - m_i).$$

Quando $n \to \infty$, temos $m_{ni} \to m_i$ e $M_{ni} \to M_i$, porque os x_i são pontos de continuidade de F e F_n converge fracamente a F. Logo,

$$\sum_{i=0}^{N-1} (m_{ni} - M_i) \xrightarrow[n\to\infty]{} \sum_{i=0}^{N-1} (m_i - M_i) = -2\varepsilon(F(b) - F(a)) \ge -2\varepsilon,$$

$$\sum_{i=1}^{N-1} (M_{ni} - m_i) \xrightarrow[n\to\infty]{} \sum_{i=0}^{N-1} (M_i - m_i) = 2\varepsilon(F(b) - F(a)) \le 2\varepsilon.$$

Daí, temos para todo n suficientemente grande,

$$-3\varepsilon < \int_a^b g dF_n - \int_a^b g dF < 3\varepsilon. \ \square$$

A recíproca do Teorema 6.1 é válida, i.e., se $\int g dF_n \to \int g dF \ \forall g$ contínua e limitada então F_n converge fracamente a F, e às vezes toma-se essa condição como definição de convergência fraca. (*Nota.* Convergência fraca de funções de distribuição nada tem a ver com a Lei Fraca dos Grandes Números.) Mas ocorre que basta a convergência das funções características associadas às funções de distribuição. É evidente como definir a função característica associada a uma função de distribuição F: define-se $\varphi(t) = \int e^{itx} dF(x)$. Assim definida, φ é função característica de alguma variável aleatória (por quê?).

TEOREMA 6.2. (*Teorema da Continuidade de Paul Lévy*). *Sejam F_1, F_2, ... funções de distribuição e φ_1, φ_2, ..., respectivamente, suas funções características. Se φ_n converge pontualmente para um limite φ e se φ é contínua no ponto zero, então*

(a) existe uma função de distribuição F tal que $F_n \to F$ fracamente e

(b) φ é a função característica de F.

Observação. Cabe destacar agora que os Teoremas 6.1 e 6.2 implicam que $X_n \xrightarrow{D} X \Leftrightarrow \varphi_{X_n} \to \varphi_X$. Mas o Teorema da Continuidade é mais forte do que a suficiência dessa proposição, porque diz que o limite de uma seqüência de funções características também é uma função característica, contanto que seja contínuo no ponto zero.

Prova. Sob as hipóteses, (a) implica (b), por Helly-Bray. Para provar que F_n converge fracamente para alguma função de distribuição, vamos provar que para toda seqüência de funções de distribuição satisfazendo as condições do teorema, existem uma subseqüência F_{n_1}, F_{n_2}, ... e uma função de distribuição F tais que $F_{n_j} \to F$ fracamente, quando $j \to \infty$. (Para ver que é suficiente provar isso, suponha que F_n *não* convirja fracamente para F, onde $F_{n_j} \to F$ fracamente. Então, existirão x, ponto de continuidade de F e uma subseqüência $F_{1'}, F_{2'}, \ldots$ tais que $F_{n'}(x) \to a \neq F(x)$. Como essa subseqüência também satisfará as condições do teorema, existirão uma subseqüência $F_{1''}, F_{2''}, \ldots$ [subseqüência de $F_{1'}, F_{2'}, \ldots$] e uma função de distribuição G tais que $F_{n''} \to G$ fracamente. Mas F e G terão a mesma função característica

$(= \varphi)$, logo $F = G$ e, em particular, $F_{n''}(x) \to G(x) = F(x)$. Mas $F_{n''}(x) \to a \neq F(x)$. Absurdo, logo F_n converge fracamente para F.)

Provaremos que

(i) existem uma subseqüência $F_{n_1}, F_{n_2}, \ldots$ e uma função $F: \mathbb{R} \to [0, 1]$ tais que F é não-decrescente e contínua à direita e $F_{n_j}(x) \to F(x)$, quando $j \to \infty$, para todo x ponto de continuidade de F, e

(ii) a F do item (i) é função de distribuição.

Observação. O item (i) é o *Teorema de Compacidade Fraca de Helly.*

Prova de (*i*). Sejam $r_1, r_2, \ldots$ os racionais da reta. Usando o método da diagonal, escolhemos uma seqüência $1 \leq n_1 < n_2 < \ldots$ de inteiros positivos tais que $F_{n_j}(r_k)$ converge, quando $j \to \infty$, para cada k fixo. Chamemos o limite de $F(r_k)$, de modo que $F_{n_j}(r_k) \xrightarrow[j \to \infty]{} F(r_k) \ \forall k$. É obvio que $0 \leq F(r_k) \leq 1$ e 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. Temos, contudo, que $F_{n_j}(x) \xrightarrow[j \to \infty]{} F(x)$ para todo x ponto de continuidade de F. (Pois: suponha x ponto de continuidade de F e sejam r', r'' racionais tais que $r' < x < r''$ e $F(r'') - \varepsilon < F(x) < F(r') + \varepsilon$. Então,

$$F(x) - \varepsilon < F(r') = \lim_{j \to \infty} F_{n_j}(r') \leq \liminf_{j \to \infty} F_{n_j}(x) \leq \limsup_{j \to \infty} F_{n_j}(x) \leq$$
$$\leq \lim_{j \to \infty} F_{n_j}(r'') = F(r'') < F(x) + \varepsilon.$$

Como ε é arbitrário, temos $F_{n_j}(x) \to F(x)$ quando $j \to \infty$.)

Podemos redefinir F, se necessário, nos seus pontos de descontinuidade para torná-la contínua à direita. Assim, (i) esta provado.

Prova de (*ii*). Resta provar $F(+\infty) = 1$, $F(-\infty) = 0$. Para g uma função característica qualquer, definamos a função característica integrada $\hat{g}$: se G é a função de distribuição correspondente a g,

$$\hat{g}(t) = \int_0^t g(s)\, ds = \int_0^t \int_{-\infty}^{\infty} e^{isx}\, dG(x)\, ds = \int_{-\infty}^{\infty} \int_0^t e^{isx}\, ds\, dG(x) =$$
$$= \int_{-\infty}^{\infty} \frac{e^{itx} - 1}{ix}\, dG(x),$$

onde justifica-se a troca da ordem de integração pelo fato do integrando ser limitado.

Para t fixo, a função $(e^{itx}-1)/(ix)$ é limitada e contínua (definir igual a t em $x=0$), e tende a zero quando $x\to+\infty$ ou $x\to-\infty$. Uma prova análoga à de Helly-Bray mostra que

$$\int_{-\infty}^{\infty}\frac{e^{itx}-1}{ix}\,dF_{n_j}(x)\xrightarrow[j\to\infty]{}\int_{-\infty}^{\infty}\frac{e^{itx}-1}{ix}\,dF(x).$$

Portanto,

$$\int_0^t\varphi_{n_j}(s)\,ds\xrightarrow[j\to\infty]{}\int_0^t\int_{-\infty}^{\infty}e^{isx}\,dF(x)\,ds.$$

Mas $\varphi_{n_j}\to\varphi$, φ contínua em zero, implica que φ é limitada e mensurável, logo pelo Teorema da Convergência Dominada (ou pelo teorema de Arzelà – veja o §3.7 – aplicado às partes real e imaginária das φ_n)

$$\int_0^t\varphi_{n_j}(s)\,ds\xrightarrow[j\to\infty]{}\int_0^t\varphi(s)\,ds.$$

Dividindo os limites iguais por t, temos

$$\frac{1}{t}\int_0^t\varphi(s)\,ds=\frac{1}{t}\int_0^t\int_{-\infty}^{\infty}e^{isx}\,dF(x)\,ds,\quad t\neq 0.$$

Fazendo $t\to 0$ e usando a continuidade em $s=0$ das duas funções $\varphi(s)$ e $\int e^{isx}\,dF(x)$ (esta é contínua pelo Teorema da Convergência Dominada), temos

$$\varphi(0)=\int_{-\infty}^{\infty}1\,dF(x)=F(+\infty)-F(-\infty).$$

Como $\varphi(0)=\lim_{n\to\infty}\varphi_n(0)=1$, segue-se que $F(+\infty)=1$, $F(-\infty)=0$. □

COROLÁRIOS. (*a*) *Sejam $X_1, X_2, \ldots$ variáveis aleatórias. Se $\varphi_{X_n}(t)\to\varphi(t)$ $\forall t\in\mathbb{R}$, e se φ é contínua no ponto zero, então φ é função característica de alguma variável aleatória, digamos $\varphi=\varphi_X$, e $X_n\xrightarrow{D}X$.*

(*b*) *Se $\varphi_{X_n}(t)\to e^{-t^2/2}$ $\forall t$, então $X_n\xrightarrow{D}N(0,1)$.*

Notação. "$X_n\xrightarrow{D}N(0,1)$" indica que X_n converge em distribuição para uma variável aleatória X que possui distribuição $N(0,1)$. Mas não é necessário que X seja explicitamente definida, e podemos interpretar a expressão "$X_n\to N(0,1)$" como indicativa da convergência fraca de F_{X_n} para Φ, a função de distribuição da $N(0,1)$. É conveniente dizermos, neste caso, que X_n *converge em distribuição para a (distribuição) normal-padrão.* Vale uma interpretação análoga para as expressões "$X_n\xrightarrow{D}$ Poisson (λ)", "$X_n\xrightarrow{D}\chi^2(1)$", etc..

APLICAÇÃO. (*O Teorema Central do Limite para variáveis aleatórias independentes e identicamente distribuídas.*) *Sejam $X_1, X_2, \ldots$ independentes e identicamente distribuídas, com média comum μ e variância comum σ^2, onde $0<\sigma^2<\infty$. Seja $S_n=X_1+\ldots+X_n$. Então*

$$\frac{S_n-ES_n}{\sqrt{\operatorname{Var}S_n}}\xrightarrow{D}N(0,1),\ \textit{i.e.},\ \frac{S_n-n\mu}{\sigma\sqrt{n}}\xrightarrow{D}N(0,1).$$

Prova. Podemos supor, sem perda de generalidade, que $\mu=0$. Pelo Teorema de Paul Lévy, basta mostrar que

$$\varphi_{S_n/(\sigma\sqrt{n})}(t)\to e^{-t^2/2},\ \forall t\in\mathbb{R}.$$

Aplicando sucessivamente as propriedades FC8, FC5 e o fato de que as variáveis X_k são identicamente distribuídas, temos

$$\varphi_{S_n/(\sigma\sqrt{n})}(t)=\varphi_{S_n}\left(\frac{t}{\sigma\sqrt{n}}\right)=\prod_{k=1}^{n}\varphi_{X_k}\left(\frac{t}{\sigma\sqrt{n}}\right)=\varphi^n\left(\frac{t}{\sigma\sqrt{n}}\right),$$

onde $\varphi=\varphi_{X_1}$.

Já que $EX_1^2<\infty$, decorre de FC9 que φ possui duas derivadas contínuas. Pela fórmula de Taylor,

$$\varphi(t)=\varphi(0)+\varphi'(0)\cdot t+\varphi''(\theta(t))\cdot\frac{t^2}{2},$$

onde $|\theta(t)|\le|t|$. Portanto,

$$\varphi(t)=\varphi(0)+\varphi'(0)\cdot t+\varphi''(0)\cdot\frac{t^2}{2}+\frac{t^2}{2}\{\varphi''(\theta(t))-\varphi''(0)\},$$

com $\varphi''(\theta(t))-\varphi''(0)\to 0$ quando $t\to 0$. Como $\varphi(0)=1$ e, por FC9, $\varphi'(0)=i\mu=0$ e $\varphi''(0)=i^2EX_1^2=-EX_1^2=-\sigma^2$, temos

$$\varphi(t)=1-\frac{\sigma^2t^2}{2}+\frac{t^2}{2}e(t),$$

onde $\lim_{t\to 0} e(t) = 0$. Por isso, para t fixo,

$$\varphi^n\left(\frac{t}{\sigma\sqrt{n}}\right) = \left[1 - \frac{t^2}{2n} + \frac{t^2}{2\sigma^2 n} e\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n =$$
$$= \left[1 - \frac{t^2}{2n}\left\{1 - \frac{1}{\sigma^2} e\left(\frac{t}{\sigma\sqrt{n}}\right)\right\}\right]^n \xrightarrow[n\to\infty]{} e^{-t^2/2},$$

pois $1 - \frac{1}{\sigma^2} e\left(\frac{t}{\sigma\sqrt{n}}\right) \to 1$ quando $n \to \infty$ e, para números complexos, $c_n \to c \Rightarrow \left(1 + \frac{c_n}{n}\right)^n \to e^c$. (Veja o exercício 12.) □

COROLÁRIO. (*Teorema Central do Limite de De Moivre e Laplace.*) *Seja S_n o número de sucessos em n ensaios binomiais independentes, com probabilidade p de sucesso em cada ensaio, onde $0 < p < 1$. Então*

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{D} N(0, 1).$$

Prova. Seja X_n o indicador de sucesso no n-ésimo ensaio. Então $S_n = X_1 + \ldots + X_n$ e $X_1, X_2 \ldots$ são independentes e, identicamente distribuídas com média $\mu = p$ e variância $\sigma^2 = p(1-p)$. □

§ 6.3 Função característica de um vetor aleatório

Nesta seção, daremos uma introdução bem abreviada às funções características multidimensionais.

DEFINIÇÃO 6.3. Seja $\underset{\sim}{X} = (X_1, \ldots, X_k)$ um vetor aleatório k-dimensional. A *função característica* de $\underset{\sim}{X}$ é a função $\varphi: \mathbb{R}^k \to \mathbb{C}$ definida por

$$\varphi(t_1, \ldots, t_k) = \varphi_{\underset{\sim}{X}}(t_1, \ldots, t_k) = Ee^{i\sum_{j=1}^{k} t_j X_j} = Ee^{i\underset{\sim}{t}\cdot\underset{\sim}{X}}$$

onde $\underset{\sim}{t}\cdot\underset{\sim}{X}$ representa o produto interno (produto escalar) dos vetores $\underset{\sim}{t} = (t_1, \ldots t_k)$ e $\underset{\sim}{X} = (X_1, \ldots, X_k)$. $\varphi_{\underset{\sim}{X}} = \varphi_{X_1, \ldots, X_k}$ é também chamada *função característica conjunta* de $X_1, \ldots, X_k$.

A função característica multivariada tem propriedades análogas à todas as propriedades enunciadas para a função característica de uma

variável aleatória. Por exemplo, propriedades FC1 a FC4, FC7 e FC8 são válidas com as óbvias modificações (a reta é substituída por $\mathbb{R}^k$). Para FC5, supõe-se $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ sejam vetores de mesma dimensão, i.e., $\underset{\sim}{X} = (X_1, \ldots, X_k)$ e $\underset{\sim}{Y} = (Y_1, \ldots, Y_k)$. Sob esta condição, a independência de $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ implica que $\varphi_{\underset{\sim}{X}+\underset{\sim}{Y}}(\underset{\sim}{t}) = \varphi_{\underset{\sim}{X}}(\underset{\sim}{t})\,\varphi_{\underset{\sim}{Y}}(\underset{\sim}{t})\ \forall \underset{\sim}{t} \in \mathbb{R}^k$. (Dizemos que os vetores $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ são *independentes* se $P(\underset{\sim}{X} \in B_1,\ \underset{\sim}{Y} \in B_2) = P(\underset{\sim}{X} \in B_1) \cdot P(\underset{\sim}{Y} \in B_2)\ \forall B_1 \in \mathscr{B}^k, B_2 \in \mathscr{B}^k$.)

Quanto a FC6, existe uma fórmula da inversão para a função característica multidimensional (veja Gnedenko [11], §40, Teorema 4). Podemos usar essa fórmula, mais uma vez, para provar o

Teorema da Unicidade: Se $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ forem vetores aleatórios k-dimensionais tais que $\varphi_{\underset{\sim}{X}}(\underset{\sim}{t}) = \varphi_{\underset{\sim}{Y}}(\underset{\sim}{t})\ \forall \underset{\sim}{t} \in \mathbb{R}^k$, então $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ têm a mesma distribuição. Em outras palavras, a função característica determina a distribuição, e podemos escrever: $\varphi_{\underset{\sim}{X}} = \varphi_{\underset{\sim}{Y}} \Leftrightarrow F_{\underset{\sim}{X}} = F_{\underset{\sim}{Y}}$. (Uma elegante prova alternativa do Teorema da Unicidade é fornecida pelo Teorema de Stone-Weierstrass. Veja Billingsley [4], Teorema 7.5.)

É possível, também, generalizar FC9 para o caso multivariado. Neste caso, podemos obter momentos mistos, ou momentos-produto, a partir da função característica multidimensional; veja os exercícios 26 e 27(b) para um exemplo.

Outra propriedade da função característica multivariada é a seguinte:

FC10. *Para obter a função característica de qualquer distribuição marginal, basta fazer todos os termos "extras" iguais a zero.* Por exemplo, para as variáveis aleatórias X, Y e Z, temos $Ee^{i(xX+yY)} = Ee^{i(xX+yY+0\cdot Z)}$, i.e.,

$$\varphi_{X,Y}(x, y) = \varphi_{X,Y,Z}(x, y, 0),\ (x, y) \in \mathbb{R}^2.$$

A definição de convergência em distribuição de vetores aleatórios é análoga à Definição 6.2:

DEFINIÇÃO 6.4. Sejam $\underset{\sim}{X}_n = (X_{n1}, \ldots, X_{nk})$, $\underset{\sim}{X} = (X_{01}, \ldots, X_{0k})$ vetores aleatórios k-dimensionais. *$\underset{\sim}{X}_n$ converge para $\underset{\sim}{X}$ em distribuição* se $F_{\underset{\sim}{X}_n}(x_1, \ldots, x_k) \xrightarrow[n\to\infty]{} F_{\underset{\sim}{X}}(x_1, \ldots, x_k)\ \ \forall \underset{\sim}{x} = (x_1, \ldots, x_k)$ ponto de continuidade de $F_{\underset{\sim}{X}}$. *Notação*: $\underset{\sim}{X}_n \xrightarrow{D} \underset{\sim}{X}$.

Como no caso unidimensional, temos convergência em distribuição se, e somente se, as funções características convergem:

TEOREMA 6.3. $\underset{\sim}{X}_n \xrightarrow{D} \underset{\sim}{X}$ se, e somente se, $\varphi_{\underset{\sim}{X}_n} \to \varphi_{\underset{\sim}{X}}$, i.e.,

$$\varphi_{\underset{\sim}{X}_n}(t_1, \ldots, t_k) \to \varphi_{\underset{\sim}{X}}(t_1, \ldots, t_k) \quad \forall\, (t_1, \ldots, t_k) \in \mathbb{R}^k.$$

Prova. Omitida. (Veja Billingsley [4], §3 e Teorema 7.6.) □

Em geral, não é fácil provar convergência de funções características k-dimensionais, para $k > 1$ (e uma aplicação direta da Definição 6.4 é menos prática ainda). Mas há um artifício, devido a Cramér e Wold, que reduz o problema de convergência em distribuição ao caso unidimensional. Para melhor entender esse artifício, observe primeiro que o Teorema da Unicidade diz que a distribuição de $\underset{\sim}{X} = (X_1, \ldots, X_k)$ é determinada pela função característica de $\underset{\sim}{X}$. Mas, por definição, a função característica de $\underset{\sim}{X}$ é determinada pelas distribuições das combinações lineares $\sum_{j=1}^{k} t_j X_j$. Em outras palavras, a distribuição conjunta de $X_1, \ldots, X_k$ é determinada pelas distribuições das combinações lineares $\sum_{j=1}^{k} t_j X_j$: distribuições unidimensionais determinam a distribuição conjunta! Então é bem natural perguntar: será que convergência em distribuição de vetores aleatórios equivale à convergência em distribuição (unidimensional) de todas as combinações lineares das coordenadas? A resposta é afirmativa:

PROPOSIÇÃO 6.1. *(Cramér-Wold). Sejam* $\underset{\sim}{X}_n = (X_{n1}, \ldots, X_{nk})$ *e* $\underset{\sim}{X} = (X_{01}, \ldots, X_{0k})$ *vetores aleatórios k-dimensionais.* $\underset{\sim}{X}_n \xrightarrow{D} \underset{\sim}{X}$ *se, e só se,*

$$\sum_{j=1}^{k} t_j X_{nj} \xrightarrow{D} \sum_{j=1}^{k} t_j X_{0j}, \text{ quando } n \to \infty,$$

para todo $(t_1, \ldots, t_k) \in \mathbb{R}^k$.

Prova. Suponhamos primeiro que $\sum_{j=1}^{k} t_j X_{nj} \xrightarrow{D} \sum_{j=1}^{k} t_j X_{0j} \quad \forall\, (t_1, \ldots, t_k)$.

Então,

$$\varphi_{\underset{\sim}{X}_n}(t_1, \ldots, t_k) = Ee^{i\sum_{j=1}^{k} t_j X_{nj}} = \varphi_{\sum_{j=1}^{k} t_j X_{nj}}(1) \xrightarrow{n \to \infty} \varphi_{\sum_{j=1}^{k} t_j X_{0j}}(1) = \varphi_{\underset{\sim}{X}}(t_1, \ldots, t_k).$$

onde utilizamos o Teorema de Helly-Bray. Como $\varphi_{\underset{\sim}{X}_n} \to \varphi_{\underset{\sim}{X}}$, decorre do Teorema 6.3 que $\underset{\sim}{X}_n \xrightarrow{D} \underset{\sim}{X}$.

Agora, suponhamos que $\underset{\sim}{X}_n \xrightarrow{D} \underset{\sim}{X}$. Para $(t_1, \ldots, t_k) \in \mathbb{R}^k$, queremos provar que $\sum_{j=1}^{k} t_j X_{nj} \xrightarrow{D} \sum_{j=1}^{k} t_j X_{0j}$. Para tanto, basta provarmos que

$$\varphi_{\sum_{j=1}^{k} t_j X_{nj}}(t) \xrightarrow{n \to \infty} \varphi_{\sum_{j=1}^{k} t_j X_{0j}}(t) \quad \forall\, t \in \mathbb{R}.$$

Mas, utilizando a outra metade do Teorema 6.3, temos

$$\varphi_{\sum t_j X_{nj}}(t) = Ee^{it\sum_j t_j X_{nj}} = Ee^{i\sum_j (tt_j) X_{nj}} =$$
$$= \varphi_{\underset{\sim}{X}_n}(tt_1, \ldots, tt_k) \xrightarrow{n \to \infty} \varphi_{\underset{\sim}{X}}(tt_1, \ldots, tt_k) = \varphi_{\sum t_j X_{0j}}(t). \quad \square$$

Veremos uma aplicação desta proposição no próximo capítulo, na demonstração da versão multivariada do Teorema Central do Limite.

Terminaremos nossa discussão de funções características multidimensionais considerando um critério para independência de (algumas) coordenadas do vetor. Suponhamos primeiro que X e Y sejam variáveis aleatórias. Então ocorre que X e Y são independentes se, e só se, a função característica do vetor (X, Y) fatora: $\varphi_{X,Y}(x, y) = \varphi_X(x)\, \varphi_Y(y) \, \forall\, (x, y) \in \mathbb{R}^2$. Novamente, temos a regra produto no caso de independência. Mais geralmente, temos o mesmo resultado quando $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ são vetores (de possivelmente diferentes dimensões):

PROPOSIÇÃO 6.2. *Sejam* $\underset{\sim}{X} = (X_1, \ldots, X_m)$ *e* $\underset{\sim}{Y} = (Y_1, \ldots, Y_n)$ *vetores aleatórios, onde* $m \geq 1$, $n \geq 1$. $\underset{\sim}{X}$ *e* $\underset{\sim}{Y}$ *são independentes se, e só se,*

$$\varphi_{X_1, \ldots, X_m, Y_1, \ldots, Y_n}(x_1, \ldots, x_m, y_1, \ldots, y_n) = \varphi_{\underset{\sim}{X}}(x_1, \ldots, x_m)\, \varphi_{\underset{\sim}{Y}}(y_1, \ldots, y_n)$$
$$\forall\, (x_1, \ldots, x_m) \in \mathbb{R}^m, \ (y_1, \ldots, y_n) \in \mathbb{R}^n.$$

Observação. Os vetores $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ são *independentes*, por definição, se

$$P(\underset{\sim}{X} \in B_1, \underset{\sim}{Y} \in B_2) = P(\underset{\sim}{X} \in B_1)\, P(\underset{\sim}{Y} \in B_2)$$

para todo $B_1 \in \mathscr{B}^m$, $B_2 \in \mathscr{B}^n$ (i.e., borelianos em $\mathbb{R}^m$ e $\mathbb{R}^n$).

Prova. Suponhamos primeiro que $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ sejam variáveis aleatórias X e Y $(m = 1, n = 1)$, com X e Y independentes. Então, temos

$$\begin{aligned}
\varphi_{X,Y}(x, y) &= Ee^{i(xX + yY)} = E(e^{ixX} e^{iyY}) = \\
&= E\left[\{\cos(xX) + i \operatorname{sen}(xX)\}\{\cos(yY) + i \operatorname{sen}(yY)\}\right] = \\
&= E\left[\cos(xX)\cos(yY) - \operatorname{sen}(xX)\operatorname{sen}(yY)\right] + \\
&\quad + iE\left[\cos(xX)\operatorname{sen}(yY) + \operatorname{sen}(xX)\cos(yY)\right] \\
&= (\text{por linearidade e independência}) = \\
&= \left[E\cos(xX) + iE\operatorname{sen}(xX)\right] \cdot \left[E\cos(yY) + iE\operatorname{sen}(yY)\right] \\
&= Ee^{ixX} \cdot Ee^{iyY} = \varphi_X(x)\, \varphi_Y(y).
\end{aligned}$$

para todo (x, y). Reciprocamente, suponha que $\varphi_{X,Y}(x, y) = \varphi_X(x)\varphi_Y(y)$ para todo $x \in \mathbb{R}$, $y \in \mathbb{R}$. Então a independência de X e Y é conseqüência 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 seria absurdo (por hipótese, possuem a função característica desejada). Logo, são independentes.

A prova no caso geral é análogo, e é deixada para o leitor (exercício 25). □

Um resultado semelhante vale para um número finito qualquer de vetores aleatórios (a prova é a mesma). Consideremos o caso mais simples em que $X_1, \ldots, X_n$ são variáveis aleatórias. Então, temos $X_1, \ldots, X_n$ independentes se, e só se,

$$\varphi_{X_1,\ldots,X_n}(t_1, \ldots, t_n) = \prod_{j=1}^{n} \varphi_{X_j}(t_j) \quad \forall\, (t_1, \ldots, t_n) \in \mathbb{R}^n.$$

§6.4 Observações e complementos

(i) *Convergência em distribuição nos casos discreto e contínuo.* Suponhamos que as variáveis aleatórias $X_1, X_2, \ldots$ e X sejam todas discretas ou todas absolutamente contínuas. Nestes casos, a convergência em distribuição de X_n para X decorre da convergência pontual das funções de probabilidade ou densidades. Consideremos primeiro o caso discreto.

PROPOSIÇÃO 6.3. *Sejam $X_1, X_2, \ldots$ e X variáveis aleatórias tomando somente os valores 0, 1, 2, ..., e sejam $p_1(k), p_2(k), \ldots$ e $p(k)$, respectivamente, as funções de probabilidade (i.e., $p_n(k) = P(X_n = k)$). Então, $X_n \xrightarrow{D} X$ se, e somente se, $p_n(k) \to p(k)$ quando $n \to \infty$ para todo $k = 0, 1, 2, \ldots$.*

Prova. Os pontos $x = 1/2,\ 3/2,\ 5/2, \ldots$ são pontos de continuidade de F_X. Portanto, se $X_n \xrightarrow{D} X$, então $F_{X_n}(k + 1/2) \to F_X(k + 1/2)$ quando $n \to \infty$ para $k = 0, 1, 2, \ldots$, e

$$p_n(k) = F_{X_n}(k + 1/2) - F_{X_n}(k - 1/2) \xrightarrow[n\to\infty]{} F_X(k + 1/2) - F_X(k - 1/2) = p(k)$$

para $k = 0, 1, 2, \ldots$ (quando $k = 0$, temos $F_{X_n}(-1/2) = F_X(-1/2) = 0$).

Reciprocamente, se $p_n(k) \to p(k)$ para todo k, então $F_{X_n}(x) \to F_X(x)$ para *todo* $x \in \mathbb{R}$, pois se $x \geq 0$

$$F_{X_n}(x) = \sum_{k=0}^{[x]} p_n(k) \xrightarrow[n\to\infty]{} \sum_{k=0}^{[x]} p(k) = F_X(x),$$

onde $[x]$ é a parte inteira de x (se $x < 0$, então $F_{X_n}(x) = F_X(x) = 0$). □

Nota. No caso geral, com k substituído por x_k, a condição da Proposição 6.3 ainda é suficiente, i.e., $p_n(x_k) \to p(x_k)\ \forall\, k \Rightarrow X_n \xrightarrow{D} X$. É necessária se os valores possíveis das X_n e X são isolados. Não é necessária no caso geral: para um contra-exemplo, sejam $X_n = 1/n$, $X = 0$.

EXEMPLO 4. (Convergência da distribuição hipergeométrica para a binomial.) Seja X_N uma variável aleatória com distribuição hipergeométrica tendo função de probabilidade

$$P(X_N = k) = \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \ldots, n,$$

onde N, D e n são inteiros não-negativos, $D \leq N$ e $n \leq N$. Lembremos que $\binom{D}{k} = 0$ se $k > D$. Tal distribuição serve de modelo, por exemplo, para o número de itens defeituosos em uma amostra de tamanho n, extraída sem reposição de um lote de N itens contendo D defeituosos.

Quando D e $N - D$ são grandes e n pequeno, X_N tem aproximadamente distribuição $b(n, D/N)$. Este resultado é intuitivo, porque nessas condições as retiradas são "quase" independentes. Com efeito, suponhamos que n seja fixo e D dependa de N de modo que $D/N \to p$ quando $N \to \infty$, onde $0 < p < 1$. Neste caso, $X_N \xrightarrow{D} b(n, p)$, como verificaremos agora. Para $k = 0, 1, \ldots, n$,

$$P(X_N = k) = \frac{D!}{k!(D-k)!} \cdot \frac{(N-D)!}{(n-k)!(N-D-n+k)!} \cdot \frac{n!(N-n)!}{N!} =$$

$$= \binom{n}{k} \frac{D(D-1)\ldots(D-k+1)\cdot(N-D)(N-D-1)\ldots(N-D-n+k+1)}{N(N-1)\ldots(N-n+1)} =$$

$$= \binom{n}{k} \cdot \frac{\frac{D}{N}\left(\frac{D}{N}-\frac{1}{N}\right)\ldots\left(\frac{D}{N}-\frac{k-1}{N}\right)\left(1-\frac{D}{N}\right)\left(1-\frac{D+1}{N}\right)\ldots\left(1-\frac{D+n-k-1}{N}\right)}{1\cdot\left(1-\frac{1}{N}\right)\cdot\ldots\cdot\left(1-\frac{n-1}{N}\right)}$$

Como $\frac{D}{N} \to p$ e n é fixo, temos

$$P(X_N = k) \xrightarrow[N\to\infty]{} \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n,$$

logo $X_N \xrightarrow{D} b(n, p)$.

Exercício. Suponha que $X_n \sim b(n, p_n)$, onde $np_n \to \lambda$, $0 < \lambda < \infty$, Mostre, de duas maneiras, que $X_n \xrightarrow{D}$ Poisson (λ), primeiro usando a Proposição 6.3, depois utilizando funções características.

PROPOSIÇÃO 6.4. *(Teorema de Scheffé). Sejam $X_1, X_2, \ldots$ e X variáveis aleatórias tendo, respectivamente, densidades $f_1, f_2, \ldots$ e f. Se $f_n(x) \to f(x)$ quando $n \to \infty$ para quase todo x relativamente à medida de Lebesgue, então $X_n \xrightarrow{D} X$.*

Nota. A condição $f_n(x) \to f(x)$ para quase todo x, ou $f_n \to f$ em quase toda parte, significa que o conjunto $\{x: f_n(x) \not\to f(x)\}$ tem medida de Lebesgue nula, i.e., comprimento zero.

Prova. Omitida (usa conceitos da Teoria da Medida. Veja, por exemplo, Lehmann [13], p. 351.). □

EXEMPLO 5. Suponha que $X_n \sim N(\mu_n, \sigma_n^2)$, $n = 1, 2, \ldots$. Se $\mu_n \to \mu$ e $\sigma_n^2 \to \sigma^2$ quando $n \to \infty$, onde $\mu \in \mathbb{R}$ e $0 < \sigma^2 < \infty$, então $X_n \xrightarrow{D} N(\mu, \sigma^2)$. É conseqüência imediata do Teorema de Scheffé, pois

$$f_{X_n}(x) = \frac{1}{\sqrt{2\pi}\,\sigma_n} e^{-(x-\mu_n)^2/2\sigma_n^2} \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad \forall x \in \mathbb{R}.$$

(*Exercício.* Verifique este resultado através de funções características.)

EXEMPLO 6. Se $X_n \sim U[0, a_n]$, onde $a_n \to a$, $0 < a < \infty$, então $X_n \xrightarrow{D} U[0, a]$. Pois temos

$$f_{X_n}(x) = \frac{1}{a_n} I_{[0, a_n]}(x) \xrightarrow[n\to\infty]{} \frac{1}{a} I_{[0, a]}(x),$$

para todo $x \in \mathbb{R}$, exceto possivelmente para $x = a$. Logo o Teorema de Scheffé é aplicável.

EXEMPLO 7. A recíproca não vale necessariamente. Daremos um exemplo de uma seqüência $X_1, X_2, \ldots$ tal que $X_n \xrightarrow{D} U[0, 1]$ mas para quase todo $x \in [0, 1]$, $f_{X_n}(x) \not\to 1$. Para tanto, sejam $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}_{[0,1]}$, P = probabilidade uniforme em $[0, 1]$, i.e., $P(A)$ = comprimento de A, para A boreliano em $[0, 1]$. Se $x \in [0, 1]$, seja $0, x_1 x_2 x_3 \ldots$ a sua expansão binária e definamos

$$f_n(x) = \begin{cases} 2x_n, & x \in [0, 1] \\ 0, & x \notin [0, 1]. \end{cases}$$

Como $\int_0^1 f_n(x)\,dx = 1$, f_n é uma densidade (é densidade da distribuição uniforme nos intervalos binários "pares" de ordem n). Para x normal com relação à base 2 e, portanto, para quase todo x (veja o §5.3), a seqüência $f_n(x)$ oscila entre 0 e 2 e não converge.

Mas se X_n possui densidade f_n, então $X_n \xrightarrow{D} U[0, 1]$, como podemos verificar assim: se $r = j/2^m$ é um racional binário em $[0, 1]$, então $P(X_n \le r) = \int_0^r f_n(x)\,dx = r$ se $n > m$. Logo, $F_{X_n}(x) \to x$ se x é racional binário em $[0, 1]$, o que implica $F_{X_n}(x) \to x$ para todo $x \in [0, 1]$ e $X_n \xrightarrow{D} U[0, 1]$.

(ii) *Relação entre os tipos de convergência.* Sejam $X_1, X_2, \ldots$ e X variáveis aleatórias definidas em um espaço de probabilidade $(\Omega, \mathcal{A}, P)$. Temos três tipos de convergência de X_n para X: convergência quase certa, em probabilidade e em distribuição. Já sabemos que convergência quase certa implica convergência em probabilidade e que não vale a recíproca. É razoável perguntarmos agora se há alguma relação entre convergência em distribuição e os outros dois tipos de convergência. A resposta a esta pergunta é que convergência em distribuição é a mais fraca das três convergências:

PROPOSIÇÃO 6.5. *Se $X_n \to X$ em probabilidade, então $X_n \xrightarrow{D} X$.*

Prova. Suponha que $X_n \xrightarrow{P} X$ e seja x um ponto de continuidade de F_X. Queremos provar que $F_{X_n}(x) \to F_X(x)$ quando $n \to \infty$.

Como para $\varepsilon > 0$, $X_n \le x \Rightarrow X \le x + \varepsilon$ ou $X - X_n > \varepsilon$, temos $[X_n \le x] \subset [X \le x + \varepsilon] \cup [|X_n - X| > \varepsilon]$. Logo,

$$F_{X_n}(x) = P(X_n \le x) \le F_X(x + \varepsilon) + P(|X_n - X| > \varepsilon).$$

Por outro lado, $X \le x - \varepsilon \Rightarrow X_n \le x$ ou $X_n - X > \varepsilon$, de modo que

$$F_X(x - \varepsilon) \le F_{X_n}(x) + P(|X_n - X| > \varepsilon).$$

Juntando as duas desigualdades, temos $\forall \varepsilon > 0, \forall n$,

$$F_X(x-\varepsilon) - P(|X_n - X| > \varepsilon) \le F_{X_n}(x) \le F_X(x+\varepsilon) + P(|X_n - X| > \varepsilon).$$

Fazendo inicialmente $n \to \infty$ e depois $\varepsilon \to 0$, temos primeiro (pois $X_n \xrightarrow{P} X$)

$$F_X(x-\varepsilon) \le \liminf_{n\to\infty} F_{X_n}(x) \le \limsup_{n\to\infty} F_{X_n}(x) \le F_X(x+\varepsilon)$$

e, portanto (x é ponto de continuidade de F_X), $F_X(x) = \lim_{n\to\infty} F_{X_n}(x)$. □

Embora a implicação recíproca não valha em geral, é válida quando X é constante:

PROPOSIÇÃO 6.6. *Se $X_n \xrightarrow{D} c$ constante, então $X_n \xrightarrow{P} c$.*

Prova. A função de distribuição de uma variável aleatória constante c é

$$F_c(x) = \begin{cases} 1 & \text{se } x \ge c \\ 0 & \text{se } x < c. \end{cases}$$

Como x é ponto de continuidade de F_c se $x \neq c$, segue-se pela convergência em distribuição que $F_{X_n}(x) \to 0$ quando $n \to \infty$, se $x < c$, e $F_{X_n}(x) \to 1$ quando $n \to \infty$, se $x > c$. Logo, para $\varepsilon > 0$,

$$P(|X_n - c| \le \varepsilon) = P(c - \varepsilon \le X_n \le c + \varepsilon) \ge P(c - \varepsilon < X_n \le c + \varepsilon) =$$
$$= F_{X_n}(c+\varepsilon) - F_{X_n}(c-\varepsilon) \xrightarrow[n\to\infty]{} 1.$$

Em outras palavras, $P(|X_n - c| > \varepsilon) \to 0 \quad \forall \varepsilon > 0$ e $X_n \xrightarrow{P} c$. □

EXEMPLO 8. Em geral, $X_n \xrightarrow{D} X \not\Rightarrow X_n \xrightarrow{P} X$. Sejam $X, X_1, X_2, \ldots$ independentes com distribuição comum $N(0, 1/2)$. Então $X_n \xrightarrow{D} X$, pois todas as distribuições são iguais. Mas $X_n - X \sim N(0, 1)$ e $P(|X_n - X| \ge \varepsilon) = 2 - 2\Phi(\varepsilon)$, onde Φ é a função de distribuição da normal-padrão. Logo, para $\varepsilon > 0$, $P(|X_n - X| \ge \varepsilon) \not\to 0$ e $X_n \not\xrightarrow{P} X$.

Resumindo, temos que os tipos de convergência possuem a seguinte relação:

conv. quase certa $\Rightarrow$ conv. em probabilidade $\Rightarrow$ conv. em distribuição.

(iii) *O Teorema de Slutsky.* Nesta seção consideramos alguns casos em que convergência de variáveis aleatórias acarreta convergência de funções, somas ou produtos das variáveis. Iniciaremos nosso estudo

com a proposição: todo tipo de convergência é preservado por funções contínuas.

PROPOSIÇÃO 6.7. *Sejam $X_1, X_2, \ldots$ e X variáveis aleatórias e $g: \mathbb{R} \to \mathbb{R}$ uma função contínua. Então*

(a) $X_n \to X$ quase certamente $\Rightarrow g(X_n) \to g(X)$ quase certamente;
(b) $X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X)$; e
(c) $X_n \xrightarrow{D} X \Rightarrow g(X_n) \xrightarrow{D} g(X)$.

Prova. Provaremos em ordem crescente de dificuldade:

(a) Se $X_n \to X$ quase certamente, então existe $A_0 \in \mathcal{A}$ tal que $P(A_0) = 1$ e para todo $\omega \in A_0$, $X_n(\omega) \to X(\omega)$. Como g é contínua, $g(X_n(\omega)) \to g(X(\omega)) \ \forall \omega \in A_0$, logo $g(X_n) \to g(X)$ quase certamente.

(c) Suponha que $X_n \xrightarrow{D} X$. Para que $g(X_n) \xrightarrow{D} g(X)$, basta a convergência das respectivas funções características. Por definição, $\varphi_{g(X_n)}(t) = Ee^{itg(X_n)} = E\cos(tg(X_n)) + iE\,\mathrm{sen}(tg(X_n))$.

Como as funções $\cos(tg(x))$ e $\mathrm{sen}(tg(x))$ são contínuas e limitadas na reta, para t fixo, decorre do Teorema de Helly-Bray que

$$\varphi_{g(X_n)}(t) \xrightarrow[n\to\infty]{} E\cos(tg(X)) + iE\,\mathrm{sen}(tg(X)) = \varphi_{g(X)}(t), \quad t \in \mathbb{R}.$$

(b) Suponha que $X_n \xrightarrow{P} X$, e seja $\varepsilon > 0$. Para todo $m > 0$, g é uniformemente contínua em $[-m, m]$; escolhamos m suficientemente grande tal que $P(-m/2 \le X \le m/2) > 1 - \varepsilon$ (isto é possível porque $P(|X| \le m/2) \to 1$ quando $m \to \infty$). Pela continuidade uniforme, existe δ tal que $0 < \delta \le m/2$ e *se* $|x| \le m$, $|y| \le m$ e $|x - y| < \delta$ *então* $|g(x) - g(y)| < \varepsilon$.

Como $P(|X_n - X| < \delta) \to 1$, temos $P(|X| \le m/2, |X_n - X| < \delta) \xrightarrow[n\to\infty]{} P(|X| \le m/2) > 1 - \varepsilon$ (observe que se $P(A_n) \to 1$, então $P(A \cap A_n) \to P(A)$). Mas

$$[|X| \le m/2, |X_n - X| < \delta] \subset [|X| \le m, |X_n| \le m, |X_n - X| < \delta] \subset$$
$$\subset [|g(X_n) - g(X)| < \varepsilon],$$

logo $P(|g(X_n) - g(X)| < \varepsilon) > 1 - \varepsilon$ para n suficientemente grande.

Por isso, se $0 < \xi < \varepsilon$, então para n suficientemente grande,

$$P(|g(X_n) - g(X)| < \varepsilon) \ge P(|g(X_n) - g(X)| < \xi) > 1 - \xi.$$

Em outras palavras, $P(|g(X_n) - g(X)| < \varepsilon) \to 1$ quando $n \to \infty$, i.e., $g(X_n) \xrightarrow{P} g(X)$. □

Como conseqüências desta proposição temos, por exemplo, $X_n \xrightarrow{D} N(0,1) \Rightarrow X_n^2 \xrightarrow{D} \chi_1^2$; $X_n \xrightarrow{D} N(0,1) \Rightarrow cX_n \xrightarrow{D} N(0, c^2)$; $X_n \xrightarrow{P} c > 0 \Rightarrow$ $\Rightarrow \log X_n \xrightarrow{P} \log c$.

Observação. Quando g não é definida e contínua em toda a reta, por exemplo, $g(x) = \log x$ ou $g(x) = 1/x$, pode-se provar que a proposição ainda vale se para algum conjunto aberto $A \subset \mathbb{R}$, g é contínua em A e $P(X \in A) = 1$. Se $g(X_n)$ não é finita com probabilidade 1, como é o caso se $g(x) = 1/x$ e $P(X_n = 0) > 0$, $g(X_n)$ pode ser arbitrariamente definida quando não é finita, de maneira a convertê-la em variável aleatória. Analogamente, a condição de que $P(Y_n \neq 0) = 1$ no item (d) do teorema seguinte não é necessário, se X_n/Y_n é arbitrariamente definida quando $Y_n = 0$, digamos $X_n/Y_n = 0$.

TEOREMA 6.4. *(Teorema de Slutsky). Sejam $X, X_1, X_2, \ldots$ e $Y_1, Y_2, \ldots$ variáveis aleatórias tais que $X_n \xrightarrow{D} X$ e $Y_n \xrightarrow{P} c$, onde c é uma constante. Então*

(a) $X_n + Y_n \xrightarrow{D} X + c$;
(b) $X_n - Y_n \xrightarrow{D} X - c$;
(c) $Y_n X_n \xrightarrow{D} cX$; *e*
(d) se $c \neq 0$ *e* $P(Y_n \neq 0) = 1$, $\dfrac{X_n}{Y_n} \xrightarrow{D} \dfrac{X}{c}$.

Prova. A prova utilizará funções características. Para uma prova mais "elementar" – partindo da definição de convergência em distribuição – veja Bickel e Doksum [3], p. 461.

(a) Pelo Teorema de Paul Lévy, basta provarmos que $\varphi_{X_n + Y_n}(t) \to \varphi_{X+c}(t)$ quando $n \to \infty$, para todo $t \in \mathbb{R}$. Temos

$$\varphi_{X_n+Y_n}(t) = Ee^{it(X_n+Y_n)} = Ee^{it(X_n+c)} + Ee^{itX_n}(e^{itY_n} - e^{itc}).$$

Como $Ee^{itX_n} = \varphi_{X_n}(t) \to \varphi_X(t)$, a primeira parcela do termo à direita tende a $e^{itc}\varphi_X(t) = \varphi_{X+c}(t)$. Logo, para terminarmos a prova basta que provemos a convergência para zero do segundo termo. Como

$$|Ee^{itX_n}(e^{itY_n} - e^{itc})| \leq E|e^{itX_n}(e^{itY_n} - e^{itc})| = E|e^{itY_n} - e^{itc}|,$$

porque $|e^{itX_n}| = 1$, basta provarmos que $E|e^{itY_n} - e^{itc}| \to 0$ quando $n \to \infty$. Mas isto é conseqüência do Teorema da Convergência Dominada, que vale também no caso de convergência em probabilidade (não provaremos esta extensão do teorema): como $Y_n \xrightarrow{P} c$ implica que $Z_n \overset{def}{=} |e^{itY_n} - e^{itc}| \xrightarrow{P} 0$, pela proposição anterior com $g(y) = |e^{ity} - e^{itc}|$, e como a seqüência é dominada pela constante 2, pois $|Z_n| \leq |e^{itY_n}| + |e^{itc}| = 2$, temos $EZ_n \to 0$. Portanto, o item (a) está provado.

(Podemos provar que $EZ_n \to 0$ diretamente, sem apelar para o Teorema da Convergência Dominada:

$$0 \leq EZ_n = EZ_n I_{[Z_n \leq \varepsilon]} + EZ_n I_{[Z_n > \varepsilon]} \leq E(\varepsilon) + 2E\, I_{[Z_n > \varepsilon]} = \\ = \varepsilon + 2P(Z_n > \varepsilon) < 2\varepsilon$$

para n suficientemente grande, pois $Z_n \xrightarrow{P} 0$.)

(b) Conseqüência de (a), pois $-Y_n \xrightarrow{P} -c$.

(c) Suponhamos primeiro que $c = 0$. Queremos mostrar que $Y_n X_n \xrightarrow{D} 0$. Para tanto, basta provarmos que $Y_n X_n \xrightarrow{P} 0$, pois convergência em distribuição decorre da convergência em probabilidade.

Agora sejam $\varepsilon, \delta > 0$ 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 $X_n \xrightarrow{D} X$, temos $P(x < X_n \leq y) = F_{X_n}(y) - F_{X_n}(x) > 1 - \delta$ para n suficientemente grande. Definamos $M = \max(y, -x)$; então a convergência em probabilidade de Y_n para zero implica que $P(|Y_n| < \varepsilon/M) > 1 - \delta$ para n suficientemente grande. Logo, para n suficientemente grande (observação: $P(A \cap B) \geq \geq 1 - P(A^c) - P(B^c)$), temos

$$P\left(x < X_n \leq y,\ |Y_n| < \frac{\varepsilon}{M}\right) > 1 - 2\delta.$$

Como $x < X_n \leq y$ e $|Y_n| < \varepsilon/M$ implicam $|X_n Y_n| < \varepsilon$, temos $P(|X_n Y_n| < \varepsilon) > 1 - 2\delta$ para n suficientemente grande. Portanto, para todo $\varepsilon > 0$, $P(|X_n Y_n| < \varepsilon) \to 1$, i.e., $X_n Y_n \xrightarrow{P} 0$, como queríamos demonstrar.

Agora consideremos o caso de c geral. Como $Y_n X_n = cX_n + (Y_n - c)X_n$ e $Y_n - c \xrightarrow{P} 0$, segue-se da conclusão no caso $c = 0$ que $(Y_n - c)X_n \xrightarrow{P} 0$. Além disso, temos $\varphi_{cX_n}(t) = \varphi_{X_n}(ct) \to \varphi_X(ct) = \varphi_{cX}(t)$, $\forall\, t \in \mathbb{R}$, i.e. $cX_n \xrightarrow{D} cX$ (este resultado decorre também da proposição anterior). Como $Y_n X_n$ é soma de dois termos, o primeiro dos quais convergente para cX em distribuição, e o segundo para zero em probabilidade, o resultado é conseqüência do item (a).

(d) Pela proposição anterior, $1/Y_n \xrightarrow{P} 1/c$. Agora basta aplicar o item (c). □

PROPOSIÇÃO 6.8. *Sejam $Y_1, Y_2, \ldots$ variáveis aleatórias tais que $\sqrt{n}(Y_n - \mu) \xrightarrow{D} N(0, \sigma^2)$. Se $g(y)$ é uma função derivável no ponto μ, então $\sqrt{n}(g(Y_n) - g(\mu)) \xrightarrow{D} N(0, \sigma^2(g'(\mu))^2)$.*

Observação. O resultado vale quando $g'(\mu) = 0$, se a distribuição $N(0, 0)$ é interpretada como massa pontual em 0, ou seja, $0 \sim N(0,0)$.

Prova. A idéia da prova é esta: sob as condições dadas, $Y_n \to \mu$ em probabilidade quando $n \to \infty$. Como g é derivável em μ e, na vizinhança de μ, $g(y) - g(\mu) \approx g'(\mu)(y - \mu)$, temos $\sqrt{n}(g(Y_n) - g(\mu)) \approx \sqrt{n}\, g'(\mu)(Y_n - \mu)$. Esta última variável converge em distribuição para o produto de $g'(\mu)$ com uma variável $N(0, \sigma^2)$, de modo que o limite é $N(0, \sigma^2(g'(\mu))^2)$.

Verifiquemos primeiro que $Y_n \xrightarrow{P} \mu$. Para tanto, sejam $\varepsilon > 0$ e m um inteiro positivo. Como

$$P(|Y_n - \mu| < \varepsilon) = P(|\sqrt{n}(Y_n - \mu)| < \sqrt{n}\,\varepsilon) \geq P(-m < \sqrt{n}(Y_n - \mu) \leq m)$$

para n suficientemente grande, temos

$$\liminf_{n \to \infty} P(|Y_n - \mu| < \varepsilon) \geq \Phi(m) - \Phi(-m),$$

onde Φ é a função de distribuição da $N(0, 1)$. Como isto vale para todo m e o termo à direita converge para 1 quando $m \to \infty$, segue-se que $P(|Y_n - \mu| < \varepsilon) \to 1$ quando $n \to \infty$, para todo $\varepsilon > 0$, i.e., $Y_n \xrightarrow{P} \mu$.

Já que $g(y) - g(\mu) = g'(\mu)(y - \mu) + e(y)(y - \mu)$, onde $\lim_{y \to \mu} e(y) = 0$, temos

$$\sqrt{n}(g(Y_n) - g(\mu)) = (g'(\mu) + e(Y_n))\sqrt{n}(Y_n - \mu).$$

Agora basta provar que $g'(\mu) + e(Y_n) \xrightarrow{P} g'(\mu)$, ou equivalentemente, $e(Y_n) \xrightarrow{P} 0$, para poder aplicar Slutsky e terminar a prova.

Para provar que $e(Y_n) \xrightarrow{P} 0$, seja $\varepsilon > 0$ e considere $P(|e(Y_n)| < \varepsilon)$. Como $\lim_{y \to \mu} e(y) = 0$, existe $\delta > 0$ tal que $|e(y)| < \varepsilon$ se $|y - \mu| < \delta$. Logo, se $|Y_n - \mu| < \delta$, então $|e(Y_n)| < \varepsilon$, de modo que $P(|Y_n - \mu| < \delta) \leq P(|e(Y_n)| < \varepsilon)$. Como $P(|Y_n - \mu| < \delta) \to 1$, pois $Y_n \xrightarrow{P} \mu$, a seqüência $P(|e(Y_n)| < \varepsilon)$ necessariamente converge para 1 quando $n \to \infty$, como queríamos demonstrar. □

EXEMPLO 9. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas, com $EX_n = \mu$ e $\operatorname{Var} X_n = \sigma^2$, onde $0 < \sigma^2 < \infty$. Então pelo Teorema Central do Limite,

$$\frac{S_n - n\mu}{\sqrt{n}\,\sigma} \xrightarrow{D} N(0, 1).$$

Definindo $\overline{X}_n = \dfrac{S_n}{n} = \dfrac{X_1 + \ldots + X_n}{n}$ ($\overline{X}_n$ é a *média amostral* da amostra de tamanho n), podemos reformular o Teorema Central do Limite,

dividindo o numerador e o denominador por n. A conclusão fica

$$\sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right) \xrightarrow{D} N(0, 1),$$

ou ainda (pela Proposição 6.7 ou por Slutsky)

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{D} N(0, \sigma^2).$$

Pela Proposição 6.8, temos, por exemplo:

(a) Seja $g(x) = x^2$. Então $g'(\mu) = 2\mu$ e

$$\sqrt{n}(\overline{X}_n^2 - \mu^2) \xrightarrow{D} N(0, 4\sigma^2\mu^2).$$

(b) Seja $g(x) = 1/x$. Se $\mu = 0$, g não é derivável em μ e não podemos aplicar a proposição. Mas se $\mu \neq 0$, então $g'(\mu) = -1/\mu^2$ e

$$\sqrt{n}\left(\frac{1}{\overline{X}_n} - \frac{1}{\mu}\right) \xrightarrow{D} N\left(0, \frac{\sigma^2}{\mu^4}\right).$$

(Para que o termo à esquerda seja uma variável aleatória, podemos definir $1/\overline{X}_n$ arbitrariamente quando $\overline{X}_n = 0$.)

Notemos que, no início da prova da Proposição 6.8, demonstramos que se $\sqrt{n}(Y_n - \mu) \xrightarrow{D} N(0, \sigma^2)$, então $Y_n \xrightarrow{P} \mu$. Portanto, *o Teorema Central do Limite, no caso de variáveis independentes e identicamente distribuídas, implica a Lei Fraca dos Grandes Números*, i.e.,

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{D} N(0, 1) \Rightarrow \overline{X}_n \xrightarrow{P} \mu.$$

(iv) *Uma caracterização das funções características.* Suponha que temos uma função φ, definida na reta real e tomando valores complexos, ou seja, $\varphi: \mathbb{R} \to \mathbb{C}$. Consideremos o problema de determinar se φ é uma função característica. Partindo da suposição de que φ seja contínua e satisfaça $\varphi(0) = 1$, há uma solução teórica geral para este problema, devida a Bochner e Khintchin: φ é característica se, e só se, é definida positiva.

DEFINIÇÃO 6.5. Uma função $\varphi: \mathbb{R} \to \mathbb{C}$ é chamada *definida positiva* se para todo $n = 1, 2, \ldots$ e toda n-upla $(c_1, \ldots, c_n)$ de números complexos temos

$$\sum_{k=1}^{n} \sum_{j=1}^{n} \varphi(t_j - t_k)\, c_j \bar{c}_k \geq 0, \ \forall (t_1, \ldots, t_n) \in \mathbb{R}^n.$$

Para ver que uma função característica é necessariamente definida positiva, é preciso recordar primeiro que se c é complexo, então $c\bar{c} = |c|^2$.

Então, se φ é a função característica da variável aleatória X, temos para $t_1, \ldots, t_n$ reais e $c_1, \ldots, c_n$ complexos,

$$\sum_{j,k} \varphi(t_j - t_k)\, c_j \bar{c}_k = \sum_{j,k} c_j \bar{c}_k\, E e^{i(t_j - t_k)X} =$$

$$= \sum_{j,k} E(c_j e^{it_j X}\, \bar{c}_k\, e^{-it_k X}) = E(\sum_{j,k} c_j e^{it_j X}\, \overline{c_k e^{it_k X}}) =$$

$$= E((\sum_j c_j e^{it_j X})(\overline{\sum_k c_k e^{it_k X}})) = E(|\sum_j c_j e^{it_j X}|^2) \geq 0.$$

TEOREMA 6.5. (*Bochner-Khintchin*). *Seja* $\varphi: \mathbb{R} \to \mathbb{C}$ *uma função contínua tal que* $\varphi(0) = 1$. φ *é uma função característica (i.e., função característica de alguma variável aleatória) se, e somente se,* φ *é definida positiva.*

Prova. A necessidade da condição foi demonstrada acima. Omitimos a prova da suficiência. Veja Feller [9], §XIX.2 ou Gnedenko [11], §39. □

Na prática, costuma ser difícil verificar se uma dada função é definida positiva. No entanto, o critério é de utilidade teórica, como no seguinte exemplo.

EXEMPLO 10. Sejam $\varphi_1, \varphi_2, \ldots$ funções características. Então toda média ponderada das φ_n é também característica, onde por média ponderada entenda-se função φ da forma

$$\varphi(t) = \sum_m p_m \varphi_m(t),$$

com $p_m \geq 0$ e $\sum_m p_m = 1$. Isto vale porque $\varphi(0) = \sum_m p_m \varphi_m(0) = \sum_m p_m = 1$ e φ é contínua ($\sum_{m=1}^{k} p_m \varphi_m$ é contínua para todo k e tende uniformemente a φ quando $k \to \infty$). Resta verificar se φ é definida positiva:

$$\sum_{j,k} \varphi(t_j - t_k) c_j \bar{c}_k = \sum_{j,k} (\sum_m p_m \varphi_m(t_j - t_k))\, c_j \bar{c}_k =$$

$$= \text{(a série é absolutamente convergente)} =$$

$$= \sum_m p_m (\sum_{j,k} \varphi_m(t_j - t_k)\, c_j \bar{c}_k) \geq 0,$$

pois cada φ_m é definida positiva.

EXERCÍCIOS DO CAPÍTULO 6

§6.1

1. (a) Se $X \sim b(n, p)$, qual a função característica de X?
   (b) Mostre, usando funções características, que se $X \sim b(m, p)$, $Y \sim b(n, p)$, e X e Y são independentes, então $X + Y \sim b(m+n, p)$.

2. Mostre que se $X_1, \ldots, X_n$ são independentes com, cada uma, distribuição simétrica em torno de 0, então $\sum_{j=1}^{n} a_j X_j$ possui distribuição simétrica em torno de 0, para toda escolha das constantes $a_j \in \mathbb{R}$.

3. Seja φ uma função característica. Mostre que $\psi(t) = e^{\lambda(\varphi(t)-1)}$, onde $\lambda > 0$, também é função característica. (*Sugestão.* Sejam $N, X_1, X_2, \ldots$ independentes tais que $N \sim$ Poisson (λ) e as X_n são identicamente distribuídas com $\varphi_{X_n} = \varphi$. Defina $Y = S_N$, onde $S_n = X_1 + \ldots + X_n$; N é um "tempo de parada" para a seqüência de somas parciais. Então $\varphi_Y = \psi$. A distribuição de Y é chamada *distribuição composta de Poisson.* A distribuição comum de Poisson corresponde ao caso $X_n = 1$, i.e., $P(X_n = 1) = 1$.)

4. Calcule EX^3 e EX^4, onde $X \sim N(\mu, \sigma^2)$. (*Sugestão.* Calcule primeiro para a $N(0, 1)$ e use linearidade.)

5. (a) Mostre que se X tem distribuição Cauchy-padrão, então $\varphi_{2X}(t) = \varphi_X^2(t)$. (Pode usar, sem provar: $\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\cos(xt)}{1+x^2}\, dx = e^{-|t|}$)
   Utilize esse resultado para provar que

   $$\varphi_{X+Y}(t) = \varphi_X(t)\, \varphi_Y(t),\ \forall\, t \in \mathbb{R} \not\Rightarrow X \text{ e } Y \text{ independentes,}$$

   e portanto,

   $$F_{X+Y}(z) = F_X * F_Y(z),\ \forall\, z \in \mathbb{R} \not\Rightarrow X \text{ e } Y \text{ independentes.}$$

   ($F_X * F_Y$ é a convolução de F_X com F_Y.)
   (b) Sejam $X_1, \ldots, X_n$ independentes e identicamente distribuídas, com distribuição comum Cauchy-padrão. Demonstre que a média amostral $\frac{S_n}{n} = \frac{X_1 + \ldots + X_n}{n}$ também é Cauchy-padrão.

6. Sejam X e Y variáveis aleatórias com a mesma distribuição. Demonstre:
   (a) Se X e Y são independentes, então $X - Y$ tem distribuição simétrica em torno de zero.
   (b) Se X e Y tomam só dois valores, então $X - Y$ tem distribuição simétrica em torno de zero.

7. (a) Suponha que $X \sim \exp(\lambda)$ e mostre que a função característica de X é $\varphi(x) = \dfrac{\lambda}{\lambda - it} = \dfrac{\lambda^2 + it\lambda}{\lambda^2 + t^2}$.
   (b) Seja Y exponencial dupla com densidade
   $$f_Y(y) = \frac{\lambda}{2}\, e^{-\lambda|y|},\ y \in \mathbb{R}.$$
   Calcule a função característica de Y. (*Sugestão.* Use simétria e o item (a)).
   (c) Demonstre: se Z e W são independentes e identicamente distribuídas, com $Z \sim \exp(\lambda)$, então $Z - W$ é exponencial dupla.

8. Use a função característica do exercício anterior para mostrar que se $X \sim \Gamma(n, \beta)$, *então* $\varphi_X(t) = \left(\dfrac{\beta}{\beta - it}\right)^n$.

9. Demonstre:
   (a) Se φ é uma função característica e existe $\lambda \neq 0$ tal que $\varphi(\lambda) = 1$, então a distribuição correspondente a φ está concentrada nos pontos $\pm k\left(\dfrac{2\pi}{\lambda}\right)$, $k = 0, 1, \ldots$.
   (b) Se φ é uma função característica e existe $\delta > 0$ tal que $\varphi(t) = 1$ para todo t com $|t| < \delta$, então $\varphi(t) = 1\ \forall t$. (Qual a distribuição correspondente a φ?)

10. A *função geradora de momentos* de uma variável aleatória X é definida por
    $$\psi_X(t) = Ee^{tX},\ t \in \mathbb{R}.$$
    (É permitido a ψ_X assumir o valor $+\infty$.) Demonstre que se $Ee^{\delta|X|} < \infty$ para algum $\delta > 0$, então:
    (a) $\psi(t)$ é finito para $t \in [-\delta, \delta]$;
    (b) todos os momentos de X são finitos; e
    (c) ψ possui derivadas contínuas de toda ordem em $(-\delta, \delta)$, e $\psi^{(k)}(0) = EX^k$ para $k = 1, 2, \ldots$. (*Sugestão.* Use o método de prova da propriedade FC9.)

11. Obtenha a função geradora de momentos (definida no exercício anterior) das variáveis aleatórias seguintes:
    (a) $X \sim$ Poisson (λ), onde $\lambda > 0$.
    (b) $X \sim$ Cauchy-padrão.
    (c) $X \sim \exp(\lambda)$, onde $\lambda > 0$. Utilize o resultado para calcular os momentos EX^k, $k = 1, 2, \ldots$. Confira com os momentos obtidos no exemplo 5 do Capítulo 3 (§3.4).

§6.2

12. Verifique se $c_1, c_2, \ldots$ e c são números complexos tais que $c_n \to c$, então $\left(1 + \dfrac{c_n}{n}\right)^n \to e^c$. (*Sugestão.* Considere o logaritmo principal de $1 + \dfrac{c_n}{n}$.)

13. (a) Suponha que $X_n \xrightarrow{D} N(0, 1)$, $Y_n \xrightarrow{D} N(0, 1)$ e, para todo n, X_n seja independente de Y_n. Mostre que $X_n + Y_n \xrightarrow{D} N(0, 2)$.
    (b) Generalize (a), provando que se $X_n \xrightarrow{D} F$ e $Y_n \xrightarrow{D} G$, onde X_n e Y_n são independentes $\forall n$, então $X_n + Y_n \xrightarrow{D} F * G$. Aqui, F e G são funções de distribuição e $F * G$ é sua convolução.

14. Qual a distribuição de X se X tem função característica $\varphi_X(t) = \cos^2(t)$? (Veja o exemplo 3.)

15. Mostre que é possível para uma seqüência de funções de distribuição convergir em todo ponto sem o limite ser uma função de distribuição. (*Sugestão.* Considere as variáveis aleatórias constantes $X_n = n$.)

16. Prove: se $F_n \to F$ fracamente e F é contínua, então $F_n(x)$ converge para $F(x)$ uniformemente na reta.

17. Utilize funções características para provar: se $X_n \xrightarrow{D} N(0, 1)$ e $(a_n)_{n \geq 1}$ é uma seqüência de números reais tal que $a_n \to a$ finito, então $X_n + a_n \xrightarrow{D} N(a, 1)$.

18. Sejam $X_1, X_2, \ldots$ variáveis aleatórias, cada uma tendo distribuição simétrica em torno de zero. Demonstre que se $X_n \xrightarrow{D} X$, então X também tem distribuição simétrica em torno de zero.

19. Sejam $X_1, X_2, \ldots$ independentes e identicamente distribuídas, com $X_n \sim U[0, 1]$, e sejam $Y_n = \min(X_1, \ldots, X_n)$, $Z_n = \max(X_1, \ldots, X_n)$,

$U_n = nY_n$, $V_n = n(1 - Z_n)$. Mostre que, quando $n \to \infty$:
(a) $Y_n \xrightarrow{P} 0$ e $Z_n \xrightarrow{P} 1$.
(b) $U_n \xrightarrow{D} W$ e $V_n \xrightarrow{D} W$, onde W tem distribuição exponencial de parâmetro 1.

20. Seja $(X_n)_{n \geq 1}$ uma seqüência de variáveis aleatórias independentes e identicamente distribuídas, tais que $P(X_n = 1) = \frac{1}{2} = P(X_n = -1)$, e seja

$$Y_n = \sum_{k=1}^{n} \frac{1}{2^k} X_k.$$

Mostre que $Y_n \xrightarrow{D} U[-1, 1]$. (*Sugestão.* Use a igualdade $\cos\theta = \operatorname{sen}(2\theta)/(2 \operatorname{sen}\theta)$.)

21. Sejam $X_1, X_2, \ldots$ variáveis aleatórias cujas funções características $\varphi_1, \varphi_2, \ldots$ convergem pontualmente, Mostre que se existe $\delta > 0$ tal que $\varphi_n(t) \to 1$ para todo t com $|t| < \delta$, então $X_n \xrightarrow{P} 0$. (*Sugestão.* Use o resultado do exercício 9.)

§6.3

22. Dizemos que $\underset{\sim}{X} = (X_1, \ldots, X_k)$ tem *distribuição simétrica em torno de zero* se $\underset{\sim}{X}$ e $-\underset{\sim}{X}$ possuem a mesma distribuição. Demonstre que $\underset{\sim}{X}$ tem distribuição simétrica em torno de zero se, e somente se, $\varphi_{\underset{\sim}{X}}(t)$ é real para todo $t \in \mathbb{R}^k$.

23. Sejam X, Y, U e V variáveis aleatórias definidas em um espaço de probabilidade $(\Omega, \mathcal{A}, P)$. Suponha que os vetores (X, Y) e (U, V) sejam independentes, que X seja independente de Y e que U seja independente de V. Mostre que X, Y, U e V são coletivamente independentes. Generalize esse resultado para o caso de n vetores aleatórios independentes, cada qual composto de componentes independentes.

24. Sejam X, Y, U e V variáveis aleatórias, e sejam $Z = X + iY$ e $W = U + iV$. Demonstre:
(a) Se os vetores (X, Y) e (U, V) são independentes, então X e U são independentes.
(b) Se (X, Y) e (U, V) são independentes e X, Y, U e V são integráveis, então $E(ZW) = EZ \cdot EW$.

25. Prove a Proposição 6.2 para m e n quaisquer. (Pode usar o seguinte fato, sem prová-lo: se $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ são vetores aleatórios independentes, então $g_1(\underset{\sim}{X})$ e $g_2(\underset{\sim}{Y})$ são independentes, onde g_1 e g_2 são funções reais mensuráveis.)

26. (Propriedade FC9 para vetores). Seja $\underset{\sim}{X} = (X_1, \ldots, X_k)$ um vetor aleatório. Sob condições de regularidade, o momento misto $E(X_1^{m_1} X_2^{m_2} \ldots X_k^{m_k})$, onde os m_j são inteiros não-negativos, poderá ser obtido derivando-se a função característica de $\underset{\sim}{X}$:

$$\frac{\partial^{m_1 + \cdots + m_k}}{\partial t_1^{m_1} \ldots \partial t_k^{m_k}} \varphi_{\underset{\sim}{X}}(t_1, \ldots, t_k) \Big|_{\underset{\sim}{t} = \underset{\sim}{0}} = i^{m_1 + \cdots + m_k} E(X_1^{m_1} \ldots X_k^{m_k}).$$

Demonstre este resultado no caso em que $k = 2$ e $m_1 = m_2 = 1$, sob a condição de que EX_1, EX_2 e EX_1X_2 sejam finitas, i.e., demonstre que

$$\frac{\partial^2}{\partial t_1 \, \partial t_2} \varphi_{(X_1, X_2)}(t_1, t_2) \Big|_{t_1 = t_2 = 0} = -EX_1X_2.$$

27. Suponha que $\underset{\sim}{X} = (X_1, \ldots, X_k)$ tenha distribuição multinomial com parâmetros $p_1, \ldots, p_k$ e n (esta distribuição é definida no exercício 7 do Capítulo 4).
(a) Mostre que a função característica de $\underset{\sim}{X}$ é

$$\varphi_{\underset{\sim}{X}}(\underset{\sim}{t}) = (p_1 e^{it_1} + \ldots + p_k e^{it_k})^n, \quad \underset{\sim}{t} \in \mathbb{R}^k.$$

(b) Utilize o exercício 26 para calcular a covariância entre X_j e X_ℓ, $j \neq \ell$.

§6.4.

28. (Generalização da Proposição 6.3.) (a) Demonstre que se $X_1, X_2, \ldots$ e X tomam somente os valores $x_1, x_2, \ldots$, então convergência pontual das funções de probabilidade implica convergência em distribuição, i.e., demonstre que se $\sum_k p(x_k) = 1 = \sum_k p_n(x_k)$, $\forall n$ e $p_n(x_k) \xrightarrow[n \to \infty]{} p(x_k)$, $\forall k$, então $X_n \xrightarrow{D} X$.

(*Sugestão.* Para $\varepsilon > 0$, escolha N tal que $\sum_{k=1}^{N} p(x_k) > 1 - \varepsilon$ e conclua que para n suficientemente grande, e todo $x \in \mathbb{R}$.

$$-2\varepsilon + F_X(x) < -\varepsilon + \sum_{k \leq N \,:\, x_k \leq x} p(x_k) < F_{X_n}(x) <$$
$$< 2\varepsilon + \sum_{k \leq N \,:\, x_k \leq x} p(x_k) \leq 2\varepsilon + F_X(x).)$$

(b) Demonstre que vale a recíproca de (a) se os pontos x_k são isolados, i.e., se para todo k existe um intervalo aberto (a_k, b_k) que contém x_k e não contém outro x_j.

29. Cem mil passas são misturadas em uma massa que posteriormente será dividida em partes iguais para fazer dois mil bolos tipo inglês. Mais tarde, um bolo será escolhido ao acaso e será contado o número de passas contidas nele.
(a) Explique porque você pode modelar esse experimento utilizando a distribuição de Poisson. Qual o parâmetro?
(b) O modelo em (a) é uma aproximação. Qual seria a distribuição *exata* do resultado do experimento?
(c) Qual a probabilidade de não encontrar passa alguma no bolo, segundo cada um dos dois modelos? As duas probabilidades são realmente aproximadamente iguais?

30. Sejam $X_1, X_2, \ldots$ independentes e identicamente distribuídas tais que $X_n \sim U[0, \theta]$, onde $\theta > 0$. Demonstre que $Y_n = \sqrt{n}\{\log(2\bar{X}_n) - \log\theta\}$ converge em distribuição para a $N\left(0, \frac{1}{3}\right)$, onde $\bar{X}_n = \frac{X_1 + \ldots + X_n}{n}$.

31. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas tais que $EX_1 = 0$. Ache o limite, quando $n \to \infty$, da função característica de $Y_n = \cos(\bar{X}_n)$ onde $\bar{X}_n = \frac{X_1 + \ldots + X_n}{n}$.

32. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas, com $EX_1 = 0$ e $EX_1^2 = 2$. Ache o limite em distribuição das seguintes seqüências:

(a) $Y_1, Y_2, \ldots$, onde $Y_n = \frac{\sqrt{n}(X_1 + \ldots + X_n)}{X_1^2 + \ldots + X_n^2}$.

(b) $Z_1, Z_2 \ldots$, onde $Z_n = \frac{X_1 + \ldots + X_n}{\sqrt{X_1^2 + \ldots + X_n^2}}$.

33. Sejam $X, X_1, X_2, \ldots$ e $Y_1, Y_2, \ldots$ variáveis aleatórias tais que $P(X_n = 0) = 0 = P(X = 0)$, $X_n \xrightarrow{D} X$ e $Y_n \xrightarrow{P} c$, onde c é constante. Mostre que $\frac{Y_n}{X_n} \xrightarrow{D} \frac{c}{X}$.

34. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas tais que $EX_1 = 0$ e $\text{Var}\, X_1 = \sigma^2$, onde $0 < \sigma^2 < \infty$. Sejam $Y_1, Y_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas tais que $EY_1 = \mu$, onde $\mu \in \mathbb{R}$. Prove que $\bar{Y}_n + \sqrt{n}\,\bar{X}_n \xrightarrow{D} N(\mu, \sigma^2)$, onde $\bar{X}_n = \frac{X_1 + \ldots + X_n}{n}$ e $\bar{Y}_n = \frac{Y_1 + \ldots + Y_n}{n}$.

35. Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas, com média μ_X e variância $\sigma_X^2 < \infty$; $Y_1, \ldots, Y_n$ variáveis aleatórias independentes e identicamente distribuídas, com média μ_Y e variância $\sigma_Y^2 < \infty$. Suponha que as X_j e Y_k sejam independentes e que $\mu_X \neq 0$. Ache o limite em distribuição de $Z_n = \sqrt{n}\left(\frac{\bar{Y}_n}{\bar{X}_n} - \frac{\mu_Y}{\mu_X}\right)$, onde

$$\bar{X}_n = \frac{X_1 + \ldots + X_n}{n} \text{ e } \bar{Y}_n = \frac{Y_1 + \ldots + Y_n}{n}.$$

(*Sugestão.* $Z_n = \sqrt{n}\left(\frac{\mu_X \bar{Y}_n - \mu_Y \bar{X}_n}{\mu_X \bar{X}_n}\right)$. Use o exercício 13.)

36. (Este resultado é útil na Estatística Não-Paramétrica.) Sejam $T_1, T_2, \ldots$ e $S_1, S_2, \ldots$ seqüências de variáveis aleatórias. Mostre que se (i) $\frac{T_n - ET_n}{\sqrt{\text{Var}\, T_n}} \xrightarrow{D} N(0, 1)$ e (ii) $\frac{E(T_n - S_n)^2}{\text{Var}\, T_n} \to 0$, segue-se que $\frac{S_n - ES_n}{\sqrt{\text{Var}\, S_n}} \xrightarrow{D} N(0, 1)$. (*Sugestão.* Prove primeiro de (ii) que $\frac{ET_n - ES_n}{\sqrt{\text{Var}\, T_n}} \to 0$ e $\frac{(T_n - S_n)}{\sqrt{\text{Var}\, T_n}} \xrightarrow{P} 0$. Depois, use a desigualdade do triângulo várias vezes para provar que $\frac{\text{Var}\, S_n}{\text{Var}\, T_n} \to 1$.)

37. (Continuação do exemplo 10). (a) Dê uma demonstração alternativa do fato de que $\Sigma p_n \varphi_n$ é função característica, provando que $\Sigma p_n F_n$ é função de distribuição, onde F_n é a função de distribuição correspondente a φ_n.

(b) Prove que $\varphi(t) = \frac{1 + e^{-t/2}}{2}$ é uma função característica. Exiba a função de distribuição correspondente. Essa distribuição é de que tipo?

(c) Mostre que φ definida por $\varphi(t) = \sum_{n=1}^{\infty} p_n \cos(a_n t)$ é função característica, onde $a_n \in \mathbb{R}$, $p_n \geq 0$ e $\Sigma p_n = 1$.

38. O resultado do exemplo 10 e o exercício 37(a) pode ser ainda demonstrado por um terceiro método: sejam $X_1, X_2, \ldots$ e X variáveis aleatórias independentes tais que X é discreta com $P(X = n) = p_n$, $n \geq 1$, e X_n possui função de distribuição F_n, $n \geq 1$. Seja Y a variável aleatória que é igual a X_n quando $X = n$, de modo que

$$Y = \sum_{n \geq 1} X_n I_{[X=n]}.$$

Mostre que $\varphi = \Sigma p_n \varphi_n$ é a função característica de Y, usando esperança condicional.

CAPÍTULO VII

O TEOREMA CENTRAL DO LIMITE

§7.1 O Teorema Central do Limite para seqüências de variáveis aleatórias

Consideremos uma seqüência de variáveis aleatórias independentes, $X_1, X_2, \ldots$, definidas no mesmo espaço de probabilidade $(\Omega, \mathscr{A}, P)$, e seja $S_1, S_2, \ldots$ a seqüência de somas parciais, definidas por $S_n = X_1 + \ldots + X_n$. Recordemos que a Lei dos Grandes Números trata da convergência de $\frac{1}{n}(S_n - ES_n)$ para zero, quando $n \to \infty$, supondo que as variáveis X_n 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 independentes e identicamente distribuídas, para concentrar-se em torno de sua média. Veremos neste capítulo que sob certas hipóteses gerais, a sua distribuição quando padronizada tende à normal.

O *problema central do limite* trata da convergência em distribuição das somas parciais normalizadas, $\frac{S_n - ES_n}{\sqrt{\text{Var } S_n}}$, para a distribuição normal-padrão $N(0,1)$. Para tanto, supõe-se que todas as variâncias sejam finitas e que pelo menos uma delas seja estritamente positiva. O problema consiste em achar condições sob as quais

$$\frac{S_n - ES_n}{\sqrt{\text{Var } S_n}} \xrightarrow{D} N(0,1).$$

Já vimos uma solução para o problema central do limite no caso de variáveis independentes e identicamente distribuídas: se as X_n pos-

suem média μ e variância σ^2, onde $0 < \sigma^2 < \infty$, então

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1).$$

Notemos, então, a diferença entre o Teorema Central do Limite e a Lei dos Grandes Números neste caso. A Lei dos Grandes Números diz que a média amostral $\frac{S_n}{n}$ converge para μ, em probabilidade ou quase certamente, i. e., a diferença $\frac{S_n}{n} - \mu$ tende para zero, e o Teorema Central do Limite diz que esta diferença, quando multiplicada pela raiz quadrada de n, converge em distribuição para uma normal:

$$\sqrt{n}\left(\frac{S_n}{n} - \mu\right) \xrightarrow{D} N(0, \sigma^2).$$

Enunciaremos agora o Teorema Central do Limite de Lindeberg, que dá condições gerais para validade da convergência normal (a prova será dada mais adiante).

TEOREMA 7.1. (*Teorema Central do Limite de Lindeberg.*) *Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes tais que $EX_n = \mu_n$ e $Var\ X_n = \sigma_n^2$, onde $\sigma_n^2 < \infty$ e pelo menos um $\sigma_n^2 > 0$. Sejam $F_n = F_{X_n}$, $S_n = X_1 + \ldots + X_n$ e $s_n = \sqrt{Var\ S_n} = \sqrt{\sigma_1^2 + \ldots + \sigma_n^2}$. Então, para que*

$$\frac{S_n - ES_n}{s_n} \xrightarrow{D} N(0,1) \text{ quando } n \to \infty,$$

é suficiente que a seguinte condição, chamada condição de Lindeberg, esteja satisfeita:

$$\forall\, \varepsilon > 0,\ \lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x-\mu_k| > \varepsilon s_n} (x - \mu_k)^2\, dF_k(x) = 0.$$

Em outras palavras, se a condição de Lindeberg está satisfeita, vale a convergência normal.

Observação. A notação $\int_{|x-\mu_k| > \varepsilon s_n}$ significa que a integração é feita em $\{x : |x - \mu_k| > \varepsilon s_n\} = (-\infty, \mu_k - \varepsilon s_n) \cup (\mu_k + \varepsilon s_n, +\infty)$. Lembre-se que os dois extremos não estão incluídos na região de integração, pois os intervalos são abertos. Se X_k for discreta, com função de probabilidade $p_k(x_i)$, então

$$\int_{|x-\mu_k| > \varepsilon s_n} (x - \mu_k)^2\, dF_k(x) = \sum_{i:\, |x_i - \mu_k| > \varepsilon s_n} (x_i - \mu_k)^2\, p_k(x_i).$$

Por outro lado, se X_k tiver densidade $f_k(x)$, então

$$\int_{|x-\mu_k| > \varepsilon s_n} (x - \mu_k)^2 dF_k(x) = \int_{-\infty}^{\mu_k - \varepsilon s_n} (x - \mu_k)^2 f_k(x)dx + \int_{\mu_k + \varepsilon s_n}^{+\infty} (x - \mu_k)^2 f_k(x)dx.$$

Notemos também que

$$\sigma_k^2 = \int (x - \mu_k)^2 dF_k(x) = \int_{|x-\mu_k| > \varepsilon s_n} (x - \mu_k)^2 dF_k(x) + \int_{|x-\mu_k| \le \varepsilon s_n} (x - \mu_k)^2 dF_k(x),$$

de modo que a condição de Lindeberg pode ser escrita da seguinte forma, utilizando-se o fato de que $s_n^2 = \sigma_1^2 + \ldots + \sigma_n^2$:

$$\forall\, \varepsilon > 0,\ \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x-\mu_k| \le \varepsilon s_n} (x - \mu_k)^2 dF_k(x) \to 1 \text{ quando } n \to \infty.$$

A condição de Lindeberg significa, basicamente, que as parcelas $\frac{X_k - \mu_k}{s_n}$ da soma $\frac{S_n - ES_n}{s_n}$ são uniformemente pequenas para n grande. Por exemplo, a condição de Lindeberg implica

$$\max_{1 \le k \le n} \frac{\sigma_k^2}{s_n^2} \to 0 \text{ quando } n \to \infty,$$

ou seja, para n grande, as variâncias das parcelas são uniformemente pequenas em relação à variância da soma. Para ver isto, observe que para todo k,

$$\frac{\sigma_k^2}{s_n^2} = \frac{1}{s_n^2} \int_{|x-\mu_k| \le \varepsilon s_n} (x - \mu_k)^2 dF_k(x) + \frac{1}{s_n^2} \int_{|x-\mu_k| > \varepsilon s_n} (x - \mu_k)^2 dF_k(x) \le$$

$$\leq \frac{1}{s_n^2}\int_{|x-\mu_k|\leq \varepsilon s_n} \varepsilon^2 s_n^2\, dF_k(x) + \frac{1}{s_n^2}\sum_{j=1}^{n}\int_{|x-\mu_j|>\varepsilon s_n}(x-\mu_j)^2 dF_j(x) \leq$$

$$\leq \frac{1}{s_n^2}\int_{-\infty}^{\infty} \varepsilon^2 s_n^2\, dF_k(x) + \frac{1}{s_n^2}\sum_{j=1}^{n}\int_{|x-\mu_j|>\varepsilon s_n}(x-\mu_j)^2 dF_j(x).$$

Este último termo não depende de k, pois a primeira parcela é igual a ε^2. Portanto, temos

$$\max_{1\leq k\leq n}\frac{\sigma_k^2}{s_n^2} \leq \varepsilon^2 + \frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x-\mu_k|>\varepsilon s_n}(x-\mu_k)^2 dF_k(x),$$

que converge para ε^2, pela condição de Lindeberg. Como vale para todo $\varepsilon > 0$, temos $\max(\sigma_k^2/s_n^2) \to 0$.

Uma vez que essa condição quer dizer que as parcelas $\dfrac{X_k - \mu_k}{s_n}$ possuem variâncias uniformemente pequenas quando n é grande, podemos dizer que nenhuma parcela tem muito peso na soma $\dfrac{S_n - ES_n}{s_n}$. Do ponto de vista intuitivo, isso serve para justificar a afirmação: a soma de um grande número de pequenas quantidades independentes e de média zero tem aproximadamente a distribuição normal.

Observemos que a condição de Lindeberg é formalmente mais forte que a mera condição dada acima sobre o máximo das variâncias. Como $s_n^2 = \sum_{k=1}^{n}\int(x-\mu_k)^2 dF_k(x)$, a condição de Lindeberg diz que quando n é grande, é pequena a parte da variância da soma devida às "caudas" das X_k situadas a mais de ε desvios-padrão s_n das suas respectivas médias μ_k.

É interessante, porém, que na presença da condição sobre o máximo, a condição de Lindeberg torna-se necessária para a validade do Teorema Central do Limite. Essa *recíproca para o Teorema de Lindeberg* deve-se a Feller: *se $X_1, X_2, \ldots$ são independentes com variâncias finitas $\sigma_1^2, \sigma_2^2, \ldots$, se pelo menos um $\sigma_n^2 > 0$ e se $\max_{1\leq k\leq n}(\sigma_k^2/s_n^2) \to 0$ quando $n \to \infty$, então a condição de Lindeberg é consequência da convergência normal, i. e.,*

$$\frac{S_n - ES_n}{s_n} \xrightarrow{D} N(0,1) \Rightarrow \forall\, \varepsilon > 0,$$

$$\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x-\mu_k|>\varepsilon s_n}(x-\mu_k)^2 dF_k(x) \underset{n\to\infty}{\to} 0.$$

(Referência: Feller [9], Teorema 2 do §XV.6. Notemos que a condição dada por Feller, a saber, $s_n^2 \to \infty$ e $\sigma_n^2/s_n^2 \to 0$, é equivalente à condição sobre o máximo. Veja o exercício 6.)

Observação histórica. A distribuição normal era chamada historicamente de *lei dos erros*. Foi usada por Gauss para modelar erros em observações astronômicas, e por isso é freqüentemente chamada de *distribuição gaussiana*. Gauss derivou a distribuição normal, não como limite de somas de variáveis aleatórias independentes, mas a partir de certas hipóteses consideradas naturais para a distribuição de erros, entre elas a de considerar a média aritmética das observações o "valor mais provável" da quantidade sob observação (veja Maistrov [15], §III.10).

Hoje em dia 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, muitas situações em que se pode justificar o uso da normal através do Teorema Central do Limite, embora não necessariamente sejam casos sujeitos a erros de observações. Por exemplo, a distribuição de alturas de homens adultos de certa cidade pode ser considerada aproximadamente normal, pois altura pode ser pensada como soma de muitos efeitos pequenos e independentes.

A distribuição normal não originou com Gauss. Apareceu, pelo menos discretamente, nos trabalhos de De Moivre, que provou o Teorema Central do Limite para o caso de ensaios de Bernoulli com $p = \dfrac{1}{2}$ (parte do Teorema de De Moivre-Laplace).

Antes de provar o Teorema de Lindeberg (não provaremos a recíproca de Feller), consideremos dois corolários e um exemplo. Primeiro, um resultado já provado diretamente no Capítulo 6.

COROLÁRIO 1. *Se $X_1, X_2, \ldots$ são variáveis aleatórias independentes e identicamente distribuídas com $EX_n = \mu$ e $Var\, X_n = \sigma^2$, onde $0 < \sigma^2 < \infty$, então*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1) \text{ quando } n \to \infty.$$

Prova. Verifiquemos a condição de Lindeberg: $s_n^2 = n\sigma^2$ e, para $\varepsilon > 0$,

$$\frac{1}{n\sigma^2} \sum_{k=1}^{n} \int_{|x-\mu| \le \varepsilon\sigma\sqrt{n}} (x-\mu)^2 dF_k(x) = \text{(são identicamente}$$

$$\text{distribuídas)} = \frac{1}{\sigma^2} \int_{|x-\mu| \le \varepsilon\sigma\sqrt{n}} (x-\mu)^2 dF_1(x)$$

$$\xrightarrow[n\to\infty]{} \frac{1}{\sigma^2} \int_{-\infty}^{\infty} (x-\mu)^2 dF_1(x) = \frac{\sigma^2}{\sigma^2} = 1,$$

onde a convergência decorre da definição da integral imprópria de Riemann-Stieltjes. □

O segundo corolário é o Teorema de Liapunov, que pode ser muito útil quando as variáveis X_n possuem momentos finitos de ordem maior que 2. Afirma este teorema que vale a convergência normal se a soma dos momentos centrais absolutos de ordem $2+\delta$ é assintoticamente pequena em relação a $s_n^{2+\delta}$.

COROLÁRIO 2. (*Teorema Central do Limite de Liapunov.*) *Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes tais que $EX_n = \mu_n$ e Var $X_n = \sigma_n^2 < \infty$, com pelo menos um $\sigma_n^2 > 0$. Seja $s_n^2 =$ = Var $S_n = \sigma_1^2 + \ldots + \sigma_n^2$. Se existir $\delta > 0$ tal que*

$$\frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} E|X_k - \mu_k|^{2+\delta} \xrightarrow[n\to\infty]{} 0,$$

então

$$\frac{S_n - ES_n}{S_n} \xrightarrow{D} N(0,1).$$

Prova. Verifiquemos a condição de Lindeberg, supondo satisfeita a condição de Liapunov. Para $\varepsilon > 0$, se $|x-\mu_k| > \varepsilon s_n$ então $|x-\mu_k|^\delta/(\varepsilon^\delta s_n^\delta) > 1$, de modo que

$$\frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x-\mu_k| > \varepsilon s_n} (x-\mu_k)^2 dF_k(x) \le$$

$$\le \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x-\mu_k| > \varepsilon s_n} (x-\mu_k)^2 \cdot \frac{|x-\mu_k|^\delta}{\varepsilon^\delta s_n^\delta} dF_k(x) \le$$

$$\le \frac{1}{\varepsilon^\delta s_n^{2+\delta}} \sum_{k=1}^{n} \int_{|x-\mu_k| > \varepsilon s_n} |x-\mu_k|^{2+\delta} dF_k(x) \le$$

$$\le \frac{1}{\varepsilon^\delta s_n^{2+\delta}} \sum_{k=1}^{n} \int_{-\infty}^{\infty} |x-\mu_k|^{2+\delta} dF_k(x)$$

$$= \frac{1}{\varepsilon^\delta s_n^{2+\delta}} \sum_{k=1}^{n} E|X_k - \mu_k|^{2+\delta} \to 0 \text{ quando } n \to \infty. \quad \square$$

EXEMPLO 1. Sejam $X_1, X_2, \ldots$ independentes, $X_n \sim U[-n, n]$. Mostraremos de duas maneiras que $\frac{S_n - ES_n}{S_n} \xrightarrow{D} N(0,1)$, verificando as condições de Lindeberg e Liapunov.

Primeiro, a condição de Lindeberg. Como $\mu_k = EX_k = 0$ e $\sigma_k^2 =$ $= \text{Var } X_k = EX_k^2 = \frac{1}{2k} \int_{-k}^{k} x^2 dx = \frac{k^2}{3}$, temos

$$s_n^2 = \sum_{k=1}^{n} \frac{k^2}{3}.$$

Acontece que não precisamos lembrar, nem calcular, o valor desta soma $\left(\text{é igual a } \frac{n(n+1)(2n+1)}{18}\right)$. De fato, veremos que o importante é a ordem de s_n^2 e não o seu valor exato. Usaremos o seguinte lema, que trata da ordem de séries do tipo Σn^λ:

LEMA 7.1. *Para $\lambda > 0$,*

$$\frac{1}{n^{\lambda+1}} \sum_{k=1}^{n} k^\lambda \to \frac{1}{\lambda+1} \text{ quando } n \to \infty,$$

de maneira que $\sum_{k=1}^{n} k^\lambda$ é da ordem de $n^{\lambda+1}$.

Prova do lema. Como $x^\lambda \le k^\lambda$ se $k-1 \le x \le k$ e $k^\lambda \le x^\lambda$ se $k \le x \le k+1$, segue-se que

$$\int_{k-1}^{k} x^\lambda dx \le \int_{k-1}^{k} k^\lambda dx = k^\lambda = \int_{k}^{k+1} k^\lambda dx \le \int_{k}^{k+1} x^\lambda dx$$

e, portanto, somando-se em k de 1 até n,

$$\int_0^n x^\lambda dx \le \sum_{k=1}^{n} k^\lambda \le \int_1^{n+1} x^\lambda dx.$$

Logo,

$$\frac{n^{\lambda+1}}{\lambda+1} \leq \sum_{k=1}^{n} k^{\lambda} \leq \frac{(n+1)^{\lambda+1}-1}{\lambda+1} \leq \frac{(n+1)^{\lambda+1}}{\lambda+1},$$

ou equivalentemente,

$$\frac{1}{\lambda+1} \leq \frac{1}{n^{\lambda+1}} \sum_{k=1}^{n} k^{\lambda} \leq \frac{1}{\lambda+1} \cdot \left(\frac{n+1}{n}\right)^{\lambda+1}.$$

Como $\left(\frac{n+1}{n}\right)^{\lambda+1} \to 1$ quando $n \to \infty$, o lema está provado. □

Voltando ao exemplo, consideremos a parcela

$$\int_{|x-\mu_k|>\varepsilon s_n} (x-\mu_k)^2 dF_k(x) = \int_{|x|>\varepsilon s_n} x^2 dF_k(x).$$

Como a densidade de X_k toma o valor zero fora do intervalo $[-k, k]$, e $\varepsilon s_n > n$ para n suficientemente grande (pelo lema, s_n^2 é da ordem de n^3, logo s_n é da ordem de $n^{3/2}$), temos que, para n suficientemente grande, a densidade assume o valor zero na região de integração, para todo $k \leq n$. Isto quer dizer que para todo n suficientemente grande, todas as parcelas são nulas e a condição de Lindeberg está satisfeita.

Vejamos uma demonstração formal, utilizando propriedades da integral de Stieltjes: para $1 \leq k \leq n$,

$$\int_{|x|>\varepsilon s_n} x^2 dF_k(x) = \int x^2 I_{\{x \in \mathbb{R}:\, |x|>\varepsilon s_n\}}(x) dF_k(x) = (\text{pela densidade}) = \frac{1}{2k} \int_{-k}^{k} x^2 I_{\{x \in \mathbb{R}:\, |x|>\varepsilon s_n\}}(x) dx,$$

e esta última integral é nula se $n < \varepsilon s_n$, pois, neste caso, o integrando toma o valor zero em $[-k, k]$. Como o lema implica que $s_n^2/n^3 \to \frac{1}{9}$, temos

$$\frac{s_n^2}{n^2} = \frac{s_n^2}{n^3} \cdot n \to +\infty,$$

de modo que $n < \varepsilon s_n$ para n suficientemente grande.

Verificaremos, agora, a condição de Liapunov para $\delta = 1$. Temos

$$E|X_k - \mu_k|^3 = E|X_k|^3 = \frac{1}{2k} \int_{-k}^{k} |x|^3 dx = \frac{1}{k} \int_0^k x^3 dx = \frac{k^3}{4}.$$

Aplicando o lema, temos que $\sum_{k=1}^{n} E|X_k - \mu_k|^3$ é da ordem de n^4. Foi visto acima que s_n é da ordem de $n^{3/2}$, logo s_n^3 é da ordem de $n^{9/2}$. Como s_n^3 é de maior ordem que a soma dos momentos, a condição de Liapunov está satisfeita. (Demonstração formal, utilizando o lema:

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} E|X_k - \mu_k|^3}{n^4} = \frac{1}{16},$$

$$\lim_{n\to\infty} \frac{s_n^3}{n^{9/2}} = \lim_{n\to\infty} \left(\frac{s_n^2}{n^3}\right)^{3/2} = \left(\frac{1}{9}\right)^{3/2} = \frac{1}{27}.$$

Portanto,

$$\lim_{n\to\infty} \frac{1}{s_n^3} \sum_{k=1}^{n} E|X_k - \mu_k|^3 = \lim_{n\to\infty} \left(\frac{n^{9/2}}{s_n^3} \cdot \frac{\sum_{k=1}^{n} E|X_k - \mu_k|^3}{n^4} \cdot \frac{1}{n^{1/2}}\right) = 27 \cdot \frac{1}{16} \cdot \lim_{n\to\infty} \frac{1}{n^{1/2}} = 0.\Big)$$

Observação. Podemos verificar a condição de Liapunov para todo δ, neste exemplo. Mas basta verificar para *um* valor de δ, e o valor 1 é muito conveniente.

Agora, provaremos o Teorema de Lindeberg. Para facilitar a leitura da prova, recordemos a condição de Lindeberg:

$$\forall\, \varepsilon > 0,\ \lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x-\mu_k|>\varepsilon s_n} (x-\mu_k)^2 dF_k(x) = 0.$$

Prova do Teorema de Lindeberg. Mostraremos que as funções características das somas parciais padronizadas convergem para a função característica da $N(0, 1)$, i. e., $\forall t \in \mathbb{R}$,

$$\varphi_{(S_n - ES_n)/s_n}(t) = (\text{por independência}) = \prod_{k=1}^{n} Ee^{it((X_k-\mu_k)/s_n)} \xrightarrow[n\to\infty]{} e^{-t^2/2}. \tag{7.1}$$

Para tanto, fixemos $t \in \mathbb{R}$. Usaremos duas versões da fórmula de Taylor aplicada à função $g(x) = e^{itx}$ (veja Feller [9], Lema 1 do §XV.4):

$$e^{itx} = 1 + itx + \theta_1(x)\frac{t^2x^2}{2}, \text{ onde } |\theta_1(x)| \leq 1,$$

$$e^{itx} = 1 + itx - \frac{t^2x^2}{2} + \theta_2(x)\frac{t^3x^3}{6}, \text{ onde } |\theta_2(x)| \leq 1.$$

Seja $\varepsilon > 0$. Usando a primeira fórmula para $|x| > \varepsilon$ e a segunda para $|x| \leq \varepsilon$, podemos escrever e^{itx} da seguinte forma geral:

$$e^{itx} = 1 + itx - \frac{t^2x^2}{2} + r_\varepsilon(x),$$

onde

$$r_\varepsilon(x) = \begin{cases} \{1 + \theta_1(x)\}\dfrac{t^2x^2}{2}, \text{ se } |x| > \varepsilon \\ \\ \theta_2(x)\dfrac{t^3x^3}{6}, \text{ se } |x| \leq \varepsilon. \end{cases}$$

Portanto,

$$E\exp\left\{it\left(\frac{X_k - \mu_k}{s_n}\right)\right\} = \int \exp\left\{it\left(\frac{x - \mu_k}{s_n}\right)\right\} dF_k(x) =$$

$$= \int\left\{1 + it\left(\frac{x - \mu_k}{s_n}\right) - \frac{t^2}{2}\left(\frac{x - \mu_k}{s_n}\right)^2 + r_\varepsilon\left(\frac{x - \mu_k}{s_n}\right)\right\} dF_k(x) =$$

$$= (\text{por linearidade}) = 1 + it\,E\left(\frac{X_k - \mu_k}{s_n}\right) - \frac{t^2}{2}E\left(\frac{X_k - \mu_k}{s_n}\right)^2 +$$

$$+ \frac{t^2}{2}\int_{|x-\mu_k| > \varepsilon s_n}\left\{1 + \theta_1\left(\frac{x - \mu_k}{s_n}\right)\right\}\left(\frac{x - \mu_k}{s_n}\right)^2 dF_k(x) +$$

$$+ \frac{t^3}{6}\int_{|x-\mu_k| \leq \varepsilon s_n} \theta_2\left(\frac{x - \mu_k}{s_n}\right)\left(\frac{x - \mu_k}{s_n}\right)^3 dF_k(x).$$

Como $EX_k = \mu_k$ e $\text{Var } X_k = \sigma_k^2$, temos

$$E\exp\left\{it\left(\frac{X_k - \mu_k}{s_n}\right)\right\} = 1 - \frac{t^2\sigma_k^2}{2s_n^2} + e_{n,k}.$$

onde o resto $e_{n,k}$ satisfaz (lembremo-nos que $|\theta_1(x)| \leq 1$ e $|\theta_2(x)| \leq 1$)

$$|e_{n,k}| \leq t^2 \int_{|x-\mu_k| > \varepsilon s_n}\left(\frac{x - \mu_k}{s_n}\right)^2 dF_k(x) +$$

$$+ \frac{|t|^3}{6}\int_{|x-\mu_k| \leq \varepsilon s_n} \varepsilon \cdot \left(\frac{x - \mu_k}{s_n}\right)^2 dF_k(x) \leq$$

$$\leq \frac{t^2}{s_n^2}\int_{|x-\mu_k| > \varepsilon s_n}(x - \mu_k)^2 dF_k(x) + \frac{\varepsilon|t|^3}{6s_n^2}\int_{-\infty}^{\infty}(x - \mu_k)^2 dF_k(x).$$

Temos, então,

$$\sum_{k=1}^{n}|e_{n,k}| \leq \frac{t^2}{s_n^2}\sum_{k=1}^{n}\int_{|x-\mu_k| > \varepsilon s_n}(x - \mu_k)^2 dF_k(x) + \frac{\varepsilon|t|^3}{6}.$$

Pela condição de Lindeberg, a primeira parcela do termo à direita tende a zero quando $n \to \infty$. Logo, para n suficientemente grande, $\sum_{k=1}^{n}|e_{n,k}| \leq \frac{\varepsilon|t|^3}{3}$.

Vamos, então, escolher uma seqüência de ε's que converge para zero. Para $\varepsilon = \frac{1}{m}$, existe n_m tal que para $n \geq n_m$, $\sum_{k=1}^{n}|e_{n,k}| \leq \frac{|t|^3}{3m}$, onde os restos $e_{n,k}$ são os determinados pela fórmula baseada em $\varepsilon = \frac{1}{m}$. Portanto, existe uma seqüência $n_1 < n_2 < n_3 < \ldots$ de inteiros positivos tal que

$$\sum_{k=1}^{n}|e_{n,k}| \leq \frac{|t|^3}{3m},$$

para $n_m \leq n < n_{m+1}$, onde para estes valores de n os restos $e_{n,k}$ são baseados em $\varepsilon = \frac{1}{m}$. Por economia de notação, a dependência de m não será expressa, mas é importante lembrar durante o restante da prova que o valor de ε que determina o resto $e_{n,k}$ depende da posição de n em relação aos n_m. Temos, então,

$$\sum_{k=1}^{n}|e_{n,k}| \to 0 \text{ quando } n \to \infty. \tag{7.2}$$

Substituindo em (7.1), vemos que

$$\varphi_{(S_n - ES_n)/s_n}(t) = \prod_{k=1}^{n}\left(1 - \frac{t^2\sigma_k^2}{2s_n^2} + e_{n,k}\right).$$

com os $e_{n,k}$ satisfazendo (7.2). Para provar que o termo à direita converge para $e^{-t^2/2}$, usaremos o seguinte lema sobre números complexos, que generaliza o resultado já utilizado para provar o Teorema Central do Limite no caso de variáveis independentes e identicamente distribuídas, de que $c_n \to c$ implica que $\left(1 + \frac{c_n}{n}\right)^n \to e^c$.

LEMA 7.2. *Sejam $c_{n,k}$ números complexos tais que $\sum_{k=1}^{n} c_{n,k} \to c$ quando $n \to \infty$. Se*

$$\max_{1 \le k \le n} |c_{n,k}| \to 0 \text{ quando } n \to \infty$$

e

$$\sum_{k=1}^{n} |c_{n,k}| \le M < \infty,$$

onde M é uma constante que não depende de n, então

$$\prod_{k=1}^{n} (1 + c_{n,k}) \to e^c \text{ quando } n \to \infty.$$

Prova do lema. Veja Chung [7], §7.1. □

Em nosso caso, sejam $c_{n,k} = -\{t^2\sigma_k^2/(2s_n^2)\} + e_{n,k}$ e $c = -t^2/2$. Por (7.2),

$$\sum_{k=1}^{n} |c_{n,k}| \le \frac{t^2}{2} + \sum_{k=1}^{n} |e_{n,k}| \xrightarrow[n \to \infty]{} \frac{t^2}{2},$$

logo $\sum_{k=1}^{n} |c_{n,k}|$ é uniformemente limitado (i.e., existe $M < \infty$ tal que $\forall n$, $\sum_{k=1}^{n} |c_{n,k}| \le M$). Para aplicar o lema, resta verificarmos a condição sobre o máximo:

$$\max_{1 \le k \le n} |c_{n,k}| \le \max_{1 \le k \le n} \frac{t^2\sigma_k^2}{2s_n^2} + \max_{1 \le k \le n} |e_{n,k}| \le$$

$$\le \frac{t^2}{2} \max_{1 \le k \le n} \frac{\sigma_k^2}{s_n^2} + \sum_{k=1}^{n} |e_{n,k}|,$$

com o segundo termo tendendo a zero por (7.2). Como já foi visto que a condição de Lindeberg implica que $\max_{1 \le k \le n} (\sigma_k^2/s_n^2) \to 0$, a prova está terminada. □

§7.2 A distribuição normal multivariada

Antes de estendermos o Teorema Central do Limite ao caso de uma seqüência de vetores aleatórios independentes, consideremos a extensão da definição de distribuição normal ao caso vetorial. Por consistência de notação, trabalharemos somente com vetores-linha em vez de coluna. Na literatura da análise multivariada, é comum o uso de vetores-coluna para representar vetores aleatórios, mas o leitor não deverá sentir dificuldade na "tradução": as modificações necessárias deverão ser óbvias.

Diremos que um vetor aleatório tem *distribuição normal multivariada* se possui a mesma distribuição de uma transformação afim de normais-padrão independentes. Isso significa o seguinte: se $X_1, \ldots, X_n$ são independentes com distribuição comum $N(0, 1)$, então o vetor aleatório $\underset{\sim}{Y} = (Y_1, \ldots, Y_n)$, onde

$$Y_j = a_{1j}X_1 + \ldots + a_{nj}X_n + \mu_j,$$

para $j = 1, \ldots, n$, possui distribuição normal n-variada. Aqui, as constantes a_{ij} e μ_j são números reais quaisquer.

Colocando as equações que definem $\underset{\sim}{Y}$ em forma matricial, escrevendo $\underset{\sim}{\mu} = (\mu_1, \ldots, \mu_n)$ e

$$A = \begin{pmatrix} a_{11} & a_{12} \ldots a_{1n} \\ a_{21} & a_{22} \ldots a_{2n} \\ \vdots & \vdots \quad \vdots \\ a_{n1} & a_{n2} \ldots a_{nn} \end{pmatrix},$$

temos

$$\underset{\sim}{Y} = \underset{\sim}{X}A + \underset{\sim}{\mu}.$$

Notemos que Y_j, sendo combinação linear de normais independentes, também é normal. De fato, é evidente que

$$Y_j \sim N\left(\mu_j, \sum_{k=1}^{n} a_{kj}^2\right).$$

Além disso, a matriz de covariâncias de $\underset{\sim}{Y}$ é $\Sigma_{\underset{\sim}{Y}} = \Sigma = A'A$, onde A' é a matriz transposta de A e a *matriz de covariâncias* de um vetor aleatório $\underset{\sim}{Y}$ é, por definição, a matriz das covariâncias entre os componentes de Y, i. e.,

$$\Sigma_{ij} = \text{Cov}(Y_i, Y_j), \quad 1 \le i \le n, \quad 1 \le j \le n.$$

Verificando este resultado, temos (notemos que $EX_iX_j = 1$ ou 0, dependendo de i ser ou não igual a j)

$$\text{Cov}(Y_i, Y_j) = E((Y_i - \mu_i)(Y_j - \mu_j)) =$$
$$= E((a_{1i}X_1 + \ldots + a_{ni}X_n)(a_{1j}X_1 + \ldots + a_{nj}X_n)) = \text{(linearidade)} =$$
$$= \sum_{k=1}^{n} \sum_{\ell=1}^{n} a_{ki}a_{\ell j}\, E(X_kX_\ell) = \sum_{k=1}^{n} a_{ki}a_{kj} = (A'A)_{ij}.$$

DEFINIÇÃO 7.1. Sejam $X_1, \ldots, X_n$ independentes e identicamente distribuídas tais que $X_j \sim N(0, 1)$, e seja $\underset{\sim}{Y}$ o vetor aleatório obtido de $\underset{\sim}{X} = (X_1, \ldots, X_n)$ através da transformação

$$\underset{\sim}{Y} = \underset{\sim}{X}A + \underset{\sim}{\mu},$$

onde A é uma matriz real $n \times n$ e $\underset{\sim}{\mu}$ é um vetor real n-dimensional. Então dizemos que $\underset{\sim}{Y}$ tem *distribuição normal n-variada* com média $\underset{\sim}{\mu}$ e matriz de covariâncias $\Sigma = A'A$. *Notação*: $\underset{\sim}{Y} \sim N(\underset{\sim}{\mu}, \Sigma)$.

A matriz de covariâncias é *simétrica* e *definida não-negativa*. A primeira propriedade decorre do fato de que $\text{Cov}(Y_i, Y_j) = \text{Cov}(Y_j, Y_i)$. Para verificar a segunda, basta observar que se $\underset{\sim}{x} \in \mathbb{R}^n$, então

$$\underset{\sim}{x}\,\Sigma\,\underset{\sim}{x}' = \underset{\sim}{x}\,A'A\,\underset{\sim}{x}' = (\underset{\sim}{x}\,A')(\underset{\sim}{x}\,A')' = |\underset{\sim}{x}\,A'|^2 \geq 0,$$

estando satisfeita a definição de matriz definida não-negativa. (Nota: a propriedade de ser simétrica e definida não-negativa não é apenas da matriz de covariâncias de um vetor com distribuição normal multivariada, mas também de toda matriz de covariâncias finita. Veja o exercício 13.)

Se uma matriz $n \times n$ Σ é definida não-negativa, então ela possui a representação $\Sigma = A'A$ para alguma matriz $n \times n$ A. Portanto, chegamos à conclusão de que toda matriz definida não-negativa é matriz de covariâncias de um vetor normal multivariado, pois basta utilizar a matriz A na Definição 7.1. Resumindo, então, temos a seguinte

PROPOSIÇÃO 7.1. *Uma matriz $n \times n$ Σ é matriz de covariâncias de algum vetor normal n-variado se, e somente se, é definida não-negativa.*

Quando o posto da matriz A é igual a n e, portanto, A é invertível, sabemos de Álgebra Linear que a transformação

$$\underset{\sim}{y} = T(\underset{\sim}{x}) = \underset{\sim}{x}\,A + \underset{\sim}{\mu}$$

é uma correspondência biunívoca entre $\mathbb{R}^n$ e $\mathbb{R}^n$. Neste caso, o jacobiano $J(\underset{\sim}{y}, \underset{\sim}{x}) = \det(A) \neq 0$, e a transformação inversa é dada por $\underset{\sim}{x} = (\underset{\sim}{y} - \underset{\sim}{\mu})A^{-1}$, de modo que

$$\sum_{i=1}^{n} x_i^2 = \underset{\sim}{x} \cdot \underset{\sim}{x} = \underset{\sim}{x}\,\underset{\sim}{x}' = (\underset{\sim}{y} - \underset{\sim}{\mu})A^{-1}(A^{-1})'(\underset{\sim}{y} - \underset{\sim}{\mu})' =$$
$$= (\underset{\sim}{y} - \underset{\sim}{\mu})A^{-1}(A')^{-1}(\underset{\sim}{y} - \underset{\sim}{\mu})' = (\underset{\sim}{y} - \underset{\sim}{\mu})(A'A)^{-1}(\underset{\sim}{y} - \underset{\sim}{\mu})'.$$

Portanto, pelo método do jacobiano, $\underset{\sim}{Y} = (Y_1, \ldots, Y_n)$ possui densidade dada por

$$f_{\underset{\sim}{Y}}(y_1, \ldots, y_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-(\underset{\sim}{y}-\underset{\sim}{\mu})(A'A)^{-1}(\underset{\sim}{y}-\underset{\sim}{\mu})'/2} \cdot \frac{1}{|\det(A)|} =$$
$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-(\underset{\sim}{y}-\underset{\sim}{\mu})\Sigma_{\underset{\sim}{Y}}^{-1}(\underset{\sim}{y}-\underset{\sim}{\mu})'/2} \cdot \frac{1}{\sqrt{\det(\Sigma_{\underset{\sim}{Y}})}};$$

para $\underset{\sim}{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$, onde a última passagem é conseqüência do fato de que $\Sigma_{\underset{\sim}{Y}} = A'A$ e $\det(\Sigma_{\underset{\sim}{Y}}) = \det(A'A) = (\det(A))^2 > 0$.

Se A é *ortonormal*, i. e., $A'A = I_n$ = matriz identidade $n \times n$, então $\Sigma_{\underset{\sim}{Y}} = I_n$, $\det(\Sigma_{\underset{\sim}{Y}}) = 1$ e a densidade conjunta de $Y_1, \ldots, Y_n$ torna-se

$$f_{\underset{\sim}{Y}}(y_1, \ldots, y_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-(\underset{\sim}{y}-\underset{\sim}{\mu})(\underset{\sim}{y}-\underset{\sim}{\mu})'/2} =$$
$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu_i)^2}, \quad (y_1, \ldots, y_n) \in \mathbb{R}^n.$$

Neste caso, a densidade é fatorável e vemos que $Y_1, \ldots, Y_n$ são independentes, com $Y_i \sim N(\mu_i, 1)$.

EXEMPLO 2. Sejam $\underset{\sim}{\mu} = (0, 0)$ e

$$A = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix},$$

de maneira que A é ortonormal. Então se X_1 e X_2 forem independentes e $N(0, 1)$, $Y_1 = \dfrac{X_1}{\sqrt{2}} + \dfrac{X_2}{\sqrt{2}}$ e $Y_2 = \dfrac{X_1}{\sqrt{2}} - \dfrac{X_2}{\sqrt{2}}$ também serão independentes e $N(0, 1)$. Logo $X_1 + X_2$ e $X_1 - X_2$ serão independentes e identicamente distribuídas, com distribuição comum $N(0, 2)$.

Analogamente, se A é *ortogonal*, i. e., $A'A$ é uma matriz diagonal com elementos diagonais $d_1, \ldots, d_n$, todos eles positivos ($d_i > 0$, $1 \le i \le n$), então $\det(\Sigma_{\underset{\sim}{Y}}) = \prod_{i=1}^{n} d_i$ e a densidade conjunta torna-se

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \frac{1}{(\sqrt{2\pi})^n \prod_{i=1}^{n} d_i} e^{-(\underset{\sim}{y}-\underset{\sim}{\mu})\Sigma_{\underset{\sim}{Y}}^{-1}(\underset{\sim}{y}-\underset{\sim}{\mu})'/2} =$$

$$= \frac{1}{(\sqrt{2\pi})^n \prod_{i=1}^{n} d_i} e^{-\frac{1}{2}\sum_{i=1}^{n} \frac{(y_i-\mu_i)^2}{d_i}},$$

onde utilizamos o fato de que $\Sigma_{\underset{\sim}{Y}}^{-1}$ também é matriz diagonal, tendo elementos diagonais $\frac{1}{d_1}, \ldots, \frac{1}{d_n}$. Neste caso, vemos que $Y_1, \ldots, Y_n$ são independentes, com $Y_i \sim N(\mu_i, d_i)$.

Se a matriz A não é invertível, o jacobiano é nulo e não há densidade. Nesse caso a imagem da transformação $\underset{\sim}{y} = T(\underset{\sim}{x}) = \underset{\sim}{x}A + \underset{\sim}{\mu}$ é um hiperplano no $\mathbb{R}^n$, de dimensão menor que n, e dizemos que $\underset{\sim}{Y}$ tem distribuição normal n-variada *degenerada*. (Veja o exercício 14 para um exemplo.)

Calculemos agora a *função característica* de um vetor normal n-variado $\underset{\sim}{Y}$. Para tanto, sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes e $N(0, 1)$, de modo que $\underset{\sim}{Y}$ possui a mesma distribuição de uma transformação afim de $\underset{\sim}{X} = (X_1, \ldots, X_n)$, ou seja,

$$\underset{\sim}{Y} \sim \underset{\sim}{X}A + \underset{\sim}{\mu}.$$

Então, a função característica de $\underset{\sim}{Y}$ é

$$\varphi_{\underset{\sim}{Y}}(t_1, \ldots, t_n) = E\, e^{i\underset{\sim}{t}\cdot\underset{\sim}{Y}} = E\, e^{i\underset{\sim}{t}\underset{\sim}{Y}'} = E\, e^{i\underset{\sim}{t}(\underset{\sim}{X}A+\underset{\sim}{\mu})'} =$$
$$= e^{i\underset{\sim}{t}\underset{\sim}{\mu}'} E\, e^{i\underset{\sim}{t}A\underset{\sim}{X}'}.$$

Notemos que se uma variável aleatória Z é combinação linear das X_j, digamos $Z = \sum_{j=1}^{n} b_j X_j$, então Z possui distribuição normal com média $\mu = 0$ e variância $\sigma^2 = \sum_{j=1}^{n} b_j^2 = \underset{\sim}{b}\underset{\sim}{b}'$. Por isso, a variável aleatória $\underset{\sim}{t}A'\underset{\sim}{X}'$, sendo combinação linear das X_j com coeficientes $b_j = (\underset{\sim}{t}A')_j = \sum_{i=1}^{n} t_i A'_{ij}$, tem distribuição normal com média 0 e variância

$$\sigma^2 = \underset{\sim}{b}\,\underset{\sim}{b}' = (\underset{\sim}{t}A')(\underset{\sim}{t}A')' = \underset{\sim}{t}A'A\underset{\sim}{t}' = \underset{\sim}{t}\Sigma\,\underset{\sim}{t}',$$

ou seja, $\underset{\sim}{t}A'\underset{\sim}{X}' \sim N(0, \underset{\sim}{t}\Sigma\,\underset{\sim}{t}')$.

Já que a função característica da $N(0, \sigma^2)$ é $\varphi(t) = \exp\left(-\frac{t^2\sigma^2}{2}\right)$, temos

$$E\, e^{i\underset{\sim}{t}A\underset{\sim}{X}'} = \varphi_{\underset{\sim}{t}A\underset{\sim}{X}'}(1) = e^{-\sigma^2/2} = e^{-\underset{\sim}{t}\Sigma\underset{\sim}{t}'/2},$$

de modo que a função característica de $\underset{\sim}{Y}$ é

$$\varphi_{\underset{\sim}{Y}}(\underset{\sim}{t}) = \exp\left\{i\,\underset{\sim}{t}\,\underset{\sim}{\mu}' - \frac{\underset{\sim}{t}\Sigma\,\underset{\sim}{t}'}{2}\right\}, \quad \underset{\sim}{t} \in \mathbb{R}^n.$$

Observação. Pelo Teorema da Unicidade, a função característica determina a distribuição. Logo a distribuição de um vetor normal n-variado é determinada pela média μ e pela matriz de covariâncias Σ. A matriz A entra na distribuição somente através da matriz $\Sigma = A'A$, e duas matrizes A_1 e A_2 tais que $A_1'A_1 = A_2'A_2$ dão origem à mesma distribuição normal multivariada. Assim, justifica-se o uso da notação $N(\underset{\sim}{\mu}, \Sigma)$ na Definição 7.1.

§7.3 O Teorema Central do Limite — caso multivariado

O clássico Teorema Central do Limite do caso univariado diz que a soma de um grande número de variáveis aleatórias independentes e identicamente distribuídas, com variância comum finita e estritamente positiva, tem distribuição aproximadamente normal. A versão apresentada no exemplo 9 do §6.4, que diz respeito à normalidade assintótica da média amostral $\bar{X}_n = \frac{X_1 + \ldots + X_n}{n}$, é a versão mais usada na prática. Essa versão diz que quando $X_1, X_2, \ldots$ são independentes e identicamente distribuídas com média comum μ e variância comum σ^2, onde $\sigma^2 < \infty$, então

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} N(0, \sigma^2).$$

(Se $\sigma^2 = 0$, interpretamos $N(0, \sigma^2)$ como massa pontual em 0.)

Ocorre que vale uma versão análoga desse resultado para uma seqüência de *vetores* aleatórios independentes e identicamente distribuídos. Neste caso, a média amostral, agora um vetor, quando apropriamente normalizada através da subtração da média μ seguida da multiplicação pela raiz de n, converge em distribuição para a normal

multivariada com média $\underset{\sim}{0}$ e matriz de covariâncias Σ, onde Σ é a matriz de covariâncias (suposta finita) comum aos vetores.

Como no caso de variáveis aleatórias, dizemos que os vetores aleatórios $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots$ são *independentes* se $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ são independentes para todo $n \geq 2$; eles são *identicamente distribuídos* se possuem a mesma distribuição, e para tanto, é evidente que os vetores precisam ser de mesma dimensão. Portanto, se $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots$ são vetores aleatórios k-dimensionais, definidos no mesmo espaço de probabilidade $(\Omega, \mathscr{A}, P)$, eles são independentes e identicamente distribuídos se, e somente se,

$$P(\underset{\sim}{X}_1 \in B_1, \ldots, \underset{\sim}{X}_n \in B_n) = \prod_{j=1}^{n} P(\underset{\sim}{X}_j \in B_j),$$

para todo $n = 2, 3, \ldots$ e para toda escolha dos borelianos k-dimensionais $B_1, \ldots, B_n$ (i. e., $\forall B_j \in \mathscr{B}^k$, $\forall j \leq n$, $\forall n \geq 2$).

Se $\underset{\sim}{X} = (X_1, \ldots, X_k)$ é um vetor aleatório k-dimensional, dizemos que tem média finita se $EX_j = \mu_j$ finito para $j \leq k$, e neste caso $\underset{\sim}{\mu} = (\mu_1, \ldots, \mu_k)$ é chamado *média* de $\underset{\sim}{X}$. Dizemos que $\underset{\sim}{X}$ tem *variância finita* se a sua matriz de covariâncias Σ é finita, i. e., se $\mathrm{Cov}(X_i, X_j)$ é finita para todo par (i, j), $1 \leq i \leq k$, $1 \leq j \leq k$. Então, uma condição necessária e suficiente para $\underset{\sim}{X}$ possuir variância finita é que a variância de X_j seja finita $\forall j$, $1 \leq j \leq k$, e é óbvio que neste caso $\underset{\sim}{X}$ também possui média finita.

Quando a matriz de covariâncias Σ for finita, ela será definida não-negativa (exercício 13), e pela discussão da última seção, existirá a distribuição normal k-variada $N(\underset{\sim}{0}, \Sigma)$. Assim, estará garantida a existência da distribuição limite no teorema que se segue.

TEOREMA 7.2. (*Teorema Central do Limite para vetores.*) *Sejam $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots$ vetores aleatórios k-dimensionais, independentes e identicamente distribuídos. Suponha que $\underset{\sim}{X}_1$ tenha variância finita, e sejam $\underset{\sim}{\mu}$ a média e Σ a matriz de covariâncias de $\underset{\sim}{X}_1$. Seja $\overline{\underset{\sim}{X}}_n$ a média amostral, definida como a média aritmética dos vetores $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$. Então*

$$\sqrt{n}(\overline{\underset{\sim}{X}}_n - \underset{\sim}{\mu}) \xrightarrow{D} N(\underset{\sim}{0}, \Sigma), \text{ quando } n \to \infty.$$

Observação. Se $\underset{\sim}{X}_j = (X_{j1}, \ldots, X_{jk})$, para $j = 1, 2, \ldots$, então

$$\overline{\underset{\sim}{X}}_n = \frac{\underset{\sim}{X}_1 + \ldots + \underset{\sim}{X}_n}{n} = \left(\frac{X_{11} + \ldots + X_{n1}}{n}, \ldots, \frac{X_{1k} + \ldots + X_{nk}}{n} \right).$$

Prova. Pelo Teorema 6.3, basta provar que a função característica k-dimensional de $\sqrt{n}(\overline{\underset{\sim}{X}}_n - \underset{\sim}{\mu})$ converge para a função característica da $N(\underset{\sim}{0}, \Sigma)$ em todo ponto $\underset{\sim}{t} \in \mathbb{R}^k$, ou seja, basta provar que

$$\varphi_{\sqrt{n}(\overline{\underset{\sim}{X}}_n - \underset{\sim}{\mu})}(\underset{\sim}{t}) \xrightarrow[n \to \infty]{} e^{-\underset{\sim}{t}\Sigma\underset{\sim}{t}'/2}, \quad \forall \underset{\sim}{t} \in \mathbb{R}^k.$$

Pelo método de Cramér e Wold (Proposição 6.1), basta provar que temos a apropriada convergência em distribuição para toda combinação linear das coordenadas do vetor $\sqrt{n}(\overline{\underset{\sim}{X}}_n - \underset{\sim}{\mu})$. Especificamente, basta provar que se $\underset{\sim}{X} = (X_1, \ldots, X_k)$ possui distribuição $N(\underset{\sim}{0}, \Sigma)$, então $\forall \underset{\sim}{t} \in \mathbb{R}^k$,

$$\sqrt{n} \sum_{j=1}^{k} t_j \left(\frac{X_{1j} + \ldots + X_{nj}}{n} - \mu_j \right) \xrightarrow{D} \sum_{j=1}^{k} t_j X_j \text{ quando } n \to \infty. \qquad (7.3)$$

Agora, o termo à direita de (7.3) possui distribuição normal univariada. De fato, é fácil obtermos a sua distribuição a partir da definição de distribuição normal multivariada. Para tanto, notemos que existe uma matriz $k \times k$, A, tal que $\underset{\sim}{X}$ tem a mesma distribuição que $\underset{\sim}{Y}A$, onde $A'A = \Sigma$ e $\underset{\sim}{Y} = (Y_1, \ldots, Y_k)$, com as Y_j independentes e $N(0, 1)$. Isto significa $\underset{\sim}{X} \sim \underset{\sim}{Y}A$ e

$$\sum_{j=1}^{k} t_j X_j = \underset{\sim}{t}\underset{\sim}{X}' \sim \underset{\sim}{t}A'\underset{\sim}{Y}',$$

de maneira que $\sum_{j=1}^{k} t_j X_j$ tem distribuição normal com média zero e variância igual à soma dos quadrados das coordenadas do vetor $\underset{\sim}{t}A'$, i. e.,

$$\mathrm{Var}\left(\sum_{j=1}^{k} t_j X_j \right) = \underset{\sim}{t}A'(\underset{\sim}{t}A')' = \underset{\sim}{t}A'A\underset{\sim}{t}'.$$

Como $A'A = \Sigma$, temos, então,

$$\sum_{j=1}^{k} t_j X_j \sim N(0, \underset{\sim}{t}\Sigma\underset{\sim}{t}').$$

Rearranjando as parcelas do termo à esquerda de (7.3), obtemos

$$\sqrt{n} \sum_{j=1}^{k} \sum_{m=1}^{n} t_j \left(\frac{X_{mj} - \mu_j}{n} \right) = \frac{1}{\sqrt{n}} \sum_{m=1}^{n} \sum_{j=1}^{k} t_j (X_{mj} - \mu_j) = \frac{1}{\sqrt{n}} \sum_{m=1}^{n} Z_m,$$

onde $Z_m = \sum_{j=1}^{k} t_j (X_{mj} - \mu_j)$.

Como $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots$ são independentes e identicamente distribuídos, as variáveis aleatórias Z_m também o são, pois são obtidas dos $\underset{\sim}{X}_m$ através da mesma função $g(\underset{\sim}{x}) = \underset{\sim}{t}(\underset{\sim}{x} - \underset{\sim}{\mu})'$. Além disso, as Z_m possuem média zero (pois $EX_{mj} = \mu_j$) e variância $\underset{\sim}{t}\,\underset{\sim}{\Sigma}\,\underset{\sim}{t}'$, como verificaremos agora:

$$\operatorname{Var} Z_1 = EZ_1^2 = E\left[\left(\sum_{i=1}^{k} t_i(X_{1i} - \mu_i)\right)\left(\sum_{j=1}^{k} t_j(X_{1j} - \mu_j)\right)\right] =$$

$$= (\text{linearidade}) = \sum_{i=1}^{k}\sum_{j=1}^{k} t_i t_j E[(X_{1i} - \mu_i)(X_{1j} - \mu_j)] =$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{k} t_i t_j \operatorname{Cov}(X_{1i}, X_{1j}) = \sum_{i=1}^{k}\sum_{j=1}^{k} t_i t_j \Sigma_{ij} = \underset{\sim}{t}\,\underset{\sim}{\Sigma}\,\underset{\sim}{t}'.$$

Aplicando o Teorema Central do Limite (univariado), temos

$$\frac{Z_1 + \ldots + Z_n}{\sqrt{n}} \xrightarrow{D} N(0, \underset{\sim}{t}\,\underset{\sim}{\Sigma}\,\underset{\sim}{t}') \text{ quando } n \to \infty.$$

Assim, (7.3) está verificada. □

EXERCÍCIOS DO CAPÍTULO 7

§7.1

1. Na geração de números aleatórios por computador, o objetivo do programador é conseguir que os números satisfaçam as condições (i) cada número tem distribuição uniforme em $[0, 1]$ e (ii) os números são independentes. Sejam $X_1, X_2 \ldots$ variáveis aleatórias representando uma seqüência de números aleatórios gerados por computador (números "pseudo-aleatórios"). Supondo que as condições estejam satisfeitas, enuncie versões da Lei dos Grandes Números e do Teorema Central do Limite para este caso. Dê uma explicação intuitiva do significado dos dois teoremas neste caso.
2. Fregueses chegam em certo supermercado segundo um processo de Poisson com intensidade média de dez por minuto. Sejam $T_1, T_2, \ldots$ os tempos entre chegadas de fregueses, de modo que $T_1 + \ldots + T_n$ é o tempo de chegada do n-ésimo freguês.

(a) Utilizando o Teorema Central do Limite, ache um número entre 0 e 1 que seja aproximadamente igual à probabilidade do milésimo freguês chegar depois de 100 minutos.
(b) Como você calcularia o valor *exato* da probabilidade no item (a)? (Não se aceita uma integral em $\mathbb{R}^{1000}$.)

3. Seja $(X_n)_{n \geq 1}$ uma seqüência de variáveis aleatórias independentes tais que X_n tem distribuição uniforme em $[-0, n]$, $\forall n$. Mostre que a condição de Lindeberg está satisfeita e enuncie o Teorema Central do Limite resultante. (Calcule os parâmetros!)
4. Suponha que $X_1, X_2, \ldots$ sejam variáveis aleatórias independentes tais que $P(X_n = -n) = \frac{1}{2} = P(X_n = n)$. Mostre que a seqüência satisfaz o Teorema Central do Limite mas não obedece à Lei Forte dos Grandes Números.
5. Usando o Teorema Central do Limite para variáveis aleatórias de Poisson, mostre que

$$\lim_{n \to \infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} = \frac{1}{2}.$$

6. (Duas condições equivalentes no enunciado do Teorema do Feller.) Sejam $\sigma_1^2, \sigma_2^2, \ldots$ números não-negativos tais que pelo menos um $\sigma_n^2 > 0$, e seja $s_n^2 = \sigma_1^2 + \ldots + \sigma_n^2$. Mostre que $\max_{1 \leq k \leq n} \frac{\sigma_k^2}{s_n^2} \to 0$ quando $n \to \infty$ se, e somente se, $s_n^2 \to \infty$ e $\frac{\sigma_n^2}{s_n^2} \to 0$. (*Sugestão.* Para a necessidade da primeira condição, mostre que o limite não depende dos primeiros N termos, para qualquer N fixo, e use o fato de que $\frac{\sigma_k^2}{s_n^2} \leq \frac{\sigma_k^2}{s_k^2}$.)
7. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes, X_n tendo densidade

$$f_n(x) = \frac{1}{2n} e^{-|x|/n}, \quad x \in \mathbb{R}.$$

Seja $S_n = X_1 + \ldots + X_n$. Demonstre que

$$\frac{S_n - ES_n}{\sqrt{\operatorname{Var} S_n}} \xrightarrow{D} N(0, 1).$$

(*Sugestão.* Use Liapunov.)

8. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes tais que

$$P(X_n = n^\alpha) = P(X_n = -n^\alpha) = \frac{1}{2}.$$

Mostre que se $\alpha > -\frac{1}{2}$ então vale o Teorema Central do Limite. (*Sugestão.* Use Liapunov.)

9. Seja $X_1, Y_1, X_2, Y_2, X_3, \ldots$ uma seqüência de variáveis aleatórias independentes, as X_n sendo identicamente distribuídas com distribuição $U[0, 1]$ e as Y_n sendo identicamente distribuídas com distribuição $U[0, 2]$. Seja S_n a soma dos n primeiros termos da seqüência, de modo que $S_1 = X_1$, $S_2 = X_1 + Y_1$, $S_3 = X_1 + Y_1 + X_2$, etc.

(a) Mostre que $\frac{S_n}{n}$ converge quase certamente e ache o seu limite.

(b) Mostre que $\frac{S_n - ES_n}{\sqrt{\text{Var } S_n}} \xrightarrow{D} N(0, 1)$.

10. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas tais que $EX_1 = \mu$ e $\text{Var } X_1 = \sigma^2$, onde $0 < \sigma^2 < \infty$. Seja $(a_n)_{n \geq 1}$ uma seqüência qualquer de números reais. Mostre que

$$P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq a_n\right) - \Phi(a_n) \to 0, \text{ quando } n \to \infty,$$

onde Φ é a função distribuição da $N(0, 1)$. Note que esse resultado é valido até no caso de não convergência de $(a_n)_{n \geq 1}$. (*Sugestão.* Use o resultado do exercício 16 do Capítulo 6.)

11. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes tais que $X_k \sim b(n_k, p)$, onde $0 < p < 1$ (p fixo). Vale o Teorema Central do Limite neste caso? Por quê? (*Sugestão.* Não tente verificar a condição de Lindeberg.)

12. Explique se a seqüência $Z_1, Z_2, \ldots$ do exercício 18 do Capítulo 5 satisfaz o Teorema Central do Limite. Se satisfaz, calcule todos os parâmetros.

§7.2

13. Seja $\underset{\sim}{X} = (X_1, \ldots, X_k)$ um vetor aleatório tal que $\text{Var } X_j < \infty$ para $j = 1, \ldots, k$.
(a) Prove que a matriz de covariâncias de $\underset{\sim}{X}$ é finita.
(b) Seja Σ a matriz de covariâncias de $\underset{\sim}{X}$. Mostre que Σ é simétrica e definida não-negativa. (*Observação.* Σ é definida não-negativa, por definição, se $\forall\, \underset{\sim}{x} \in \mathbb{R}^k$ $\underset{\sim}{x}\, \Sigma\, \underset{\sim}{x}' \geq 0$.)

14. Sejam X e Y independentes e identicamente distribuídas, com distribuição $N(0, 1)$.
(a) Mostre que o vetor (X, X) possui distribuição normal bivariada degenerada. (*Sugestão.* Qual a matriz A neste caso?) Qual o hiperplano (aqui é uma reta) que contém os valores do vetor?
(b) Mostre que $(X - Y,\ 2X - 2Y + 2)$ tem distribuição normal bivariada degenerada. Qual a reta contendo os valores do vetor?

15. Mostre que um vetor aleatório com distribuição $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, como definida no exemplo 15 do Capítulo 2 (§2.5), possui distribuição normal bivariada segundo a Definição 7.1.

16. Demonstre que se $\underset{\sim}{Y} = (Y_1, \ldots, Y_n)$ é normal n-variada e a matriz de covariâncias $\Sigma_{\underset{\sim}{Y}}$ é diagonal, então $Y_1, \ldots, Y_n$ são independentes. Portanto, se $Y_1, \ldots, Y_n$ têm distribuição conjunta normal n-variada e $\text{Cov}(Y_i, Y_j) = 0$ para $i \neq j$, então $Y_1, \ldots, Y_n$ são independentes. (*Sugestão.* Função característica.)

17. Se $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ são vetores aleatórios k-dimensionais tais que $\underset{\sim}{X}$ e $\underset{\sim}{Y}$ são independentes com $\underset{\sim}{X} \sim N(\underset{\sim}{\mu}, \Sigma_{\underset{\sim}{X}})$, $\underset{\sim}{Y} \sim N(\underset{\sim}{\xi}, \Sigma_{\underset{\sim}{Y}})$, qual a distribuição de $\underset{\sim}{Z} = \underset{\sim}{X} + \underset{\sim}{Y}$? Generalize esse resultado para a soma de n vetores independentes.

18. Sejam X e Y variáveis aleatórias independentes, cada uma com distribuição normal. Prove que $X + Y$ e $X - Y$ são independentes se, e somente se, $\text{Var } X = \text{Var } Y$. (Compare com o exercício 29 do Capítulo 3.)

19. Seja $\underset{\sim}{Y} = (Y_1, \ldots, Y_n)$ um vetor aleatório tendo distribuição normal n-variada com média $\underset{\sim}{\mu}$ e matriz de covariâncias $\Sigma_{\underset{\sim}{Y}} = A'A$.
(a) Se B é uma matriz real $n \times k$ e $\underset{\sim}{\xi}$ um vetor real k-dimensional, mostre que o vetor aleatório $\underset{\sim}{Z}$ definido por $\underset{\sim}{Z} = \underset{\sim}{Y}B + \underset{\sim}{\xi}$ tem distribuição normal k-variada com média $\underset{\sim}{\mu}B + \underset{\sim}{\xi}$ e matriz de co-

variâncias $\Sigma_{\underset{\sim}{Z}} = B'A'AB$. (*Sugestão.* Obtenha a função característica de $\underset{\sim}{Z}$).

(b) Partindo do item (a), obtenha a distribuição de Y_j, para $j = 1, \ldots, n$. Confira sua solução com a obtida no início do §7.2.

(c) Mostre que todo vetor de dimensão $m < n$, cujos componentes são componentes diferentes de $\underset{\sim}{Y}$ tem distribuição normal m-variada (por exemplo, (Y_2, Y_5) é normal bivariada), e mostre que sua matriz de covariâncias é uma submatriz de $\Sigma_{\underset{\sim}{Y}}$.

20. Mostre que um vetor aleatório $\underset{\sim}{X}$ possui distribuição normal multivariada se, e somente se, toda combinação linear dos componentes de $\underset{\sim}{X}$ tem distribuição normal univariada (i.e., $\underset{\sim}{X} = (X_1, \ldots, X_n)$ é normal multivariada $\Leftrightarrow \sum_{k=1}^{k} a_k X_k$ é normal $\forall (a_1, \ldots, a_n) \in \mathbb{R}^n$).

§7.3

21. Seja $(X_{j1}, \ldots, X_{jk})$, $j = 1, \ldots, n$, uma amostra aleatória de tamanho n de uma distribuição k-dimensional com média $\underset{\sim}{\mu} = (\mu_1, \ldots, \mu_k)$ finita e matriz de covariâncias $\Sigma = (\sigma_{ij})$. Seja $(\bar{X}_1, \ldots, \bar{X}_k)$ o vetor de médias (média amostral). Demonstre que se $\delta_j > 0$, $j = 1, \ldots, k$, temos

$$P(|\bar{X}_j - \mu_j| < \delta_j,\ j = 1, \ldots, k) \geq 1 - \frac{1}{n}\left(\sum_{j=1}^{k} \frac{\sigma_{jj}}{\delta_j^2}\right).$$

22. (O Teorema Central do Limite para a distribuição multinomial.) Seja $\underset{\sim}{X} = (X_1, \ldots, X_k)$ um vetor aleatório tal que os componentes X_j assumem somente os valores 0 e 1, com $\sum_{j=1}^{k} X_j = 1$. Seja $p_j = EX_j$, $j = 1, \ldots, k$.

(a) Mostre que $\mathrm{Cov}(X_j, X_\ell) = -p_j p_\ell$, se $j \neq \ell$.

(b) Prove que se $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ são vetores aleatórios independentes e identicamente distribuídos, cada um tendo a mesma distribuição que $\underset{\sim}{X}$, então $\underset{\sim}{S}_n = \underset{\sim}{X}_1 + \ldots + \underset{\sim}{X}_n$ tem distribuição multinomial com parâmetros $p_1, \ldots, p_k$ e n.

(c) Suponha que $\underset{\sim}{N} = (N_1, \ldots, N_k)$ possua distribuição multinomial com parâmetros $p_1, \ldots, p_k$ e n. Calcule a covariância entre N_j e N_ℓ, onde $j \neq \ell$, utilizando o item (b) e a bilinearidade da covariância. (Compare com o resultado obtido no exercício 27 do Capítulo 6.)

(d) Enuncie um Teorema Central do Limite para os vetores aleatórios $\underset{\sim}{S}_1, \underset{\sim}{S}_2, \ldots$ definidos no item (b).

REFERÊNCIAS

[1] AHLFORS, L.V. — *Complex Analysis*, 2ª edição. McGraw-Hill, N. York, 1966.

[2] APOSTOL, T.M. — *Mathematical Analysis*. Addison-Wesley, Reading, Massachusetts, 1957.

[3] BICKEL, P.J. E DOKSUM, K.A. — *Mathematical Statistics*. Holden-Day, S. Francisco, 1977.

[4] BILLINGSLEY, P. — *Convergence of Probability Measures*. Wiley, N. York, 1968.

[5] BREIMAN, L. — *Probability*. Addison-Wesley, Reading, Massachusetts, 1968.

[6] BREIMAN, L. — *Probability and Stochastic Processes*. Houghton Mifflin, Boston, 1969.

[7] CHUNG, K.L. — *A Course in Probability Theory*, 2ª edição. Academic Press, N. York, 1974.

[8] FELLER, W. — *Introdução à Teoria das Probabilidades e Suas Aplicações, Parte I*. Edgard Blücher, São Paulo, 1976. Tradução parcial do Volume I, 3ª edição americana (1968). Traduzido por Flávio Wagner Rodrigues e Maria Eliza Fini.

[9] FELLER, W. — *An Introduction to Probability Theory and Its Applications, Volume II*. Wiley, N. York, 1966.

[10] FERNÁNDEZ, P.J. — *Introdução à Teoria das Probabilidades*. Livros Técnicos e Científicos, Rio de Janeiro; 1973. Coleção Elementos de Matemática, IMPA.

[11] GNEDENKO, B. — *The Theory of Probability*. Mir, Moscou, 1976. Tradução para o inglês de *Kurs Teoriia Veroiatnostei*.

[12] KOLMOGOROV, A.N. — *Foundations of the Theory of Probability*. Chelsea, N. York, 1950. Tradução para o inglês de *Grundbegriffe der Wahrscheinlichkeits rechnung* (1933).

[13] LEHMANN, E.L. — *Testing Statistic Hypotheses*. Wiley, N. York, 1959.

[14] LIMA, E.L. — *Curso de Análise, Volume 1*. Projeto Euclides, CNPq, 1976.

[15] MAISTROV, L.E. — *Probability Theory: A Historical Sketch*. Academic Press, N. York, 1974. Tradução para o inglês de *Teoriia Veroiatnostei Istoricheskii Ocherk* (1967).

[16] RUDIN, W. — *Princípios de Análise Matemática*. Ao Livro Técnico, Rio de Janeiro, 1971. Traduzido por Eliana Rocha Henriques de Brito da 2ª edição americana (1964).

[17] RUDIN, W. — *Real and Complex Analysis*. McGraw-Hill, N. York, 1966.

ÍNDICE DE NOTAÇÕES

O símbolo "$\sim$" significa "tem como distribuição" ou "está distribuído como" (por exemplo: $X \sim N(0, 1)$).

Ω	1	$N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$	63, 283
$\mathbb{R}$	1	I_G	65
$\emptyset$	3	$U(G)$	65
$\mathscr{A}$	5	$g^{-1}(B)$	67
$A - B$	6	$\sim \exp(\lambda)$	68
$\mathscr{P}(\Omega)$	7	$\chi^2(n)$	85
$\mathscr{B}_{[0,1]}$	8	EX	107
$\mathscr{B}_1$	8	$\mathrm{Var}X$	121
$\mathscr{B}_2$	8	σ_X	121
$\mathbb{R}^2$	8	$\mathrm{Cov}(X, Y)$	129
$\mathscr{B}^n$	9	$\rho(X, Y)$	131
$P(A)$	9	$F_X(x\|A)$	145
$A_n \downarrow \emptyset$	11	$E(X\|A)$	145
$A_n \uparrow A$, $A_n \downarrow A$	13	$F_X(x\|Y = y)$	147, 162
$P(A\|B)$	14	$E(X\|Y = y)$	147, 174
$\log x$ (= $\log_e x$)	24	$E(X\|Y)$	147, 174
$\limsup A_n$	29, 197	$P(X \in B\|Y = y)$	155, 162
$\liminf A_n$	29, 197	$f(x\|y)$	159
$\lim A_n$	29, 198	$Y_n \xrightarrow{P} Y$	192
c, $\tilde{c}$	35	[A_n infinitas vezes]	198
$\neq A$	35	e^{ix}	221
$[X \leqslant x]$	36	φ_X	221
F_X	37	$\bar{c}$	222
$F(x-)$	38, 112	$\mathbb{C}$	229
$N(0, 1)$	40, 50	$X_n \xrightarrow{D} X$	231
$[x]$	40	$X_n \xrightarrow{D} N(0\ 1)$	236
$U[0, 1]$	43	$\varphi_{\underline{X}}$	238
$[X \in B]$	47	Σ_Y	273
$N(\mu, \sigma^2)$	51	$N(\underline{\mu}, \Sigma)$	274
$\Gamma(\alpha, \beta)$	52		
$b(n, p)$	53		
$F_{\underline{X}}$	55		

ÍNDICE ALFABÉTICO

Absolutamente contínua 41, 42, 58
Aditividade
  da integral de Stieltjes 100
  finita 10
Álgebra 5
  propriedades de 5
Amostra aleatória 82, 96
  da distribuição uniforme
    densidade conjunta de 84
  densidade conjunta de 83
Ao acaso 4, 91, 94
Axiomas de probabilidade 9
  satisfeitos pela distribuição 49
  satisfeitos pela distribuição condicional 144
  sistemas equivalentes de 12

Bayes, fórmula de 17
Bernoulli 191
  ensaios de 54, 191
  Lei dos Grandes Números de 196
Bienaymé-Tchebychev, desigualdade de 123
Binominal, veja distribuição binomial
Bochner-Khintchin, Teorema de 252
Bonferroni, desigualdades de 30, 32
Borel, Lei Forte de 212
  números normais de 212
Borel-Cantelli, Lema de 198
Borelianos 8

Carathéodory, Teorema da Extensão de 6
Casamentos, problema de 30
Cauchy, veja distribuição de Cauchy
Cauchy-Schwarz, desigualdade de 178
Coeficiente de correlação 131
  da normal bivariada 64, 178
Complexo conjugado 222
Condição de Lindeberg 262
  forma alternativa de 263
Continuidade
  à direita 38, 56
  de probabilidade 13
  no vazio 11
  uniforme 223
    da função característica 223
Convergência
  de séries 137
  em distribuição 231
    de vetores aleatórios 239
    não implica convergência em probabilidade 246
    no caso contínuo 244
    no caso discreto 242, 257
    para a normal-padrão 237
  em lei 231
  em probabilidade 192
    implica convergência em distribuição 245
    não implica convergência quase certa 194
  fraca de funções de distribuição 233
  preservada por funções contínuas 247
  quase certa 192
    implica convergência em probabilidade 193
Convolução
  de binomiais 253
  de densidades 71
  de funções de distribuição 180, 225
  de gamas 85
  de normais 71
  de Poissons 92
Coordenadas 36
Covariância 129
  bilinearidade de 142
  condicional 187
  matriz de 273
Cramér-Wold, artifício de 240
Critério
  para independência 60
    no caso contínuo 62
    no caso discreto 89
  para integrabilidade 115, 140

Decomposição da função de distribuição 46
Definida
  não-negativa (matriz) 274, 283
  positiva (função) 251
De Moivre 265
De Moivre-Laplace, Teorema Central do Limite de 238
Densidade 42
  como derivada 42, 84
  condicional 159, 183
    para vetores 185
  conjunta 58
    pelo método de jacobiano 76, 79
  critério para f ser 42, 84
  da soma de duas variáveis aleatórias 71
  de um vetor aleatório 58
  estelar 33
  marginal 63, 64
  não unicidade de 64
Desigualdade
  "básica" 122
  de Bonferroni 30, 32
  de Cauchy-Schwarz 178
  de Jensen 114, 116, 141
    estrita 141
    para esperança condicional 175

  de Kolmogorov 204
  de Markov 123
  de Tchebychev, clássica 123
    generalizada 122
Desvio-padrão 121, 131
Discreto 41, 59
Distribuição
  binominal 53
    como distribuição condicional de Poissons 153
    como limite de hipergeométricas 243
    convergência para a Poisson 244
    convolução de 255
  condicional
    como limite 155, 163
    dada Y discreta 147
    dado que Y = y 162
      como achar 163
    dado um evento 144
    de um vetor dada Y discreta 150
    existência de 156, 162
    para vetores 169
    regular 162
  conjunta 59
  de Cauchy 52, 138
    como quociente de duas normais 81
    padrão 52
      função característica de 253
  de Laplace 138
  de Poisson 40
    como limite de binomiais 244
    composta 253
    convolução de 92
    esperança de 104
    função característica de 228
    função de probabilidade de 53
  de Rayleigh 91
    taxa de falha de 87
  de um vetor aleatório 59
    no caso contínuo 59
    no caso discreto 59
  de uma variável aleatória 49
    no caso contínuo 49
    no caso discreto 49
    representações de 50
  de Weibull 93
    taxa de falha de 87
  exponencial 41
    como uma gama 53
    esperança de 111
    falta de memória de 86
    função característica de 254
    mínimo também é 92
    momentos de 119
  exponencial dupla 138
  F 85
  gama 52
    convolução de 85
    função característica de 254
  gaussiana 265
  geométrica 54, 180
    esperança de 112
  hipergeométrica 243
    convergência para a binomial 243
  logística 138
  multinomial 181
    função característica de 257
    Teorema Central do Limite para 284
  normal 50-51
    bivariada 63, 64, 71
      coeficiente de correlação em 178
      densidade condicional em 161
    convolução de 71
    média de 130
    multivariada 273
      caracterização de 284
    n-variada 274
      degenerada 276
      função característica de 276
    -padrão (N(0,1)) 50
      função característica de 229
      simulação de 95
    variância de 130
  qui-quadrado 85
    como uma gama 53
  simétrica 137
    critério para 227
    de vetores 256
  t de Student 85
  uniforme 65
    distribuição condicional em 184
    em [a,b] 88
    em (0,1) 43
      como distribuição de F(x) 92
    estatísticas de ordem de 84, 152
    singular 66

Ensaios 9, 54
  binomiais 54, 191, 196, 208
  de Bernoulli 54, 191
Erro
  absoluto médio 124
  quadrático médio 124
Espaço
  amostral 1
    condições satisfeitas por 2
  de probabilidade 13
  L^p 124
  produto 189
Esperança 107
  como esperança da esperança condicional 147, 175, 186
  como integral de Riemann 108
  condicional 145, 147, 174
    princípio da substituição para 175
    propriedades de 175
  de uma constante 113
  de uma distribuição condicional 145, 174

de uma função de X 118
no caso contínuo 120
no caso discreto 120
de uma função de X 127
no caso contínuo 127
no caso de independência 128
no caso discreto 127
de uma variável discreta 104
de variáveis não-negativas 110
e assumindo valores inteiros 111
do produto (se independentes) 128
infinita 107
propriedades de 113-116, 122-124
Estatísticas de ordem 82
densidade conjunta de 83
distribuição condicional quando dadas 172
Estocasticamente maior 139
Evento(s) 2, 3
aleatório 5
álgebra de 5
certo 3
disjuntos 10
elementar 3
impossível 3
incompatíveis 4
independentes 18, 21
linguagem de 4, 27
mutuamente exclusivos 4, 10
Experimento
composto 17, 189
de duas etapas 17, 148, 183
modelo matemático para 13
Exponencial, veja distribuição exponencial

Feller, Teorema Central do Limite (recíproca) 264, 281
Fórmula
da inversão 224, 239
de Bayes 17
Freqüência relativa 9, 15
Função
absolutamente contínua 42
característica 221
caracterização de 251
como função geradora de momentos 227
para vetores 257
conjunta 238
continuidade uniforme de 223
de um vetor aleatório 238
definição alternativa de 222
coordenada 36
de Cantor 44-45
de densidade de probabilidade 42
de distribuição 39
acumulada 37
condicional 145, 162
conjunta 55
de um vetor aleatório 55
de uma variável aleatória 37
exemplos de 40-41
marginal 63, 64
n-dimensional 58
propriedades de 38, 56, 58
de freqüência 41
de probabilidade 41
definida positiva 251
-escada 22
gama 52
geradora de momentos 227, 254
identidade 36
indicadora 65
mensurável 36, 67

Gama, veja distribuição gama
Gauss 265
Geométrica, veja distribuição geométrica, probabilidade geométrica

Helly, Teorema de Compacidade Fraca de 235
Helly-Bray, Teorema de 232
Hipergeométrica, veja distribuição hipergeométrica

Identicamente distribuídos 82, 190
vetores aleatórios 278
i.i.d. 190
Incrementos
estacionários 22
independentes 23
Independência
a pares 20
critério para 60
no caso contínuo 62
no caso discreto 89
de eventos 18, 21
de variáveis aleatórias 60
de vetores aleatórios 239, 241, 278
critério para 241
2 a 2 20
hereditariedade de 60, 72
Indicador 65
esperança condicional de 179
"Infinitas vezes" 198
Integrabilidade
critério para 115, 140
Integrador 97
Integral
de Lebesgue 42, 107
de Lebesgue-Stieltjes 99
de Riemann-Stieltjes 97-99, 103
de Stieltjes 99-103
múltipla 127
Integral 107
se limitada 115

Jacobiano 75
método do 74-84, 95

Jensen, desigualdade de 114, 116, 141
para esperança condicional 175

Khintchin, Lei Fraca de 197
veja Bochner-Khintchin
Kolmogorov 9
desigualdade de 204
Lei Forte de 210
recíproca para 202
primeira Lei Forte de 205

Laplace, veja De Moivre-Laplace
Lebesgue, integral de 42, 107
medida de (nula) 42
Lebesgue-Stieltjes, integral de 99
Lei (= distribuição) 50
dos erros 265
dos Grandes Números 105, 189, 195
de Bernoulli 196
Forte dos Grandes Números 195
de Borel 212
de Kolmogorov 210
primeira 205
recíproca para 202
estendida 219
Fraca dos Grandes Números 195
de Khintchin 197
de Tchebychev 196
Lema de Borel-Cantelli 198
Levy, Paul, Teorema da Continuidade de 234
Liapunov, Teorema Central do Limite de 266
Limite de uma seqüência de eventos 29, 198
inferior 29, 197
superior 29, 197
Lindeberg, condição de 262, 263
Teorema Central do Limite de 262
Linearidade
da esperança 113, 127
da esperança condicional 175
da integral de Stieltjes 100

Markov 197
desigualdade de 123
Massa pontual 164
Matriz de covariâncias 273
é simétrica e definida não-negativa 274, 283
Máximo de uma amostra aleatória 82
densidade conjunta de, com mínimo 96

Média 104
amostral 85, 252
da normal n-variada 274
de um vetor aleatório 278
ponderada 104, 117, 252
Mediana 52, 125
Medida
de Lebesgue (nula) 42
de probabilidade 10
Mensurável 36, 67
Método do jacobiano 74-84, 95
Mínimo de uma amostra aleatória 82
densidade conjunta de, com máximo 96
Modelo probabilístico 13
Momento 120-121
absoluto 121
de ordem k 120
k-ésimo 120
central 120
em torno de b 120
propriedades de 122-124
Multinomial, veja distribuição multinomial
Mutuamente exclusivos 4, 10

Não-correlacionadas 129
condições para implicar independência 142
Normal, veja distribuição normal
Normalização de uma variável aleatória 131
Números
normais 212
"pseudo-aleatórios" 95, 280

Observado 189

Padronização de uma variável aleatória 131
Parâmetro
de configuração 53
de escala 51
de locação 51
do processo de Poisson 27
Partição do espaço amostral 16, 145
Permutações, invariância para 89
Poisson, veja distribuição de Poisson, processo de Poisson
Preditor, melhor 124
Princípio
da indiferença 4
da preservação de chances relativas 159, 166
da substituição 166
para distribuição condicional 167
para esperança condicional 175
Probabilidade 10
"a priori" 18
axiomas definidores de 9
condicional 14
como esperança condicional 179
de que $X \in B$ dado que $Y = y$ 162, 163
de um evento 9
como esperança de probabilidade condicional 179
finitamente aditiva 10
geométrica 4
"posterior" 18
σ-aditiva 10
Problema central do limite 261
Processo de Poisson 22-27, 105, 149-152, 170
análogo espacial 33
fluxo de tráfego como 32

hipóteses para 22
parâmetro (taxa) de 27
probabilidades de eventos em 27
soma de dois independentes 181
variáveis aleatórias de 37

Quartil 52, 87
Quase
certamente (= com probabilidade 1) 123
toda parte 42
Qui-quadrado, veja distribuição qui-quadrado

Resultados
equiprováveis 4
favoráveis 2
possíveis 1
Riemann, somas de 97
Riemann-Stieltjes, integral de 97-99

Scheffé, Teorema de 244
Seqüência
de variáveis aleatórias independentes (existência) 201
decrescente para o vazio 11
σ-aditividade 10
σ-álgebra 6
das partes de Ω 7
de Borel
em $[0,1]$ 8
na reta 8
no plano 8
no $\mathbb{R}^n$ 9, 59
exemplos de 7-8
gerada 6, 28
Simetria de uma distribuição 137
relação com função característica 227
Simulação
de normais-padrão 95
de uniformes 95, 280
Singular 45, 66
Slutsky, Teorema de 248
Somas
de Riemann 97
parciais 194
Stieltjes, integral de 99-104
integral múltipla de 127

Taxa de falha 33, 87
Tchebychev, desigualdade de 122, 123
Lei Fraca de 196
Tempo
de espera 54
paradoxo de 183
de parada 255
Teorema
Central do Limite
de De Moivre-Laplace 238
de Liapunov 266
de Lindeberg 262
prova de 271
recíproca de (Feller) 264, 281
para variáveis i.i.d. 237, 250, 265
implica Lei Fraca 251
para vetores 278
da Continuidade de Paul Lévy 234
da Convergência Dominada 135
para esperança condicional 175
da Convergência Monótona 134, 137
para esperança condicional 175
da Extensão de Carathéodory 6
da Multiplicação 16
da Probabilidade Composta 16
da Probabilidade Total 17
da Unicidade 226, 239
de Arzelà 136
de Bochner-Khintchin 252
de Compacidade Fraca de Helly 235
de Helly-Bray 232
de Scheffé 244
de Slutsky 248

Uniforme, veja distribuição uniforme

Valor esperado 104
Variância 121
amostral 85
condicional 176, 187
de um produto 142
de uma constante 122
de uma soma 130
propriedades 122, 123
Variável aleatória (variáveis aleatórias) 36
complexa 221
contínua (= absolutamente contínua) 41
distribuição de 49
discreta 41
distribuição de 49
função de distribuição de 37
identicamente distribuídas 82, 190
independentes 60
integrável 107
n-dimensional 55
não-correlacionadas 129
não-negativa 110
"mista" 43
padronização de 131
simétrica 168
singular 45
truncada 133, 210
Vetor(es) aleatório(s) 55
contínuo (= absolutamente contínuo) 58
discreto 58
função de distribuição de 55
identicamente distribuídos 278
independência de 239, 241, 278
von Mises 9

Departamento de Estatística
Fac. Ciências Mat. e Naturais
Universidade Federal do Rio de Janeiro



barry r. james

É americano da Pennsylvania, criado em Minnesota, onde faz um frio de rachar.
Bacharelou-se ("summa cum laude") no Williams College, Massachusetts, obteve o Doutorado (Ph.D) em Estatística na Universidade da Califórnia e, um ano depois (1972), veio para o Brasil (leia-se Rio de Janeiro), numa trajetória que faz um ângulo agudo com o gradiente da temperatura, como se vê. Mantida esta condição, dificilmente sairá daqui.
É Pesquisador Associado do IMPA e suas áreas preferidas de pesquisa são Probabilidade e Estatística não-Paramétrica. Gosta de trabalhar em parceria com a esposa Kang Ling, que também doutorou-se em Estatística na Universidade da Califórnia, em Berkeley e é Pesquisadora Associada do IMPA. Ela, não sendo co-autora do livro, não concordou em sair na foto, o que sem dúvida o torna menos atraente.

probabilidade: um curso em nível intermediário

CNPq Conselho Nacional de Desenvolvimento Científico e Tecnológico
impa Instituto de Matemática Pura e Aplicada