Introdução à Teoria das Probabilidades

Publicações Matemáticas

Introdução à Teoria das Probabilidades

Pedro J. Fernandez
FGV

impa

Estrada Dona Castorina, 110
22460-320 Rio de Janeiro, RJ

Impresso no Brasil / Printed in Brazil

Capa: Noni Geiger

Publicações Matemáticas

- Introdução à Análise Funcional – César R. de Oliveira
- Introdução à Topologia Diferencial – Elon Lages Lima
- Les Équations Différentielles Algébriques et les Singularités Mobiles – Ivan Pan e Marcos Sebastiani
- Criptografia, Números Primos e Algoritmos – Manoel Lemos
- Introdução à Economia Dinâmica e Mercados Incompletos – Aloísio Araújo
- Conjuntos de Cantor, Dinâmica e Aritmética – Carlos Gustavo Moreira
- Geometria Hiperbólica – João Lucas Marques Barbosa
- Introdução à Economia Matemática – Aloísio Araújo
- Superfícies Mínimas – Manfredo Perdigão do Carmo
- The Index Formula for Dirac Operators: an Introduction – Levi Lopes de Lima
- Introduction to Symplectic and Hamiltonian Geometry – Ana Cannas da Silva
- Primos de Mersenne (e outros primos muito grandes) – Carlos Gustavo T. A. Moreira e Nicolau Saldanha
- The Contact Process on Graphs – Márcia Salzano
- Canonical Metrics on Compact almost Complex Manifolds – Santiago R. Simanca
- Introduction to Toric Varieties – Jean-Paul Brasselet
- Birational Geometry of Foliations – Marco Brunella
- Introduction to Nonlinear Dispersive Equations – Felipe Linares e Gustavo Ponce
- Introdução à Teoria das Probabilidades – Pedro J. Fernandez

Distribuição:
IMPA
Estrada Dona Castorina, 110
22460-320 Rio de Janeiro, RJ
e-mail: ddic@impa.br - http://www.impa.br
ISBN: 85-244-0240-7

PREFÁCIO

O conteúdo das presentes notas é o de um curso de 12 semanas de introdução à Teoria das Probabilidades. (ou, talvez, mais exatamente de espaços de probabilidades discretos). O pré-requisito para sua leitura é a de um curso de cálculo diferencial e integral e noções elementares de álgebra. É necessário conhecer as propriedades básicas das séries numéricas e de potências. Estas propriedades estão enunciadas no Anexo 1 para facilitar a leitura.

Estas notas foram originalmente escritas para estudantes de matemática, ou estudantes de engenharia e economia com interêsses teóricos. Essa audiência é a que determinou a orientação das exposições. Um primeiro curso de Probabilidades para alunos só com uma preparação reduzida em Matemáticas que pretende ser formal ao mesmo tempo, tem que limitar-se quase essencialmente a espaços de Probabilidades discretos. Êsse é precisamente o tema destas notas.

A Teoria das Probabilidades nasceu do estudo de certos fenômenos empíricos e ainda mantem sua conexão com êles. A apresentação dos temas tratados nestas notas pode ser feita de maneira abstrata esquecendo totalmente as situações empíricas que motivaram seu estudo. Isto roubaria certo interêsse e "emoção" à leitura. Por outro lado fazer uma apresentação totalmente ligada a situações empíricas induz quase sempre um êrro metodológico que consiste na confusão permanente entre o mundo real e o modêlo matemático projetado para estudá-lo. Decidimos então dar ao material um tratamento formal e incluir freqüentemente al gumas descrições intuitivas que têm também a vantagem de facilitar a leitura e a memorização.

Na medida do possível, procuramos dar demonstrações que são válidas em situações mais gerais ou que só necessitam pequenas modificações. Isto indubitàvelmente facilitará posteriormente o trabalho para aqueles que continuem estudando temas mais avançados de Probabilidades. As obras citadas em [2], [6] e [7] podem ser usadas para referência geral.

Esta monografia foi escrita originalmente para ser apresentada no 8º Colóquio Brasileiro de Matemática durante o mês de julho de 1971 em Poços de Caldas, MG. A revisão posterior da mesma, para ser incluida na coleção de Monografias de Matemática do IMPA apontou uma quantidade numerosa de erros de impressão. Por outra parte como acontece quase sempre com uma leitura posterior, a apresentação de certos teoremas, a ênfase sôbre algumas noções particulares, o número e tratamento de alguns exemplos, penso agora que talvez não seja totalmente adequado. Para tentar solucionar êsses problemas e utilizar o material anterior diminuindo assim os custos de impressão, uma secção de Notas aos Capítulos foi incluida. Comentários, alguns novos resultados, e mais exemplos e exercícios são o conteúdo dessas Notas. O leitor é aconselhado a corrigir os erros de cada capítulo antes de começar a leitura do mesmo para evitar perdas de tempo desagradáveis. (Ver Errata). As notas correspondentes a um capítulo devem ser consideradas parte do mesmo.

Desejo agradecer em geral a Comissão Organizadora do 8º Colóquio, e muito particularmente ao Professor Elon Lages Lima por seu cordial convite para participar do mesmo, e posteriormente para incluir essas notas entre as mo nografias do IMPA.

Desejo também agradecer especialmente a João Ismael Damasceno Pinheiro e Annibal Parracho Sant'Anna por sua desinteressada colaboração na correção destas notas.

E, finalmente, quero expressar meu reconhecimento a todos os meus colegas e em geral a todo o pessoal que integra o IMPA, por proporcionar a camaradagem e ambiente adequado para o ensino e a pesquisa, tão necessários para empreender sèriamente qualquer trabalho intelectual.

Rio de Janeiro, Agôsto de 1971

Pedro J. Fernandez

PRÉ-REQUISITOS E NOTAÇÃO BÁSICA

Para a leitura dêste livro, são necessário apenas conhecimentos elementares de Álgebra e Análise. Portanto, é conveniente que o leitor já tenha tido contacto alguma vez com os conceitos de indução completa, conjuntos enumeráveis, relações de equivalência, conjunto cociente, sucessão, operações com funções reais (soma, produto, etc.), números racionais, reais, densidade dos racionais nos reais, supremo e ínfimo de conjuntos de números reais, diferenciação das funções a uma variável real, e integral de Riemann.

Os <u>supremo</u> e <u>ínfimo</u> de dois números reais a e b, serão indicados por $a \vee b$ e $a \wedge b$, respectivamente. Verifica-se fàcilmente que:

$$a \vee b = \frac{1}{2}(a+b + |a-b|), \quad \text{e}$$
$$a \wedge b = \frac{1}{2}(a+b - |a-b|) .$$

Se a e b são números reais, teremos sempre:

$$|a+b| \leq |a|+|b|, \text{ e } \quad ||a|-|b|| \leq |a-b| .$$

Se f e g são funções reais, definimos a função <u>máximo</u> (supremo) de f e g $(f \vee g)$ e <u>mínimo</u> (ínfimo) de f e g $(f \wedge g)$ pelas fórmulas:

$$(f \vee g)(x) = f(x) \vee g(x) \qquad (f \wedge g)(x) = f(x) \wedge g(x)$$

Caso $g = 0$, escrevemos f^+ em lugar de $f \vee 0$, e f^- em lugar de $-(f \wedge 0)$. Resulta então que, $|f| = f^+ + f^-$, e $f = f^+ - f^-$.

Se $\{x_n\}_{n=1,2,\dots}$ é uma sucessão de números reais, definimos o <u>limite superior</u> e o <u>limite inferior</u> de tal sucessão, de acôrdo com as fórmulas

$$\lim_n \sup x_n = \inf_n \sup \{x_k : k \geq n\}$$

$$\lim_n \inf x_n = \sup_n \inf \{x_k : k \geq n\} \ .$$

Por outro lado, sendo $\{f_n\}_{n=1,2,\dots}$ uma sucessão de funções, podemos naturalmente definir $\lim_n \sup f_n$, e $\lim_n \inf f_n$, como as funções:

$$(\lim_n \sup f_n)(x) = \lim_n \sup f_n(x);$$

$$(\lim_n \inf f_n)(x) = \lim_n \inf f_n(x).$$

Se uma série converge de modo absoluto (série absolutamente convergente), será por vêzes indicada com a notação $\sum_{n=1}^{\infty} |x_n| < \infty$. Os teoremas sôbre séries numéricas, a que nos referiremos neste trabalho, encontram-se enunciados no Anexo 1.

Com a finalidade de fixar a notação, bem como tecer algumas considerações adicionais, o Capítulo 1 contém todos os resultados da Teoria dos Conjuntos que venhamos a empregar.

Os quantificadores $\exists$ (existe) e $\forall$ (para todo), serão usados na medida de sua conveniência.

Se $\{x_n\}_{n=1,2,\ldots}$ converge a x escreveremos $x_n \to x$. Se $\{x_n\}_{n=1,2,\ldots}$ é crescente (decrescente) e converge a x escreveremos $x_n \uparrow x$. ($x_n \downarrow x$). O maior inteiro menor ou igual ao número real x será denotado por $[x]$ (parte inteira de x).

Indicaremos com $\mathbb{R}$ e às vêzes com R^1 o conjunto dos números reais.

ÍNDICE

pag.

CAPÍTULO 1

INTRODUÇÃO À TEORIA DOS CONJUNTOS

1.1 - Conjuntos

Neste livro, tôda vez em que usarmos a palavra conjunto, queremos significar um subconjunto de um conjunto fixo, que em geral designaremos por Ω.

Ω é chamado o <u>espaço</u>, e seus elementos, <u>pontos</u>. Se A é um subconjunto de Ω, a notação $\omega \in A$ significa que ω é um elemento de A. O fato de ω não pertencer a A será indicado com a notação $\omega \notin A$. Se A e B são subconjuntos de Ω, a notação $A \subseteq B$ indicará que todo elemento de A também pertence a B. A notação $A \subset B$ indicará que $A \subseteq B$, porém existe algum ponto de B que não é elemento de A.

Com o objetivo de facilitar a notação, usaremos um conjunto que não contenha nenhum elemento, ao qual chamaremos <u>conjunto vazio</u>, e denotaremos por $\emptyset$.

Usaremos a palavra <u>classe</u> ou <u>família</u>, para um conjunto de conjuntos. A classe de todos os subconjuntos de um conjunto A, será indicada por $\mathbb{P}(A)$ (partes de A).

Para indicar um conjunto, também usaremos a notação $\{\omega : p(\omega)\}$, onde $p(\omega)$ é uma proposição concernente a ω, e o conjunto consiste de todos os elementos para os quais $p(\omega)$ é verdadeira. Por exemplo; $\{\omega : \omega = 2k;\ k = 1,2,\ldots\}$ é o conjunto de todos os inteiros positivos pares.

Conjuntos finitos serão por vêzes indicados pela designação de seus elementos. Assim, o conjunto que consiste dos números 0,2 e 4 será indicado $\{0,2,4\}$.

É importante distingüir entre o ponto ω e o conjunto $\{\omega\}$, cujo único elemento é ω.

1.2 - Operações com conjuntos

Seja Γ um certo conjunto de índices, e para cada $\gamma \in \Gamma$, consideremos A_γ, um subconjunto de Ω. Chamaremos união da família $\{A_\gamma : \gamma \in \Gamma\}$, ao conjunto de todos os pontos que pertencem a pelo menos um dos conjuntos A_γ. Êsse conjunto será simbolizado por

$$\bigcup_{\gamma \in \Gamma} A_\gamma = \{\omega : \exists \gamma \in \Gamma, \text{ tal que } \omega \in A_\gamma\}.$$

Se o conjunto Γ fôr enumerável, pode-se dizer que a família $\{A_\gamma : \gamma \in \Gamma\}$ é uma sucessão $\{A_n\}_{n=1,2,\ldots}$.

Indicaremos, neste caso, a união por $\bigcup_{n=1}^{\infty} A_n$.

Se Γ fôr finito, ou seja, $\Gamma = \{1,2,\ldots,k\}$, a união será indicada por $\bigcup\limits_{n=1}^{k} A_n$.

A <u>união</u> de <u>dois</u> conjuntos A e B, será indicada por $A \cup B$.

Se $\{A_\gamma : \gamma \in \Gamma\}$ é uma família de conjuntos, o conjunto de todos os elementos que pertencem a todos os A_γ será chamado <u>interseção da família</u>, e simbolizado por

$$\bigcap_{\gamma\in\Gamma} A_\gamma = \{\omega : \forall \gamma,\ \omega \in A_\gamma\} .$$

Como no caso da <u>união</u> de conjuntos, conforme Γ seja <u>enumerável</u>, <u>finito</u>, ou conste de <u>dois</u> elementos, usaremos as notações:

$$\bigcap_{n=1}^{\infty} A_n\ ,\quad \bigcap_{n=1}^{k} A_n\ ,\quad A \cap B\ .$$

Dois conjuntos são ditos <u>disjuntos</u>, se $A \cap B = \emptyset$. Uma família $\{A_\gamma : \gamma \in \Gamma\}$ se diz <u>disjunta</u>, se para todo par de elementos $\gamma, \gamma' \in \Gamma$; com $\gamma \neq \gamma'$, $A_\gamma \cap A_{\gamma'} = \emptyset$.

Neste caso, usaremos a notação $\sum\limits_{\gamma\in\Gamma} A_\gamma$, em lugar de $\bigcup\limits_{\gamma\in\Gamma} A_\gamma$, para indicar a <u>união</u> da família.

Dado $A \subseteq \Omega$, A^c denotará o conjunto dos pontos que não pertencem a A. Em símbolos, $A^c = \{\omega : \omega \notin A\}$. A^c é chamado <u>complemento de A</u>.

A <u>diferença</u> entre A e B, simbolizada por A-B, é <u>por definição</u>, o conjunto $A-B = A \cap B^c$.

Tal diferença será chamada <u>própria</u>, se $B \subseteq A$.

Ao conjunto $(A-B) + (B-A)$, designado por $A \Delta B$, chamaremos <u>diferença simétrica</u> entre A e B.

Apresentamos a seguir, uma série de importantes relações que serão usados sem menções explícitas, nos capítulos seguintes. Suas demonstrações são deixadas como exercício.

1.2.1. $\forall A,\ A \subseteq \Omega;\quad A \cup \emptyset = A,\quad A \cap \emptyset = \emptyset$

1.2.2. $A \subseteq B$, se e sòmente se, $A \cup B = B$

1.2.3. $A \cup (B \cup C) = (A \cup B) \cup C$

1.2.4. $A \cap (B \cap C) = (A \cap B) \cap C$

1.2.5. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

1.2.6. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

1.2.7. $A \cap \left(\bigcup\limits_{\gamma \in \Gamma} A_\gamma \right) = \bigcup\limits_{\gamma \in \Gamma} (A \cap A_\gamma)$

1.2.8. $A \cup \left(\bigcap\limits_{\gamma \in \Gamma} A_\gamma \right) = \bigcap\limits_{\gamma \in \Gamma} (A \cup A_\gamma)$

1.2.9. $A + A^c = \Omega,\quad A \cap A^c = \emptyset,\quad \emptyset^c = \Omega,\ \Omega^c = \emptyset$

1.2.10. $(A^c)^c = A;\quad A \subseteq B$, se e sòmente se, $B^c \subseteq A^c$

1.2.11. $\left(\bigcup\limits_{\gamma \in \Gamma} A_\gamma \right)^c = \bigcap\limits_{\gamma \in \Gamma} (A_\gamma^c)$

1.2.12. $\left(\bigcap\limits_{\gamma \in \Gamma} A_\gamma \right)^c = \bigcup\limits_{\gamma \in \Gamma} (A_\gamma^c)$

1.2.13. $A \Delta B \subseteq (A \Delta C) \cup (B \Delta C)$

1.3 - Limites e Indicadores

Se $\{A_n\}_{n=1,2,\dots}$ é uma sucessão de conjuntos, chamaremos <u>limite superior da sucessão</u> $\{A_n\}_{n=1,2,\dots}$; ao conjunto de todos os pontos ω que pertençam a A_n, para infinitos valores de n. É fácil ver, então, que:

$$\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i .$$

O conjunto de todos os pontos de Ω que pertencem a todos os A_n, a menos de um número finito de tais A_n, é chamado <u>limite inferior da sucessão</u> $\{A_n\}_{n=1,2,\dots}$, e simbolizado por $\liminf_n A_n$. É fácil perceber, de maneira análoga ao caso anterior, que:

$$\liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i .$$

Uma sucessão $\{A_n\}_{n=1,2,\dots}$ é dita <u>crescente</u> (respectivamente, <u>decrescente</u>), se $A_n \subseteq A_{n+1}$, $n = 1,2,\dots$ (resp. $A_n \supseteq A_{n+1}$, $n = 1,2,\dots$).

Se para uma sucessão $\{A_n\}_{n=1,2,\dots}$, $\limsup_n A_n = \liminf_n A_n$, diremos que tal sucessão tem <u>limite</u>, e então empregaremos apenas a notação $\lim_n A_n$. Indicaremos sucessões crescentes (resp. decrescentes) de conjuntos com a notação $A_n \uparrow$ (resp. $A_n \downarrow$). Escreveremos $A_n \uparrow A$ (resp. $A_n \downarrow A$) se $A_n \uparrow$ (resp. $A_n \downarrow$) e $\bigcup_{n=1}^{\infty} A_n = A$ (resp.

$\bigcap_{n=1}^{\infty} A_n = A)$.

Dado $A \subseteq \Omega$, chamaremos <u>indicador</u> de A, à função I_A definida por:

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

Existe claramente uma correspondência biunívoca (bijeção) entre subconjuntos de Ω e seus <u>indicadores</u> correspondentes. Esta noção nos será extremamente útil nos capítulos seguintes. A seguir, apresentaremos algumas fórmulas, que são deixadas a título de exercício. É importante que o leitor se familiarize com o manejo de <u>indicadores</u>.

1.3.1. $A \subseteq B$, se e sòmente se, $I_A \leq I_B$, $I_{A^c} = 1-I_A$

1.3.2 $I_{A\cap B} = I_A \wedge I_B = I_A I_B$; $I_{\bigcap_{i=1}^{n} A_i} = \prod_{i=1}^{n} I_{A_i}$

1.3.3. $I_{A\cup B} = I_A + I_B - I_A I_B = I_A \vee I_B$

1.3.4. $I_{\left(\bigcup_{i=1}^{n} A_i\right)} = 1 - \prod_{i=1}^{n}(1-I_{A_i}) =$

$$= \sum_{k=1}^{n} (-1)^{k-1} \sum_{1\leq i_1<i_2<\dots<i_n\leq n} \prod_{j=1}^{k} A_{i_j}$$

1.3.5. $I_{(A\Delta B)} = I_A + I_B - 2I_A I_B = |I_A - I_B|$

1.3.6. $I_{(A-B)} = I_A(1-I_B)$

1.3.7. $I_{\limsup_n A_n} = \limsup_n I_{A_n}$; $I_{\liminf_n A_n} = \liminf_n I_{A_n}$.

1.4 - Funçȏes

Sendo X e Y dois conjuntos, denotamos por Y^X o conjunto de tôdas as funçȏes de X em Y.

Se $A \subseteq Y$ e f é uma função, usamos a notação $f^{-1}(A)$, ou às vêzes $[f \in A]$, para indicar o conjunto $\{x: f(x) \in A\}$. Se $B \subseteq X$, $f(B)$ representará o conjunto $\{y: \exists x \in B;\ y = f(x)\}$.

Usaremos, sem menção explícita, as seguintes fórmulas fàcilmente verificáveis:

1.4.1. $f^{-1}(A^c) = [f^{-1}(A)]^c$; $A \subseteq Y$

1.4.2. $f^{-1}(\bigcap_{\alpha\in I} A_\alpha) = \bigcap_{\alpha\in I} f^{-1}(A_\alpha)$ $f^{-1}(\bigcup_{\alpha\in I} A_\alpha) = \bigcup_{\alpha\in I} f^{-1}(A_\alpha)$, onde $\{A_\alpha\}_{\alpha\in I}$ é uma família de subconjuntos de Y.

1.4.3. $f(\bigcup_{\alpha\in I} B_\alpha) = \bigcup_{\alpha\in I} f(B_\alpha)$, $\{B_\alpha\}_{\alpha\in I}$ é uma família de subconjuntos de X.

Se $A \subseteq X$ e $f: X \to Y$, a notação $f|A$ é a função pertinente a Y^A, definida por:

$$f|A(x) = f(x)$$

$f|A$ é chamada a <u>restrição</u> de f a A.

EXERCÍCIOS

1. Provar as fórmulas das seções 2, 3 e 4.

2. Descrever por palavras o conjunto:

$$\lim_n \sup A_n - \lim_n \inf A_n$$

3. Escrever a união e a diferença de dois conjuntos, utilizando as operações Δ e $\cap$.

CAPÍTULO 2

ANÁLISE COMBINATÓRIA

O presente capítulo contém os elementos de Análise Combinatória necessários aos capítulos subseqüentes. O tratamento é formal, embora seja sempre possível uma interpretação concreta, utilizando um modêlo que se valha, por exemplo, de bolinhas em uma urna. Êste modêlo é o seguinte: suponhamos uma caixa contendo n bolinhas, numeradas de 1 a n. Imaginemos que m bolinhas, $1 \leq m \leq n$, sejam extraídas de tal urna, por algum procedimento. O resultado da extração será uma sucessão de números, $x_1, x_2, \ldots, x_m$, compreendidos entre 1 e n. A esta sucessão, chamamos <u>amostra</u>.

O problema consiste em contar quantas sucessões (amostras) podem ser obtidas, ou de outra forma, quantas são as sucessões que possam traduzir o resultado do experimento. Êste número irá depender de como se faça a escolha das bolinhas, e da classificação que faremos das amostras. Aconselhamos ao leitor, fazer uma interpretação, em têrmos dêste modêlo, dos possíveis casos que possam ocorrer. Entretanto, a noção formal que apresentamos tem vantagens na demonstração de certos resultados, bem como na eliminação de ambigüidades na interpretação e solução de problemas concretos.

2.1 - Elementos básicos

Seja A um conjunto finito. Indicaremos por $\#(A)$, o número de elementos de A. Seja I_n o conjunto $\{1,2,\ldots,n\}$. Usaremos sem menção explícita, as seguintes regras: se $\{A_i\}_{i=1,2,\ldots,n}$ é uma família de conjuntos disjuntos, então

$$\#\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \#(A_i) \qquad (1)$$

Se $\prod_{i=1}^{n} A_i$ é o produto cartesiano de $A_1, A_2, \ldots, A_n$; então

$$\#\left(\prod_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \#(A_i)\,. \qquad (2)$$

2.2 - Arranjos com repetição, de n elementos, tomados m a m (ou amostras ordenadas de tamanho m com reposição).

DEFINIÇÃO - Aos elementos de $I_n^{I_m}$, denominamos arranjos com repetição de n elementos, tomados m a

m. Como $I_n^{I_m} = \underbrace{I_n \times I_n \times I_n \times \ldots \times I_n}_{m \text{ vêzes}}$, teremos de acôrdo com a fórmula (2), de 2.1:

$$\# \left(I_n^{I_m}\right) = n^m \; .$$

2.3 - Arranjos de n elementos tomados m a m (ou amostras ordenadas de tamanho m, sem reposição).

DEFINIÇÃO - Seja $m \leq n$, e

$$A_{m,n} = \{f;\ f: I_m \to I_n,\ f \ \underline{\text{biunívoca}}\} \; .$$

Aos elementos de $A_{m,n}$; denominamos arranjos de n elementos tomados m a m. Pode-se mostrar fàcilmente por indução e pela fórmula (2) de 2.1, que:

$$\# A_{m,n} = n(n-1)\ldots(n-m+1) \; .$$

2.4 - Permutações de n elementos

São assim chamados os elementos de $A_{n,n}$ $(m=n)$. Evidentemente, $\# (A_{n,n}) = n!$

2.5 - Combinações de n elementos tomados m a m (ou amostras não ordenadas de tamanho m, sem reposição).

Seja R a seguinte relação de equivalência em

$A_{m,n}$: $f \; R \; g$, se e sòmente se, $f(I_m) = g(I_m)$.

Aos elementos do conjunto cociente $C_{m,n} = A_{m,n}/R$, denominamos combinações de n elementos tomados m a m. Se $f \; R \; g$, é fácil ver que existe $h \in A_{m,m}$ tal que $g = f \circ h$, e recìprocamente. Portanto, o conjunto saturado de f, contém exatamente $m!$ elementos, ou seja, as classes de equivalência em $A_{m,n}$ têm tôdas o mesmo número ($m!$) de elementos. Logo,

$$\# \, (C_{m,n}) = \frac{\# \, (A_{m,n})}{m!} = \frac{n(n-1)\ldots(n-m+1)}{m!} =$$

$$= \frac{n!}{m!(n-m)!} \text{, que é simbolizado por } \binom{n}{m}.$$

Outra forma conveniente de formalizar a noção de combinações, é a de denominar combinações de n elementos tomados m a m, aos elementos do conjunto:

$$C'_{m,n} = \{f : f \in A_{m,n}, \quad f \text{ crescente}\}^{(*)}$$

Claramente, a interseção de $C'_{m,n}$ com cada uma das classes de equivalência determinados por R, consiste exatamente de um só elemento, e portanto:

$$\# \, (C_{m,n}) = \# \, (C'_{m,n})$$

(*) Note-se que por pertencer f a $A_{m,n}$ e ser crescente, é também estritamente crescente.

(*)
2.6 - Combinações com repetição (ou amostras não ordenadas de tamanho m, com reposição).

Chamamos assim aos elementos do conjunto

$$C^R_{m,n} = \{f;\ f:I_m \to I_n,\ \ f\ \ \text{crescente}\}.$$

PROPOSIÇÃO 2.6.1 - $\#\,(C^R_{m,n}) = \binom{m+n-1}{m}$.

<u>Demonstração</u>: Definimos $\phi:\ C^R_{m,n} \to C'_{m,m+n-1}$ da seguinte maneira: $\phi(f)(i) = f(i) + i-1$.

É fàcilmente verificável que ϕ é <u>1-1</u> e <u>sôbre</u> (<u>injetora</u> e <u>sobrejetora</u>), e portanto $\#\,(C^R_{m,n}) = \#\,(C'_{m,m+n-1}) = \binom{m+n-1}{m}$.

2.7 - Permutações com repetição

Antes de procedermos à definição, necessitamos de algumas noções preliminares.

Seja A um conjunto finito, e P_A o conjunto de aplicações <u>1-1</u> e <u>sôbre</u> (<u>injetoras</u> e <u>sobrejetoras</u>), de A em A (as <u>permutações</u> de A). (**)

(*) Note-se que há uma correspondência biunívoca entre $C_{m,n}$ (ou $C'_{m,n}$) e os subconjuntos com m elementos de I_n. Quer dizer: $\#(\{A:A \subseteq I_n,\ \#(A) = m\} = \binom{m}{n}$.

(**) Note-se que $P_{I_n} = A_{n,n}$.

Seja $\{A_i\}_{i=1,2,\ldots,p}$, uma partição de A, e chamemos $k_i = \#(A_i)$ e $n = \#(A) = \sum_{i=1}^{p} k_i$

Seja $H = \{f;\ f \in P_A$ e $\forall i,\ i=1,2,\ldots,p, f^{-1}(A_i) = A_i\}$.

Teremos, então:

PROPOSIÇÃO 2.7.1 - $\#(H) = k_1!\ldots k_p!$

Demonstração: Pode-se verificar fàcilmente que a aplicação $f:H \to P_{A_1} \times P_{A_2} \times\ldots\times P_{A_p}$, definida por $\phi(f) = (f|A_1,\ldots,f|A_p)$ é 1-1 e sôbre, e portanto

$$\#(H) = \#(P_{A_1} \times P_{A_2} \times\ldots\times P_{A_p}) = \prod_{i=1}^{p} \#(P_{A_i}) .$$

Seja $\{A_i\}_{i=1,2,\ldots,p}$ uma partição de I_n. Definimos uma relação de equivalência em P_{I_n} da seguinte maneira:

$$f\ R\ g, \text{ se e sòmente se, } \forall i\ f^{-1}(A_i) = g^{-1}(A_i) .$$

DEFINIÇÃO 2.7.1 - Os elementos de P_{I_n}/R se chamam permutações com repetição de n elementos em relação à partição $\{A_i\}_{i=1,2,\ldots,p}$.

PROPOSIÇÃO 2.7.2 - $\#(P_{I_n}/R) = \dfrac{n!}{\prod_{i=1}^{p} k_i!}$.

Demonstração: Seja, como anteriormente,

$$H = \{g: g \in P_{I_n} \text{ e } g^{-1}(A_i) = A_i,\ i=1,2,\ldots,p\}.$$

É fácil concluir que h pertence ao saturado de f com respeito a R, se e sòmente se $\exists\ g \in H$ tal que $h = g \circ f$. Portanto, tôdas as classes de equivalência têm o mesmo número de elementos $\prod_{i=1}^{p} k_i!$, pela proposição anterior e em consequência:

$$\#\,(P_{I_n}/R) = \frac{n!}{\prod_{i=1}^{p} k_i!}.$$

2.8 - Números combinatórios

Chamaremos números combinatórios aos números $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ definidos para $0 < k < n$. Como $\#(C_{n,n}) = 1$, definimos $0! = 1$ e então resulta que $\binom{n}{n} = 1$, e que a fórmula $\binom{n}{k} = \binom{n}{n-k}$ é válida para todo k, $0 \leq k \leq n$.

Os números combinatórios têm várias propriedades importantes, que passamos a estudar.

PROPOSIÇÃO 2.8.1 - $\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$, $1 \leq k \leq n$.

Demonstração: Consideremos em $C'_{k,n+1}$ os subconjuntos

$$A = \{f : f \in C_{k,n+1}, (n+1) \notin f(I_k)\} \text{ e } A^c.$$

Claramente, $\#(A) = \#(C_{k,n}) = \binom{n}{k}$. Se $f \in A^c$, então $f(k) = n+1$, e a função $\phi : A^c \to C'_{k-1,n}$ definida por $f \to f|I_{k-1}$ é uma bijeção. Resulta portanto que:

$$\#\,(C'_{k,n+1}) = \#(A) + \#(A^c) = \binom{n}{k} + \binom{n}{k-1},$$

que é o resultado desejado.

PROPOSIÇÃO 2.8.2 (Teorema Binomial) - $\forall x,y \in \mathbb{R}$; $\forall n$ inteiro positivo, $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$.

Demonstração: Se $A \subseteq I_n$ é fácil ver que

$$(x+y)^n = \sum_{A \in \mathbb{P}(I_n)} x^{\#(A)} . y^{\#(A^c)} ,$$

e portanto igual a

$$\boxed{\sum_{k=0}^{n} \sum_{\substack{A \in \mathbb{P}(I_n) \\ \#(A)=k}} {}^{\#(A)} y^{\#(A^c)}}$$

Como $\#(\{A : A \in \mathbb{P}(I_n)$ e $\#(A) = k\}) = \binom{n}{k}$, esta última expressão iguala $\sum_{k=0}^{n} \binom{n}{k} x^k \; y^{n-k}$, que é o resultado esperado.

PROPOSIÇÃO 2.8.3 - Se $\#(\Omega) = n$, então $\#(\mathbb{P}(\Omega)) = 2^n$.

Demonstração: Seja $A_k = \{A : A \subseteq \Omega,\ \#(A) = k\}$, $k=0,1,\ldots,n$, então, $\#(A_k) = \binom{n}{k}$ e teremos:

$$\#(\mathbb{P}(\Omega)) = \sum_{k=0}^{n} \#(A_k) = \sum_{k=0}^{n} \binom{n}{k} = (1+1)^n = 2^n .$$

2.9 - Aplicações

PROPOSIÇÃO 2.9.1 - O número de pares inteiros (i,j) tais que $1 \leq i < j \leq n$, é $\frac{n(n-1)}{n}$.

Demonstração: Tal número é igual a $\#(C_{2,n})$, e portanto igual a $\binom{n}{2} = \frac{n!}{2!(n-2)!} = \frac{n(n-1)}{n}$.

PROPOSIÇÃO 2.9.2 - O número de pares de inteiros (i,j) tais que $1 \leq i \leq j \leq n$ é $\frac{n(n+1)}{2}$.

Demonstração: O conjunto descrito é precisamente $C^R_{2,n}$, e portanto sua cardinalidade é $\#(C^R_{2,n}) = \frac{n(n+1)}{2}$.

PROPOSIÇÃO 2.9.3 - $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$.

Demonstração: Seja $A_k = \{(k,i): 1 \leq i \leq k\}$ e $A = \{(i,j): 1 \leq i \leq j \leq n\}$.

Então os A_k são disjuntos e $\sum_{k=1}^{n} A_k = A$.

Teremos, $\sum_{k=1}^{n} k = \sum_{k=1}^{n} \#(A_k) = \#(\sum_{k=1}^{n} A_k) = \#(A) = \frac{n(n+1)}{2}$, pela Proposição 2.9.2.

PROPOSIÇÃO 2.9.4 - O número de aplicações $f: I_m \to I_n + \{0\}$, tais que $\sum_{i=1}^{m} f(i) \leq n$ (resp. $\sum_{i=1}^{m} f(i) = n$) é dado por:

$$\binom{n+m}{n} \quad (\text{resp. } \binom{n+m-1}{m-1}) .$$

Demonstração: Seja $A(m,n)$ (resp. $B(m,n)$) o conjunto de aplicações $f: I_m \to I_n + \{0\}$, tais que $\sum_{i=1}^{m} f(i) \leq n$ (resp. $\sum_{i=1}^{m} f(i) = n$). Definamos $\phi: A(m,n) \to C^R_{m,n+1}$ da seguinte forma $\phi(f)(i) = 1 + \sum_{j \leq i} f(j)$, $i = 1,2,\ldots,m$. Claramente, $\phi(f)$ é um elemento de $C^R_{m,n+1}$. Se $f \neq g$, $\exists$ um primeiro i_o tal que $f(i_o) \neq$

$\neq g(i_o)$, e portanto:

$$\phi(f)(i_o) = 1 + \sum_{j\leq i_o} f(j) = 1 + \sum_{j<i_o} g(j) + f(i_o) \neq$$

$$\neq 1 + \sum_{j<i_o} g(j) + g(i_o) = \phi(g)(i_o).$$

(Caso $i_o = 1$, $\sum_{j<i_o} g(j)$ será tomado por 0). Portanto, ϕ é 1-1 (injetiva, biunívoca). Agora, dada $g \in C^R_{m,n+1}$, definimos

$$f(i) = \begin{cases} g(i)-1, & i = 1 \\ g(i)-g(i-1), & 1 < i \leq n \end{cases}$$

Então, $f(i) \geq 0$, para todo i, $\sum_{i=1}^{m} f(i) = (g(m)-1) \leq n$, e

$$\phi(f)(i) = \begin{cases} 1 + g(1)-1 = g(i), & i=1 \\ 1+(g(1)-1)+\sum_{j=2}^{i} g(j)-g(j-1) = g(i) & i > 1, \end{cases}$$

o que mostra que ϕ é também <u>sôbre</u>. Em outras palavras, ϕ é uma bijeção entre $A(m,n)$ e $C^R_{m,n+1}$, e portanto, $\#(A(m,n)) = \binom{m+n+1-1}{m} = \binom{m+n}{m}$.

Para calcular $\#(B(m,n))$, basta observar que a função $\xi : B(m,n) \to A((m-1),n)$ definida por $\xi(f) = f|I_{m-1}$ para $f \in B(m,n)$ estabelece uma <u>correspondência biunívoca</u> entre $B(m,n)$ e $A((m-1),n)$, e portanto:

$$\#(B(m,n)) = \binom{m+n-1}{m-1} .$$

(*) DEFINIÇÃO 2.9.1 - Uma família $A_1, A_2, \ldots, A_n$ de subconjuntos de um conjunto Ω é chamada um <u>sistema</u> de Sperner, se $\forall i,j \quad i \neq j \quad A_i \not\subset A_j$.

PROPOSIÇÃO 2.9.5 (<u>Lema de Sperner</u>) - Se $\#(\Omega) = n$ e $\{A_i\}_{i=1,\ldots,N}$ é um <u>sistema de Sperner</u>, então $N \leq \binom{n}{[\frac{n}{2}]}$.

<u>Demonstração</u>: Consideremos o conjunto de cadeias maximais em Ω. Quer dizer, famílias de conjuntos $\{B_j\}_{j=1,2,\ldots,n}$ tais que $B_j \subseteq B_{j+1}$, $j = 1,2,\ldots,n-1$, e $\#(B_j) = j$. É fácil ver que o número de <u>cadeias maximais</u> é $n!$. Se C é um conjunto não-vazio contido em Ω, o número de cadeias maximais dos quais A é membro, é $[\#(A)]![n-\#(A)]!$

Outro fato fácil de comprovar é que o valor mínimo assumido pela função $r!(n-r)!$ definida para $r=0,1,\ldots,n$ é $[\frac{n}{2}]!(n-[\frac{n}{2}])!$, onde $[x]$ indica a parte inteira de x.

Então, dado A_i, existirão $[\#(A_i)]![n-\#(A_i)]!$ cadeias maximais às quais A_i pertence, e portanto:

$$n! \geq \sum_{i=1}^{N} (\#(A_i))!(n-\#(A_i))! \geq N[\tfrac{n}{2}]!(n-[\tfrac{n}{2}])!$$

Isto é, $$N \leq \frac{n!}{[\frac{n}{2}]!(n-[\frac{n}{2}])!} = \binom{n}{[\frac{n}{2}]} .$$

(*) A Proposição 2.9.5 não será utilizada no futuro, neste livro.

Esta demonstração curta e elegante do lema de Sperner, é devida a Lubell [1].

2.10 - Fórmula de Stirling

$n! \cong \sqrt{2\pi}\ u^{n+\frac{1}{2}}\ e^{-n}$ no sentido que o quociente dos dois membros converge a 1.

Vamos aproximar $\log(n!)$. É natural interpretar $\log(n!)$ como a integral da função que vale $\log k$ entre $k - \frac{1}{2}$ e $k + \frac{1}{2}$ (ver figura 2.10.1).

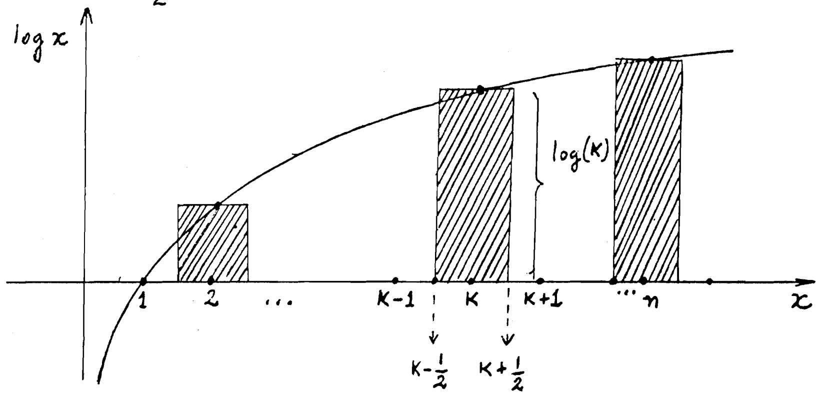


Figura 2.10.1

Chamemos $L(x)$ a esta função escada e $\ell(x) = \log(x)$. Temos que $\log(n!) = \int_{1/2}^{u+1/2} L(x)dx$. A idéia é aproximar $L(x)$ por $\ell(x)$ e estimar a diferença.

Para trabalhar com integrandos positivos poremos

$$a_k = \int_{k-1/2}^{k} L(x)dx - \int_{k-1/2}^{k} \log x\ dx = \int_{k-1/2}^{k} [\log k - \log x]dx =$$

$$= \int_{k-1/2}^{k} \log(\tfrac{k}{x})dx.$$

$$b_k = \int_k^{k+1/2} \log x - \int_k^{k+1/2} L(x)dx = \int_k^{k+1/2} [\log x - \log k]dx =$$

$$= \int_k^{k+1/2} \log(\tfrac{x}{k})dx.$$

Definamos $I(x) = \int_b^x \log y \, dy = x \log x - x$.

Temos

$$\sum_{k=1}^{n} a_k - b_k = \int_{1/2}^{n+1/2} L(x)dx - \int_{1/2}^{n+1/2} \log(x)dx .$$

Isto é

$$a_1 - b_1 + a_2 - b_2 + \ldots + a_n - [I(n+\tfrac{1}{2}) - I(n)] + \tfrac{1}{2}\log n =$$

$$= \log(n!) - [I(n+\tfrac{1}{2}) - I(\tfrac{1}{2})].$$

$$a_1 - b_1 + \ldots + a_n = \log(n!) - \tfrac{1}{2}\log n + I(\tfrac{1}{2}) - I(n)$$

$$= \log(n!) + n - (n+\tfrac{1}{2})\log n + I(\tfrac{1}{2})$$

É fácil ver que

$$a_k = \int_0^{1/2} \log \frac{1}{1-(\frac{Y}{k})} dy, \quad b_k = \int_0^{1/2} \log(1+\tfrac{y}{k})dy .$$

Destas expressões para a_k e b_k se deduz imediatamente que $a_k > b_k > a_{k+1}$ e que $a_k \to 0$. Isto pode ser "intuido" pela figura.

Portanto, $a_1 - b_1 + a_2 - b_2 + \ldots + a_n$ converge e temos

$$\sum_{k=1}^{\infty}(a_k-b_k) = \lim_{n\to\infty}\,[\log(n!) + n-(n+\tfrac{1}{2})\log n] + I(\tfrac{1}{2})$$ isto é:

$$\lim_{n\to\infty}\,[\log(n!) + n-(n+\tfrac{1}{2})\log n] = \sum_{k=1}^{\infty}(a_k-b_k) - I(\tfrac{1}{2}) = c\ ,$$

que é a fórmula de Stirling com e^c em lugar de $\sqrt{2\pi}$.

O problema agora é avaliar e^c. Isto vai ser feito mais adiante. (Ver Teorema 8.2.2).

EXERCÍCIOS

1. Provar que a soma dos u primeiros números ímpares é n^2.

2. Provar que a soma dos n primeiros números pares é $n(n+1)$.

3. Seja Ω um conjunto finito contendo n elementos e seja $g(n)$ o número de partições de conjuntos disjuntos de Ω. (Isto é familias $\{A_i\}_{i=1,\ldots,k}$ tal que $A_i \neq \emptyset$ e $\sum_{i=1}^{n} A_i = \Omega$).

   a) Demonstre que $g(u) = \sum_{k=0}^{n-1} \binom{n-1}{k} g(k)$

      $(g(0) = 1$ por definição).

   b) Demonstre que $g(n) = e^{-1}\sum_{k=0}^{\infty} \dfrac{k^n}{k!}$.

4. Para n inteiro positivo ≥ 2

   a) $1 - \binom{n}{1} + \binom{n}{2}\ \ldots\ = 0$

   b) $\binom{n}{1} + 2\binom{n}{2} + 3\binom{n}{3} + \ldots. = n\,2^{n-1}$

   c) $\binom{n}{1} - 2\binom{n}{2} + 3\binom{n}{3} - \ldots. = 0$

d) $2\binom{n}{2} + 3.2\binom{n}{3} + 4.3\binom{n}{4} + \ldots. = n(n-1)2^{n-2}$

5. a) $\sum_{v=0}^{k} \binom{n}{v}\binom{n-v}{k-v} t^v = \binom{n}{k}(1+t)^k$

b) $\binom{n}{o}\binom{n}{k} - \binom{n}{1}\binom{n-1}{k-1} + \binom{n}{2}\binom{n-2}{k-1} + \binom{n}{2}\binom{n-2}{k-2} \ldots +$

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

6. Para $a \geq 1$ $n < a$, inteiros positivos

$\sum_{v=0}^{n} (-1)^v \binom{a}{v} = (-1)^n \binom{a-1}{u}$ (usar Proposição 2.3.1)

7. $\sum_{v=m}^{n} \binom{v}{m} = \binom{n+1}{m+1}$ (usar Proposição 2.3.1)

8. a, b, ℓ, inteiros positivos

$$\binom{a+b}{\ell} = \sum_{\substack{k+j=\ell \\ 0\leq k\leq a \\ 0\leq j\leq b}} \binom{a}{k}\binom{b}{j}$$

(Sugestão: $(1+t)^a(1+t)^b = (1+t)^{a+b}$)

9. $\binom{n}{0}^2 + \binom{n}{1}^2 + \ldots. + \binom{n}{n}^2 = \binom{2n}{n}$

10. Calcular o número de monômios $X_1^{\alpha_1} \ldots X_m^{\alpha_m}$ em m indeterminadas de grau $\leq n$ e igual a n.

11. Provar que

$\sum_{k=j}^{i} \binom{k}{j}\binom{i}{k} (-1)^{j+k} = 0 \quad i > j$

(Sugestão: $\binom{i}{k}\binom{k}{j} = \binom{i}{j}\binom{i-j}{k-j}$)

12. (Êste exercício requer a Proposição 2.9.5).

DEFINIÇÃO - Sejam $A_1, A_2, \ldots, A_n$ subconjuntos não vazios de Ω. $A_1, A_2, \ldots, A_n$ são chamados qualitativamente independentes a pares (q.i.p.) se e só se $B_i \cap B_j \neq \emptyset$ para todo (i,j) $i \neq j$ e $B_i = \begin{cases} A_i \\ A_i^c \end{cases}$.

Seja $\Omega = \{1,2,\ldots,2n\}$. Provar que o número máximo de subconjuntos q.i.p. é igual a $\frac{1}{2}\binom{2n}{n}$ para $n \geq 2$. (Sugestão: Se $A_1, A_2, \ldots, A_s$ são q.i.p. então $A_1, A_2, \ldots, A_s, A_1^c, \ldots, A_s^c$ é um sistema de Sperner, o que dá uma desigualdade. Depois construir um sistema q.i.p. com $\frac{1}{2}\binom{2n}{n}$ elementos. (Tomar todos os subconjuntos com n elementos que não contém o número 1 por exemplo)).

Mais exercícios sôbre problemas combinatórios podem ser encontrados no excelente [2]. Ver também [3].

CAPÍTULO 3

MODELOS PROBABILÍSTICOS

3.1 - Experimentos Determinísticos e Aleatórios

Dirèmos em geral que um experimento é determinístico quando, repetido em condições semelhantes, conduz a resultados essencialmente idênticos. Os experimentos que, repetidos sob as mesmas condições dão resultados em geral distintos serão chamados experimentos aleatórios. A Teoria das Probabilidades é o ramo da Matemática que cria, desenvolve e, em geral, investiga modelos que podem ser utilizados para estudar experimentos aleatórios. É talvez conveniente, para manter o interêsse no assunto e ajudar a lembrar certos resultados, usar uma linguagem pitórica que provém de certos experimentos aleatórios particulares. Nós materemos êsse costume, mas o leitor deverá ter sempre presente que tôdas as afirmações que faremos serão teoremas dentro do modêlo e não necessàriamente descreverão exatamente a situação no mundo real. A seguir, relacionamos uma série de experimentos aleatórios aos quais faremos referência mais ou menos constante nos capítulos seguintes.

E_1. Jogue uma moeda e observe o resultado obtido.

E_2. Jogue um dado e observe o número mostrado na face de cima.

E_3. Jogue uma moeda dez vêzes e observe a seqüência de caras e coroas obtida.

E_4. Jogue uma moeda dez vêzes e observe o número de caras obtido.

E_5. Jogue uma moeda até obter a primeira cara e observe a seqüência obtida.

E_6. Jogue uma moeda até obter a décima cara e observe a seqüência obtida.

E_7. Jorgue uma moeda até obter uma cara e observe o número de lançamentos que foram necessários.

E_8. De uma relação completa dos habitantes de uma cidade, se escolhe um ao acaso e se registra a sua altura.

E_9. De uma substância radioativa, se mede, com um contador, o número de partículas emitidas em um minuto.

E_{10}. Usando um termógrafo, se registra a temperatura ambiente em graus centígrados, continuamente, por um período de dez horas. Observa-se o gráfico obtido.

3.2 - O Espaço Amostral

Dado um experimento, diremos que um conjunto Ω é um <u>espaço amostral para êste experimento</u> quando a cada possível resultado do mesmo possamos associar um ponto do conjunto Ω de modo que a resultados distintos do experimento correspondam pontos distintos em Ω. Decorre claramente desta definição que é possível ter vários espaços amostrais para o mesmo experimento. É, porém, conveniente tomar Ω o menor possível. Trata-se, quanto possível, de escolher Ω de modo que haja uma correspondência biunívoca entre êste conjunto e o conjunto dos resultados do experimento.

Vamos agora descrever um espaço amostral para cada um dos experimentos em 3.1. Ω_i será o espaço amostral do experimento E_i.

$\Omega_1 = \{H,T\}$. Às vêzes, usaremos também o conjunto $\{0,1\}$.

$\Omega_2 = \{1,2,3,4,5,6\}$

$\Omega_3 = \{(x_1,x_2,\ldots,x_{10}) \mid x_i=0 \text{ ou } 1 \quad i=1,2,\ldots,10\}$

$\Omega_4 = \{0,1,2,3,\ldots,10\}$

$\Omega_5 = \{(x_1,x_2,\ldots,x_n) \mid x_i=0 \quad i=1,\ldots,n-1 \quad x_n=1, \; n=1,2,\ldots\}$

$\Omega_6 = \{(x_1,x_2,\ldots,x_n) \mid x_n=1, \; \sum_{i=1}^{n} x_i = 10 \quad n=10,11,\ldots\}$

$\Omega_7 = \{1,2,3,\ldots\}$

$\Omega_8 = \{x \mid 0 \leq x \leq 3\}$

$\Omega_9 = \{0,1,2,\ldots\}$

$\Omega_{10} = \{f \mid f:[0,10] \to [-100,+\infty), \quad f \text{ contínua}\}$

É claro que é perfeitamente razoável usar, em lugar de Ω_8, $\Omega'_8 = \{x: 0,10 \leq x \leq 2,80\}$ e, em lugar de Ω_{10}, $\Omega'_{10} = \{f: f:[0,10] \to [-100,+100],\ f$ contínua e derivável$\}$.

Êstes exemplos mostram claramente a arbitrariedade da escolha do espaço amostral. Êste é um problema que não nos afeta diretamente, em um curso de teoria de probabilidades, mas é de grande importância quando se usam modelos probabilísticos para procurar dar soluções a certas situações concretas.

Nesta obra, espaços como os descritos em Ω_8 e Ω_{10} não serão tratados, porque seu estudo requer o uso de ferram‹ tas matemáticas mais avançadas. Todos os espaços amostrais que estudaremos serão finitos ou enumeráveis.

Eventos - Dado um espaço amostral Ω, todo subconjunto de Ω será chamado evento. Os subconjuntos que possuam um único elemento serão chamados eventos-elementares.

Nos seguintes exemplos, A_i é um evento de Ω_i.

$A_2 = \{1,3,5\}$, isto é, um número ímpar ocorre.

$A_3 = \{(x_1,x_2,\ldots,x_{10}): \sum_{i=1}^{10} x_i \geq 4\}$, isto é, mais de quatro caras ocorrem.

$A_9 = \{0\}$, isto é, nenhuma partícula foi emitida no intervalo considerado.

Como os eventos são subconjuntos de Ω, tôdas as operações elementares com conjuntos se podem efetuar também com eventos. É possível descrever com mais colorido as operações com eventos, do ponto de vista da teoria de probabilidades. Por exemplo, diz-se que o evento $\bigcup_{i=1}^{\infty} A_i$ ocorre se e só se algum dos eventos A_i ocorre, que o evento A^c ocorre se A não ocorre, etc. Ω é também chamado o <u>evento certo</u> e $\emptyset$ o <u>evento impossível</u>. Quando dois eventos são disjuntos diz-se também que são <u>mùtuamente exclusivos</u>.

3.3 - Espaços de Probabilidades

Até o momento só descrevemos alguns dos aspectos que compõem um experimento. Nosso problema agora é o de descrever um experimento aleatório matemàticamente. Em outras palavras, o que queremos é construir um modêlo que sirva para tratar experimentos aleatórios. Consideremos um experimento aleatório com espaço amostral Ω e seja $A \subseteq \Omega$. Se repetimos o experimento n vêzes, seja $F(A)$ o número de vêzes que A ocorre e $f(A) = F(A)/n$ sua freqüência relativa. Esta freqüência relativa tem várias propriedades: $0 \leq f(A) \leq 1$; se $A = \emptyset$, $f(A) = 0$; se $A = \Omega$, $f(A) = 1$; se $A_1 \cap A_2 = \emptyset$, então $f(A_1) + f(A_2) = f(A_1+A_2)$. O que queremos agora é atribuir um número a cada evento A, o qual avaliará quão verossímil será a ocorrência de A quando o experimento fôr realizado. Se também queremos que nosso modêlo tenha as características que correspondem às freqüências relativas, é natural exigir que o número que associamos a A, que vamos denotar com $P(A)$ e que chamaremos probabilidade de A, tenha as seguintes propriedades:

(3.3.1) $0 \leq P(A) \leq 1$; $P(\emptyset) = 0$ e $P(\Omega) = 1$;

$P(A_1+A_2) = P(A_1) + P(A_2)$ se A_1 e A_2 são mùtuamente exclusivos.

Por razões de conveniência matemática, vamos impor uma condição mais restritiva à nossa definição de probabilidade.

DEFINIÇÃO 3.3.1 - $P:\mathbb{P}(\Omega) \to [0,1]$ é uma probabilidade sôbre Ω se e só se

i) $P(\Omega) = 1$ e $P(\emptyset) = 0$

ii) Se $\{A_i\}_{i=1,\ldots}$ é uma família de conjuntos disjuntos, $P(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

DEFINIÇÃO 3.3.2 - Ao par (Ω,P) se dá o nome de espaço de probabilidades.

PROPOSIÇÃO 3.3.1 - Se $\{A_i\}_{i=1,\ldots,n}$ é uma família finita disjunta, então

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) \, .$$

<u>Demonstração</u>: Façamos

$$B_i = \begin{cases} A_i & \text{se} \quad i=1,\ldots,n \\ \emptyset & \text{se} \quad i > n \end{cases} .$$

Então

$$P\left(\sum_{i=1}^{n} A_i\right) = P\left(\sum_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i) = \sum_{i=1}^{n} P(B_i) = \sum_{i=1}^{n} P(A_i)$$

A Proposição 3.3.1 nos diz que tôda probabilidade satisfaz as condições (3.3.1). Uma função definida sôbre $\mathbb{P}(\Omega)$ que satisfaça (3.3.1) será chamada uma <u>probabilidade finitamente aditiva</u>. <u>Existem probabilidades finitamente aditivas que não são probabilidades</u>. (Ver exercício 2).

Vamos deduzir agora algumas conseqüências mais ou menos imediatas da definição de probabilidade.

PROPOSIÇÃO 3.3.2 - Se $A \subseteq B$ então $P(A) \leq P(B)$.

<u>Demonstração</u>: $B = A + (B-A)$. Portanto $P(B) = P(A)+P(B-A)$. Como $P(B-A) \geq 0$ resulta que $P(B) \geq P(A)$.

PROPOSIÇÃO 3.3.3 - $P(A^c) = 1 - P(A)$.

<u>Demonstração</u>: $\Omega = A + A^c$. Portanto $1 = P(\Omega) = P(A) + P(A^c)$ ou seja $1 - P(A) = P(A^c)$.

PROPOSIÇÃO 3.3.4 - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

<u>Demonstração</u>: $A \cup B = (A-B) + (B-A) + (A \cap B)$. Portanto

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

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

PROPOSIÇÃO 3.3.5 - $P(A\cup B\cup C) = P(A) + P(B) + P(C) -$

$- P(A\cap B) - P(A\cap C) - P(B\cap C) + P(A\cap B\cap C).$

Demonstração: $P(A\cup B\cup C) = P(A\cup B) + P(C) - P[(A\cup B) \cap C] =$

$= P(A) + P(B) + P(C) - P(A\cap B) -$

$- P[(A\cap C) \cup (B\cap C)] = P(A) + P(B) + P(C) - P(A\cap B) - P(A\cap C) -$

$- P(B\cap C) + P(A\cap B\cap C).$

Tôdas as igualdades se obtêm por aplicação da Proposição 3.3.4 e de propriedade distributiva.

Esta proposição tem uma generalização que agora enunciamos e cuja demonstração veremos mais adiante. (Ver Exemplo 6.2.4). (É possível fazer uma demonstração direta por indução).

PROPOSIÇÃO 3.3.6 - $P(\bigcup_{i=1}^{n} A_i) =$

$$= \sum_{k=1}^{n} (-1)^{k-1} \sum_{1\le i<\ldots<i_k\le n} P(A_{i_1} \cap\ldots\cap A_{i_k}) .$$

DEFINIÇÃO 3.3.3 - Uma probabilidade finitamente aditiva diz-se contínua por cima no vazio se e só se, para tôda sucessão decrescente de conjuntos $\{A_i\}_{i=1,\ldots}$ tal que $\bigcap_{i=1}^{\infty} A_i = \emptyset$, $\lim_{i\to\infty} P(A_i) = 0.$

PROPOSIÇÃO 3.3.7 - Uma probabilidade finitamente aditiva é contínua por cima no vazio se e só se é uma probabilidade.

<u>Demonstração</u>: Seja P uma probabilidade finitamente aditiva e contínua por cima no vazio e $\{A_i\}_{i=1,2,\ldots}$ uma sucessão de conjuntos disjuntos.

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

A primeira e segunda igualdades seguem de ser finitamente aditiva e propriedades de limites; a terceira, da continuidade por cima no vazio; a última por definição.

Recìprocamente, se P é probabilidade e $\{A_i\}$ é decrescente para o vazio, temos, para todo $n \geq 1$,

$$A_1 = (A_1-A_2) + (A_2-A_3) + \ldots + (A_{n-1} - A_n) + A_n$$

donde $P(A_1) = \sum_{i=1}^{n-1} P(A_i-A_{i+1}) + P(A_n)$.

Como P é probabilidade e a sucessão $\{A_i\}$ é decrescente para o vazio,

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

Portanto,

$$P(A_1) = \lim_{n\to\infty} \sum_{i=1}^{n-1} P(A_i - A_{i+1}) + \lim_{n\to\infty} P(A_n) = P(A_1) + \lim_{n\to\infty} P(A_n).$$

Isto é $\lim_{n\to\infty} P(A_n) = 0.$

3.4 - Construção de Probabilidades

Se (Ω, P) é um espaço de probabilidades e $\Omega = \{\omega_1, \omega_2, \ldots\}$, temos, devido à Definição 3.3.1 ii) que $1 = \sum_{i=1}^{\infty} P(\{\omega_i\})$. Se definimos $p_i = P(\{\omega_i\})$, temos então que a sucessão $\{p_i\}_{i=1,2,\ldots}$ satisfaz as seguintes propriedades:

$$p_i \geq 0, \quad i=1,2,\ldots \quad \text{e} \quad \sum_{i=1}^{\infty} p_i = 1 \tag{3.4.1}$$

A sucessão $\{p_i\}_{i=1,\ldots}$ determina P porque $P(A) = \sum_{i=1}^{\infty} I_A(\omega_i) P(\{\omega_i\}) = \sum_{i=1}^{\infty} I_A(\omega_i) p_i$ e, portanto, se $P \not\equiv P'$, as correspondentes sucessões $\{p_i\}$ e $\{p_i'\}$ são distintas. A pergunta natural agora é se, dada uma sucessão que satisfaça (3.4.1), $\exists$ uma probabilidade P tal que $P(\{\omega_i\}) = p_i$, $i=1,2,\ldots$. A resposta é afirmativa e está contida na seguinte proposição.

PROPOSIÇÃO 3.4.1 - Existe uma única probabilidade P tal que $P(\{\omega_i\}) = p_i$, $i=1,2,\ldots$

<u>Demonstração</u>: Tal probabilidade, caso existe, é única, pelo visto acima e deve ser tal que

$$P(A) = \sum_{i=1}^{\infty} I_A(\omega_i)p_i.$$

Definamos P então por esta fórmula e verifiquemos que é uma probabilidade. Isto claramente demonstrará a proposição.

Obviamente, $0 \leq P(A) \leq 1$, $P(\Omega) = 1$, $P(\emptyset) = 0$ e P é finitamente aditiva.

Seja agora $\{A_i\}_{i=1,\ldots}$ uma sucessão de conjuntos decrescente para o vazio. Se $\#(A_1)$ é finito, é fácil ver que existe um i_o tal que, se $i \geq i_o$, $A_i = \emptyset$ e portanto $P(A_i) = 0$ para $i \geq i_o$ e $\lim_{i\to\infty} P(A_i) = 0$. Se $\#(A_1)$ não é finito, seja $A_1 = \{\omega_{i_1}, \omega_{i_2}, \ldots\}$. Dado $\epsilon > 0$, como $\sum_{j=1}^{\infty} p_{ij}$ converge a $P(A_1)$, determinemos k de tal maneira que $\sum_{j=k}^{\infty} p_{ij} < \epsilon$. Como a sucessão $\{A_i\}_{i=1,\ldots}$ é decrescente para o vazio existe i_o tal que se $i \geq i_o$, $A_i \subseteq \{\omega_{i_k}, \omega_{i_{k+1}}, \ldots\}$ e portanto $P(A_i) \leq \sum_{j=k}^{\infty} p_{ij} < \epsilon$ para $i \geq i_o$. Isto é, $\lim_{i\to\infty} P(A_i) = 0$ e, portanto, pela Proposição 3.3.7, P é uma probabilidade.

Êste teorema nos vai permitir simplificar a apresentação de espaços de probabilidades concretos. Sempre que necessitemos introduzir um espaço (Ω, P), indicaremos simplesmente qual é Ω e logo daremos uma sucessão $\{p_n\}_{n=1,2,\ldots}$ de números tais que $p_n \geq 0$, $n=1,2,\ldots$ e

$\sum_{n=1}^{\infty} p_n = 1$. As probabilidades a que nos referiremos então serão as que correspondam à probabilidade P introduzida no teorema anterior.

Se $\{x_1, x_2, \ldots\}$ é um conjunto de números reais e $\{p_n\}_{n=1,\ldots}$ é uma sucessão que satisfaz (3.4.1), podemos definir, para todo conjunto A de números reais, $P(A) = \sum_{i=1}^{\infty} I_A(x_i) p_i$ e isto é uma probabilidade sôbre a família de todos os subconjuntos da reta real. As probabilidades construídas desta maneira serão chamadas <u>distribuições</u>.

Da infinita quantidade de probabilidades que se pode associar a um espaço amostral Ω é importante estudar, em particular, aquelas escolhas que dêm lugar a espaços de probabilidades que possam servir como modêlo para experimentos aleatórios concretos. A escolha de P nestes casos se faz fundamentalmente de duas maneiras: por considerações de simetria ou baseando-se nas freqüências relativas obtidas a partir de larga série de repetições do experimento. Para ilustrar esta situação, consideremos o espaço amostral Ω_1. Por considerações de simetria é natural tomar $P(\{H\}) = p_1 = p_2 = P(\{T\}) = \frac{1}{2}$. Êste seria o espaço de probabilidades correspondente a uma moeda equilibrada. Se, efetuada uma longa série de lançamentos de uma moeda, a freqüência relativa de caras fôra de 0,4, o espa ço das probabilidades anterior não serviria para descrever o comportamento desta moeda. Um espaço de probabilidades mais adequado seria o que se obt'm tomando $r_1 = 0.4$ e $p_2 = 0.6$. Em geral, se $0 \le p \le 1$, se constrói um espaço de probabilidades fixando $p_1 = p$ e $p_2 = 1-p$. Esta família de espaços descreveria o comportamento de tôdas as moedas possíveis, de massa bem ou màl distribuída.

Para o experimento E_2, uma escolha natural de P seria $p_1 = p_2 = \dots = p_6 = \frac{1}{6}$. Como no caso anterior, uma èscolha diferente desta corresponderia ao caso de um dado viciado. Para E_9, uma probabilidade que tem importantes aplicações se obtém fixando $p_n = \frac{e^{-\lambda}\lambda^n}{n!}$, $\lambda > 0$, $n = 0,1,\dots$.

É claro que $p_n > 0$ e $\sum_{n=0}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} =$ $= e^{-\lambda}e^{\lambda} = e^0 = 1$ e em virtude da Proposição 3.3.7 esta sucessão determina então uma probabilidade P_λ sôbre Ω_9. Observe-se que para cada valor de $\lambda > 0$ temos uma probabilidade diferente sôbre Ω_9, e, por isto, a notação P_λ. Esta probabilidade é chamada de Poisson. (Alguns valores de P_λ estão contidos na Tabela 2).

Seja agora $\Omega = \{\omega_1,\dots,\omega_n\}$. Entre tôdas as probabilidades que é possível definir sôbre $\mathbb{P}(\Omega)$ há uma que merece particular atenção. Esta probabilidade é a determi

minada fixando $P(\{\omega_i\}) = p_i = \frac{1}{n}$, $i=1,2,\dots,n$. Se $A \subseteq \Omega$,
$P(A) = \sum_{i=1}^{n} I_A(\omega_i)p_i = \frac{1}{n}\sum_{i=1}^{n} I_A(\omega_i) = \frac{\#(A)}{n} = \frac{\#(A)}{\#(\Omega)}$.

Quer dizer, a probabilidade de um conjunto A é o número de elementos de A sôbre o total de elementos em Ω. Esta expressão é também descrita dizendo que a probabilidade é o cociente do "número de casos favoráveis dividido pelo número de casos possíveis". É para êste particular espaço de probabilidades principalmente que desenvolvemos as técnicas do Capítulo 2.

Em seguida vamos apresentar vários exemplos que ilustram o uso das técnicas do Capítulo 2.

EXEMPLO 3.4.1 - Consideremos uma urna contendo n bolas das quais $n_1 \geq 1$ são brancas e $n_2 \geq 1$ são pretas e $n = n_1+n_2$. Escolhe-se, ao acaso, uma amostra de r, $1 \leq r \leq n$ e se quer saber qual é a probabilidade de que exatamente k das bolas na nossa amostra sejam brancas. Suponhamos que as bolas tenham sido numeradas de 1 a n de tal forma que as n_1 primeiras são brancas e as n_2 últimas são pretas. Como a cada amostra podemos fazer corresponder uma função estritamente crescente de I_r em I_n é natural tomar como espaço amostral $C'_{r,n}$ e como probabilidade a que faz todos os eventos dêste conjunto igualmente prováveis. Temos então que calcular o número de funções $f \in C'_{r,n}$ tal que $\#(f^{-1}(I_{n_1})) = k$. Denotemos com A_k êste conjunto de funções. Antes de efetuar o cálculo de $\#(A_k)$, vamos determinar os possíveis valores de k. Há quatro casos possíveis dependendo da relação r com n_1 e n_2

$r \le n_1$,	$r \le n_2$	$0 \le k \le r$
$r \le n_1$	$r > n_2$	$r-n_2 \le k \le r$
$r > n_1$	$r \le n_2$	$0 \le k \le n_1$
$r > n_1$	$r > n_2$	$r-n_2 \le k \le n_1$

É fácil ver que êstes 4 casos podem ser combinados na fórmula

$$(r-n_2)\vee 0 = (r-n_2)^+ \le k \le r\wedge n_1 \ .$$

Existe claramente uma correspondência biunívoca φ entre C'_{r-k,n_2} e o conjunto de tôdas as aplicações estritamente crescentes de $\{k+1,\dots,r\}$ em $\{n_1+1,\dots,n\}$.
A aplicação $f \longmapsto (f|_{I_k}, \varphi(f|_{\{k+1,\dots,r\}}))$ estabelece uma correspondência biunívoca entre A_k e $C'_{k,n_1} \times C'_{r-k,n_2}$ e portanto

$$\#(A_k) = \binom{n_1}{k}\binom{n_2}{r-k} \qquad \text{e}$$

$$P(A_k) = \frac{\binom{n_1}{k}\binom{n_2}{r-k}}{\binom{n}{r}} \ .$$

É importante observar que teria sido perfeitamente natural tomar como espaço amostral $A_{r,n}$. Neste caso, se A'_k denota o conjunto das funções de $A_{r,n}$ tal que $\#(f^{-1}(I_{n_1})) = k$ resulta que

$$\#(A'_k) = \binom{r}{k}\{n_1(n_1-1)\ldots(n_1-k+1)\}\{n_2(n_2-1)\ldots(n_2-r+k+1)\}$$

e

$$P'(A'_k) = \frac{\binom{r}{k}\{n_1(n_1-1)\ldots(n_1-k+1)\}\{n_2(n_2-1)\ldots(n_2-r+k+1)\}}{n(n-1)\ldots(n-r+1)}$$

É fácil ver que $P(A_k) = P'(A'_k)$.

Os números $p_k = P(A_k)$, $(r-n_2)\vee 0 \le k \le r\wedge n_1$ determinam uma distribuição que é chamada distribuição hipergeométrica.

EXEMPLO 3.4.2 - Suponhamos que, de n objetos escohemos r ao acaso, com reposição. Qual é a probabilidade de que nenhum objeto seja escolhido mais de uma vez?

Como espaço amostral tomemos $I_n^{I_r}$. O conjunto cuja probabilidade queremos calcular é $A_{r,n}$ e, portanto, a probabilidade procurada é

$$\frac{\#(A_{r,n})}{\#(I_n^{I_r})} = \frac{n(n-1)\ldots(n-r+1)}{n^r} = 1(1-\frac{1}{n})(1-\frac{2}{n})\ldots(1-\frac{(r-1)}{n}).$$

Uma cota superior para esta probabilidade se obtém fàcilmente usando a desigualdade $1-x \le e^{-x}$. Resulta então:

$$P(A_{r,n}) \leq \prod_{i=1}^{r-1} e^{-\frac{i}{n}} = e^{-\frac{1}{n}\sum_{i=1}^{r-1} i} = e^{-\frac{(r-1)r}{2n}} .$$

Uma aplicação interessante dêste resultado é a seguinte. Suponhamos que o aniversário de uma pessoa pode cair com igual probabilidade em qualquer um dos 365 dias do ano. Se r pessoas são escolhidas ao acaso a probabilidade de que todos façam anos em dias diferentes é dada pela fórmula anterior. A cota superior $e^{-\frac{(r-1)r}{730}}$ dá interessantes resultados. Por exemplo, para $r \geq 23$ e já menos que $\frac{1}{2}$.

EXEMPLO 3.4.3 - Uma urna contém n bolinhas numeradas de 1 a n. As bolinhas são escolhidas uma a uma, ao acaso, até esvaziar a urna. Se a bolinha r aparece na r-ésima extração, diz-se que ocorre um reencontro. Calcular a probabilidade de que ocorre pelo menos um reencontro.

Seja A_i, $i=1,2,\ldots,n$ o evento que ocorre se e só se ocorre um reencontro na i-ésima extração. Temos que calcular $P(\bigcup_{i=1}^{n} A_i)$. Pela Proposição 3.3.6 temos que

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < i_2 \ldots < i_k \leq n} P(A_{i_1} \cap \ldots \cap A_{i_k}) .$$

Agora $P(A_{i_1} \cap \ldots \cap A_{i_k}) = \frac{(n-k)!}{n!}$ e então

$$\sum_{1\le i_1<\ldots<i_k\le n} P(A_{i_1}\cap\ldots\cap A_{i_k}) = \binom{n}{k}\frac{(n-k)!}{n!} = \frac{1}{k!}$$

e portanto

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n}(-1)^{k-1}\frac{1}{k!} = 1-\sum_{k=0}^{n}\frac{(-1)^k}{k!}\ .$$

Observe que o limite do membro direito é $1-e^{-1}$ que é boa aproximação inclusive para n pequeno ($n\ge 3$).

EXEMPLO 3.4.4 - Suponhamos que se joga uma moeda n vêzes. Como espaço amostral, tomamos $\{0,1\}^{I_n}$. P é definida como de costume. A probabilidade da sucessão $(0,\ldots,0,1)$ (obter a primeira cara no tempo n) é portanto $\frac{1}{2^n}$.

Se se define $p_n = \frac{1}{2^n}$, $n=1,2,\ldots$, é natural tomar Ω_7 e a probabilidade induzida pela sucessão $\{p_n\}$ como espaço de probabilidades para descrever o experimento E_7. Temos sòmente que verificar que $\sum_{n=1}^{\infty} p_n = 1$. Mas,

$$\sum_{n=1}^{\infty}\frac{1}{2^n} = \frac{\frac{1}{2}}{1-\frac{1}{2}} = 1$$

(série geométrica de razão $\frac{1}{2}$).

3.5 - Falácias em Problemas Combinatórios

Há dois tipos de erros comuns na resolução de problemas combinatórios. O primeiro consiste em calcular a probabilidade de eventos em um espaço de probabilidades que não é o eleito ao começar o problema. Um exemplo simplies é o seguinte.

EXEMPLO 1 - De um baralho, três cartas são retiradas ao acaso. Calcular a probabilidade de não obter ases.

Solução proposta: O número total de possibilidades $\binom{52}{3}$. Agora, de que a primeira carta não seja um ás há 48 possibilidades; 47 da segunda; 46 da terceira. A probabilidade desejada é

$$\frac{48.47.46}{\binom{52}{3}}$$

Êrro: Os casos favoráveis foram calculados em $A_{3,52}$ e não em $C_{3,52}$ como deveriam já que êste é o espaço de amostras escolhido. É fácil ver que a solução correta é

$$\binom{48}{3} / \binom{52}{3}.$$

O segundo tipo de êrro não é de natureza matemática e consiste na escolha de um espaço de amostras que não é o que corresponde ao problema real. O seguinte é um exemplo dessa situação.

EXEMPLO 2 - De um baralho, três cartas são escolhidas com reposição. Qual é a probabilidade de que essas cartas sejam, em alguma ordem, o ás de espadas, o 10 de espadas e o ás de copas.

<u>Solução proposta:</u> O número de amostras não ordenadas com reposição é $\binom{52+3-1}{3}$ e portanto a probabilidade dessa amostra particular é $1/\binom{54}{3} =$

<u>Êrro</u> - O espaço amostral eleito $C'_{3,52}$ não é adequado, no sentido que, se o experimento é repetido um grande número de vêzes, as freqüências relativas vão diferir consistentemente das probabilidades dadas pelo modêlo. A resposta correta é $\frac{3!}{52^3}$.

EXERCÍCIOS

1. a) Provar que para tôda sucessão de conjuntos $\{A_n\}_{n=1,2,\ldots}$

$$P(\lim_n \inf A_n) \leq \lim_n \inf P(A_n) \leq \lim_n \sup P(A_n) \leq P(\lim_n \sup A_n)$$

b) Provar que se $\lim_n A_n$ existe, então

$$P(\lim_n A_n) = \lim_n P(A_n) .$$

2. Seja $\Omega = \{x: 0 \leq x \leq 1,\ x$ racional$\}$. Se $[a,b)$ é um intervalo racional contido em Ω definimos $P([a,b)) = b-a$. (O comprimento do intervalo). Usando técnicas fora do alcance dêste curso é possível provar que P pode ser extendida a $P(\Omega)$. E dizer, existe $P: P(\Omega) \to [0,1]$ tal que P é uma probabilidade finitamente aditiva e $P([a,b)) = b-a$. Prove que essa probabilidade finitamente aditiva não é uma probabilidade.

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

4. Uma "hat check girl" recebeu n chapéus, mas êstes ficaram totalmente misturados. Decidiu então, devolvê-los a êsmo. Calcular a probabilidade de que nenhum homem recebeu o seu. (É interessante tratar de adivinhar o comportamento dessa probabilidade quando n é grande antes de efetuar o cálculo. Depois de efetuado observe que o resultado é pràticamente independente de n, para $n \geq 4$ a variação e menor que 0,01).

5. (Futebol). Para a Copa do Mundo 16 times são divididos em quatro grupos de 4 times cada um. Calcular a probabilidade de que dois times determinados A e B se encontrem no mesmo grupo. (Na realidade a escolha não é a êsmo).

6. No exemplo 3.4.4 calcular a probabilidade de que

   a) O número de jogadas seja ≥ 4

   b) O número de jogadas seja um número par.

7. Um número com 3 dígitos é escohido a êsmo (desde 000 a 999). Calcular

   a) A probabilidade de que a soma dos dígitos seja ≥ 9.

   b) A probabilidade de que exatamente um dígito seja >7.

8. Uma caixa contém 20 peças em boas condições e 15 em más condições. Uma amostra de 10 peças é extraída. Calcular a probabilidade de que ao menos uma peça na amostra seja defeituosa.

9. Se $\{P_i\}_{i=1,\ldots,n}$ são probabilidades sôbre Ω e $\{\alpha_i\}_{i=1,\ldots,n}$ são números tais que $\alpha_i \geq 0$ e $\sum_{i=1}^{n} \alpha_i =$ $= 1$, então $\sum_{i=1}^{n} \alpha_i P_i$ e também uma probabilidade sôbre Ω.

10. Uma loteria tem n números e só um prêmio. Um jogador compara n bilhetes em uma jogada. Outro jogador compra só um bilhete em n jogos diferentes. (E dizer ambos jogadores apostam a mesma quantidade). Qual é o jogador que tem à maior probabilidade de ganhar o prêmio?

<u>Sugestão</u>: $(1 - \frac{1}{N})^m > (1 - \frac{1}{N})^{m-1} - \frac{1}{N}$).

11. Se $\sum_{n=1}^{\infty} P(A_n) < \infty$ então $P(\lim_n \sup A_n) = 0$.

12. Seja (Ω,P) um espaço de probabilidades tal que $P(\{\omega_i\}) = p_i > 0 \quad i=1,2,\dots$.

Definamos para $A \subseteq \Omega$ e $B \subseteq \Omega$

$$d(A,B) = P(A \Delta B)$$

a) Provar que $(\mathbb{P}(\Omega),d)$ é um espaço métrico.

b) Provar que as funções $(A,B) \to A\cup B$, $(A,B) \to A\cap B$, e $A \to P(A)$ são uniformemente contínuas.

c) Provar que $(\mathbb{P}(\Omega),d)$ é totalmente limitado.

13. Provar que as sucessões $p_n = \frac{q^n}{cn} \quad 0 \leq q < 1$, $p = 1-q$ $c = \log(\frac{1}{p})$ e $p'_n = \frac{1}{e^\lambda - 1} \frac{\lambda^n}{n!}$, $n = 1,2,\dots$ definem probabilidades sôbre $\Omega = \{1,2,\dots\}$. (A primeira destas é chamada distribuição logarítmica).

CAPÍTULO 4

INDEPENDÊNCIA

4.1 - Probabilidade Condicional

Consideremos o experimento que consiste em atirar um dado equilibrado. Seja $\Omega = \{1,2,\ldots,6\}$ e $p_1 = p_2 = \ldots = p_6 = \frac{1}{6}$. Sejam $A = \{2,4,6\}$ e $B = \{1,2,3\}$. Temos $P(B) = \frac{1}{2}$. Esta é nossa probabilidade de B a priori, quer dizer, antes que o experimento se realize. Suponhamos que, uma vez realizado o experimento, alguém nos informe que o resultado do mesmo é um número par, isto é, que A ocorreu. Nossa opinião sôbre a ocorrência de B se modifica com esta informação já que, se A ocorrer, só se o resultado do experimento fôr 2 poderá ter ocorrido também B; quer dizer, a probabilidade de B a posteriori, ou como nos vamos referir a ela doravante, a probabilidade condional de B dado A é $\frac{\#(\{2\})}{\#(A)} = \frac{1}{3}$.

Observe que $\frac{\#(\{2\})}{\#(A)} = \frac{P(\{2\})}{P(A)} = \frac{P(A\cap B)}{P(A)}$.

Introduzimos agora a seguinte:

DEFINIÇÃO 4.1.1 - A probabilidade condicional de B dado

A, que denotaremos $P(B/A)$, é o número $\frac{P(A\cap B)}{P(A)}$. Êste número está definido só para o caso em que $P(A) > 0$. Simbòlicamente, $P(B/A) = \frac{P(B\cap A)}{P(A)}$ se $P(A) > 0$ (4.1.1).

Esta fórmula também se escreve usualmente da seguinte maneira:

$$P(A\cap B) = P(A)P(B/A) \qquad (4.1.2)$$

Se $P(B) > 0$ também temos:

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

Fixado A, $P(B/A)$ se comporte como uma nova probabilidade sôbre $\mathbb{P}(\Omega)$. Mais precisamente, temos a seguinte:

PROPOSIÇÃO 4.1.1 - Se $P(A) > 0$ então

i) $P(\emptyset/A) = 0$ e $P(\Omega/A) = 1$

ii) Se $\{B_i\}_{i=1,2,\ldots}$ é uma família disjunta, então

$$P(\sum_{i=1}^{\infty} B_i/A) = \sum_{i=1}^{\infty} P(B_i/A) .$$

<u>Demonstração</u>: ver exercício 9.

4.2 - Teorema de Bayes

Seja $\{A_i\}_{i=1,2,\ldots,n}$ uma família finita de conjuntos disjuntos que cobrem um conjunto B $(B \subseteq \sum_{i=1}^{n} A_i)$

e tal que $P(A_i) > 0$, $i=1,2,\ldots,n$, e $P(B) > 0$. A situação é indicada na figura 4.2.1.

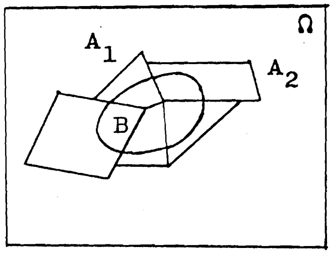


Figura 4.2.1

Agora $P(B) = \sum_{i=1}^{n} P(A_i \cap B) = \sum_{i=1}^{n} P(A_i)P(B/A_i)$ (4.2.1).

A primeira igualdade decorre de $B \subseteq \sum_{i=1}^{n} A_i$ e de que os A_i são disjuntos; a segunda igualdade é conseqüência da equação (4.1.2).

A fórmula (4.2.1) é chamada, às vêzes, Teorema da Probabilidade Total.

Temos aqui uma expressão de $P(B)$ em função das probabilidades dos conjuntos que cobrem B e das probabilidades condicionais de B dado cada um dêles. O seguinte teorema é conseqüência imediata desta fórmula.

TEOREMA 4.2.1 (Bayes):

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{\sum_{j=1}^{n} P(A_j)P(B/A_j)} \qquad i=1,2,\ldots,n \quad .$$

Demonstração:

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)\ P(B/A_i)}{P(B)} \quad .$$

Esta última igualdade segue do uso da identidade (4.1.2). O teorema se obtém agora substituindo $P(B)$ pela sua expressão na fórmula (4.2.1).

4.3 - Independência

Sejam (Ω, P) um espaço de probabilidades e A e B dois eventos. Há situações nas quais a ocorrência de A claramente não modifica nossa opinião sôbre B. Formalmente, escreveríamos $P(B/A) = P(B)$ $\qquad P(A) > 0$ e diríamos que A e B são independentes. Para evitar a restrição $P(A) > 0$, escrevemos a igualdade anterior na forma

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

e damos então a seguinte

DEFINIÇÃO 4.3.1 - Dois eventos A e B são <u>independentes</u> se e só se $P(A \cap B) = P(A)P(B)$.

Em geral, para uma família $\{A_\alpha\}_{\alpha \in \Gamma}$ qualquer, definimos

DEFINIÇÃO 4.3.2 - Dizemos que a família $\{A_\alpha\}_{\alpha \in \Gamma}$ é uma família de eventos independentes quando

$$\forall n, \forall \alpha_1, \alpha_2, \ldots, \alpha_n \in \Gamma \quad P(A_{\alpha_1} \ldots A_{\alpha_n}) = \prod_{i=1}^{n} P(A_{\alpha_i}) \quad .$$

Note-se que para provar que dois eventos são independentes só temos que verificar uma identidade. Para provar que três eventos são independentes há que verificar 4 identidades. Em geral, para provar a independência de n eventos, há que verificar 2^n-n-1 identidades. Ω e ϕ são independentes de qualquer evento.

O leitor deve convencer-se por sua conta que certos eventos que são intuitivamente independentes o são também de acôrdo com a nossa definição. Por exemplo, os resultados obtidos nas primeira e segunda jogadas de um dado arremessado duas vêzes, etc.

Vamos deduzir agora algumas conseqüências da definição de independência.

LEMA 4.3.1 - Se $\{A_1, A_2, \ldots, A_r\}$ são independentes então $P(\bigcap_{i=1}^{r} B_i) = \prod_{i=1}^{r} P(B_i)$ onde B_i é igual a A_i ou a A_i^c.

<u>Demonstração</u>: Por indução no número de conjunto que trocamos de A_i para A_i^c.

Para $n=1$, suponhamos $B_1 = A_1^c$ e $B_i = A_i$ para $i > 1$.

Temos

$$P(A_1^c\ A_2\ldots A_n) + P(A_1A_2\ldots A_n) = P(A_2A_3\ldots A_n)$$

Isto é

$$P(A_1^c\ A_2\ldots A_n) + \prod_{i=1}^{r} P(A_i) = \prod_{i=2}^{n} P(A_i)$$

$$P(A_1^c\ A_2\ldots A_n) = \prod_{i=2}^{n} P(A_i)[1-P(A_1)] = \prod_{i=2}^{n} P(A_i)P(A_1^c)$$

que é o resultado para $n=1$.

Suponhamos que o resultado é verdadeiro para qualquer família de conjuntos independentes na qual se tenham efetuado exatamente n trocas. Consideremos uma família em que se efetuam $n+1$ trocas. Suponhamos sem perda de generalidade que

$$B_1 = A_1^c\ .\ .\ .\ .\ B_{n+1} = A_{n+1}^c \qquad B_i = A_i,\quad n+1 < i \leq r$$

$$P(\bigcap_{i=1}^{n+1} A_i^c \bigcap_{i=n+2}^{r} A_i) + P(\bigcap_{i=1}^{n} A_i^c \cap A_{n+1} \cap \bigcap_{i=n+2}^{r} A_i) = P(\bigcap_{i=1}^{n} A_i^c \cap \bigcap_{i=n+2}^{r} A_i)$$

Êste último têrmo é igual, por indução, a

$$\prod_{i=1}^{n} P(A_i^c) \prod_{i=n+2}^{r} P(A_i)$$

e temos

$$P(\bigcap_{i=1}^{r} B_i) = \prod_{i=1}^{n} P(A_i^c)[1-P(A_{n+1})] \prod_{i=n+2}^{r} P(A_i) = \prod_{i=1}^{r} P(B_i)$$

que é o resultado desejado.

Êste lema dá imediatamente a seguinte:

PROPOSIÇÃO 4.3.1 - Se $\{A_1,\ldots,A_n\}$ são independentes, então $\{B_1,\ldots,B_r\}$ também são independentes, onde $r \leq n$ e B_i é igual a A_i ou A_i^c para todo $1 \leq i \leq r$.

PROPOSIÇÃO 4.3.2 - Se $\{A_1,\ldots,A_n\}$ são independentes, então

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

Demonstração:

$$P(\bigcup_{i=1}^{n} A_i) = 1-P(\bigcap_{i=1}^{n} A_i^c) = 1-\prod_{i=1}^{n} P(A_i^c) = 1-\prod_{i=1}^{n} [1-P(A_i)] \; .$$

A segunda igualdade segue da proposição 4.3.1. Usando a desigualdade $1-x \leq e^{-x}$ temos agora

$$P(\bigcup_{i=1}^{n} A_i) \geq 1 - \prod_{i=1}^{n} e^{-P(A_i)} = 1-e^{-\sum_{i=1}^{n} P(A_i)}$$

que é o resultado.

PROPOSIÇÃO 4.3.3 (Lema de Borel-Cantelli) - Se $\{A_i\}_{i=1,2,\ldots}$ é uma sucessão de eventos independentes, então, se $\sum_{i=1}^{\infty} P(A_i) = +\infty$, $P(\lim_n \sup A_n) = 1$.

Demonstração:

$$P(\lim_n \sup A_n) = P(\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i) = \lim_n P(\bigcup_{i=n}^{\infty} A_i) =$$

$$= \lim_n \lim_p P(\bigcup_{i=n}^{n+p} A_i) \geq \lim_n \lim_p [1 - e^{-\sum_{i=n}^{n+p} P(A_i)}] = 1 \; .$$

4.4 - Modelos Produto

Ocorre muito frequentemente que certos experimentos aleatórios resultam da combinação de vários experimentos distintos. Por exemplo, o experimento que consistem em atirar 3 moedas simultâneamente pode ser pensado como o resultado de efetuar 3 experimentos em cada um dos quais só uma moeda é usada. O extrair de uma urna contendo mais de duas bolinhas duas delas simultâneamente pode ser pensado como a execução de dois experimentos sucessivos consistentes cada um dêles na extração de uma só bolinha. O problema de combinar experimentos aleatórios simples para obter outros mais complicados vai ser analisado neste capítulo. Vamos começar pelo importante caso de

<u>Experimentos produto</u>. Para evitar repetições vamos descrever o problema em tôda generalidade ainda que seja aconselhável que o leitor pense em algum caso concreto como o de n repetições do experimento consistente em jogar uma moeda por exemplo.

Sejam $E_1, E_2, \ldots, E_n$ n experimentos com seus correspondentes espaços de probabilidade $(\Omega_1, P_1), \ldots, (\Omega_n, P_n)$ onde $\Omega_i = \{\omega_{i1}, \omega_{i2}, \ldots\}$.

Definamos

$$p_{ij} = P_i(\{\omega_{ij}\}) \qquad i=1,2,\ldots,n \qquad j=1,2,\ldots$$

É claro que $p_{ij} \geq 0$ e $\sum_{j=1}^{\infty} p_{ij} = 1$.

É natural tomar como espaço amostral para o experimento composto o conjunto $\Omega = \prod_{i=1}^{n} \Omega_i$.

Seja $A_{ij} = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_{i-1} \times \{\omega_{ij}\} \times \Omega_{i+1} \times \ldots \times \Omega_n$.

Quer dizer, A_{ij} é o evento que ocorre se e só se o experimento i-ésimo dá como resultado ω_{ij}. O que queremos agora é definir P sôbre $\mathbb{P}(\Omega)$ de tal maneira que os eventos A_{ij_i}, $i=1,\ldots,n$ resultem independentes.
Isto é equivalente a

$$P\left(\bigcap_{i=1}^{n} A_{ij_i}\right) = \prod_{i=1}^{n} P(A_{ij_i})$$

Mas

$$\bigcap_{i=1}^{n} A_{ij_i} = (\omega_{1j_1}, \omega_{2j_2}, \ldots, \omega_{nj_n})$$

e $P(A_{ij_i})$ deveria ser igual a p_{ij_i}. Ou seja

$$P(\omega_{1j_1}, \omega_{2j_2}, \ldots, \omega_{nj_n}) = \prod_{i=1}^{n} p_{ij_i} .$$

Esta equação define P para todo elemento de Ω.

Há que verificar agora que P é uma probabilidade, para o que, pela Proposição 3.4.1, basta ver que

$$\sum_{j_1, j_2, \ldots, j_n} \prod_{i=1}^{n} p_{ij_i} = 1 .$$

Procedemos por indução. Para $n=1$, a igualdade é trivial.

Agora,

$$\sum_{j_1,j_2,\ldots,j_n,j_{n+1}} \prod_{i=1}^{n+1} p_{ij_i} =$$

$$\sum_{j_1,j_2,\ldots,j_n} \prod_{i=1}^{n} p_{ij_i} \left(\sum_{j_{n+1}=1} p_{(n+1)j_{(n+1)}} \right) =$$

$$= \sum_{j_1,j_2,\ldots,j_n} \prod_{i=1}^{n} p_{ij_i} = 1 \quad .$$

A segunda igualdade segue do fato que P_{n+1} é uma probabilidade e a última, da hipótese indutiva. Resumindo, temos a seguinte

PROPOSIÇÃO 4.4.1 - Existe uma única probabilidade sôbre $\mathbb{P}(\Omega)$ que converte os eventos A_{ij_i}, $i=1,\ldots,n$ em eventos independentes. Esta probabilidade é dada pela fórmula

$$P(\omega_{1j_1},\omega_{2j_2},\ldots,\omega_{nj_n}) = \prod_{i=1}^{n} p_{ij_i} \ .$$

Como exemplo, tomemos o caso em que

$$\Omega_1 = \Omega_2 = \ldots = \Omega_n = \{0,1\}$$

Ω é então igual a

$$\underbrace{\{0,1\} \times \ldots \times \{0,1\}}_{n} = \{0,1\}^{I_n}$$

quer dizer, Ω consiste de tôdas as sucessões de zeros e uns de comprimento n. Se $(e_s,\ldots,e_n)$ é uma dessas sucessões, $P((e_1,e_2,\ldots,e_n)) = \prod_{i=1}^{n} \beta_i$ onde

$$\beta_i = \begin{cases} p_i & \text{se } \epsilon_i = 1 \\ 1-p_i & \text{se } \epsilon_i = 0 \end{cases}$$

Se tôdas as probabilidades p_i, $i=1,\ldots,n$, são iguais (isto é, se $p = p_1 = p_2 = \ldots = p_n$), resulta que

$$P((\epsilon_1,\epsilon_2,\ldots,\epsilon_n)) = p^{\sum_{i=1}^{n} \epsilon_i} (1-p)^{n-\sum_{i=1}^{n} \epsilon_i} .$$

Note que $\sum_{i=1}^{n} \epsilon_i$ é igual ao número de uns (caras) e $n-\sum_{i=1}^{n} \epsilon_i$ é igual ao número de zeros (coroas).

É comum em muitas aplicações calcular probabilidades de eventos sem precisar qual é o espaço de probabilidades com que se está trabalhando. A existência de tal espaço é muitas vêzes garantida pela Proposição 4.4.1. Vejamos uma dessas aplicações.

EXEMPLO 4.4.1 - A probabilidade de fechamento de cada relé do circuito apresentado na figura 4.4.1 é igual a p, $0 < p < 1$. Se tôdas ao relés funcionam independentemente, qual é a probabilidade de que haja corrente circulando entre os terminais C e B.

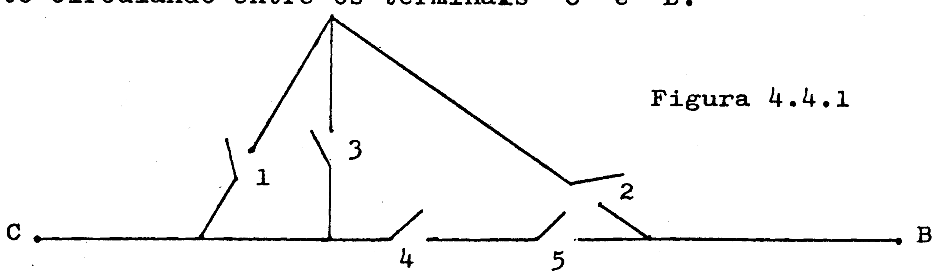


Figura 4.4.1

Podemos representar o estado de cada relé com o conjunto $\{0,1\}$ (0: aberto, 1: fechado). Portanto $\Omega = \{0,1\}^{I_5}$ seria o espaço amostral dêste experimento e, como se supõe que ao relés funcionam independentemente

$$P((e_1,e_2,\ldots,e_5)) = p^{\sum_{i=1}^{5} e_i} (1-p)^{5-\sum_{i=1}^{5} e_i}$$

Isto é uma probabili ade, pela Proposição 4.4.1.

Seja $A_i = \{(e_1,\ldots,e_5): e_i = 1\}$, $i=1,2,\ldots,5$ e seja A_i o evento que ocorre se e só se há corrente entre os terminais C e B. Então

$$A = (A_1A_2) \cup (A_4A_5) \cup (A_2A_3)$$

$$\begin{aligned} P(A) &= P(A_1A_2) + P(A_4A_5) + P(A_2A_3) - P(A_1A_2A_4A_5) - \\ &\quad - P(A_1A_2A_3) - P(A_2A_3A_4A_5) + P(A_1A_2A_3A_4A_5) = \\ &= p^2+p^2+p^2-p^4-p^3-p^4+p^5 = 3p^2 - 2p^4 - p^3 + p^5 \end{aligned}$$

Para êste experimento, descrevemos formalmente o espaço de probabilidades subjacente. Não vamos seguir êste procedimento fielmente no futuro, devido a seu caráter rotineiro; mas se aconselha o leitor a repetí-lo até familiarizar-se com o mesmo.

4.5 - Independência a Pares e Independência

DEFINIÇÃO 4.5.1 - Sejam (Ω, P) um espaço amostral e $A_1, A_2, \ldots, A_n$ n subconjuntos de Ω. Dizemos que $A_1, A_2, \ldots, A_n$ são independentes a pares se e só se $\forall i$ e $\forall j \in \{1, \ldots, n\}$, $i \neq j$, $P(A_i \cap A_j) = P(A_i)P(A_j)$.

A seguir vamos dar um exemplo em que 3 eventos são independentes a pares mas não são independentes.

Seja Ω o espaço amostral apresentado na Figura 4.5.1, possuindo 4 pontos e P a probabilidade que associa a cada ponto o valor $\frac{1}{4}$.

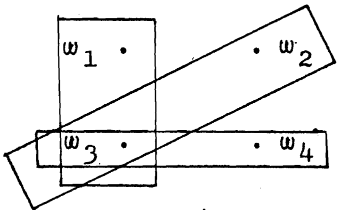


Figura 4.5.1

Sejam $C = \{\omega_1, \omega_3\}$, $F = \{\omega_3, \omega_4\}$ e $D = \{\omega_2, \omega_3\}$ três eventos, correspondentes à primeira coluna, à segunda fita é a diagonal, respectivamente. Resulta que

$$P(C) = P(F) = P(D) = \frac{1}{2}$$
$$P(CF) = P(CD) = P(FD) = P(\{\omega_3\}) = \frac{1}{4}$$

e portanto êstes três eventos são independentes a pares.

Mas $P(CFD) = P(\{\omega_3\}) = \frac{1}{4} \neq (\frac{1}{2})^3$ e conseqüentemente não são independentes.

O que se passa é que dois dêles já determinam o resultado do experimento.

4.6 - *Árvores*

Na seção 4.4 consideramos uma sucessão de experimentos na qual cada um era realizado sem ter em conta os resultados anteriores. É fácil imaginar uma sucessão de experimentos na qual as possibilidades de um experimento dependem dos resultados de experimentos anteriormente realizados. Vamos descrever a situação em geral. Sejam $\{T_i\}_{i=0,\dots,n}$ e $\{S_i\}_{i=0,1,\dots,n-1}$ duas famílias de conjuntos não vazios tais que $T_i \cap S_i = \emptyset$, $i=0,\dots,n-1$. Sejam $\Omega_i = T_i + S_i$, $i=0,\dots,n-1$ e $\Omega_n = T_n$. Sejam μ uma probabilidade sôbre Ω_o e, para todo $k=1,\dots,n-1$ e $(i_o,\dots,i_{k-1}) \in S_o \times S_1 \times\dots\times S_{k-1}$, $p_{i_o\dots i_{k-1}}$ uma probabilidade sôbre Ω_k.

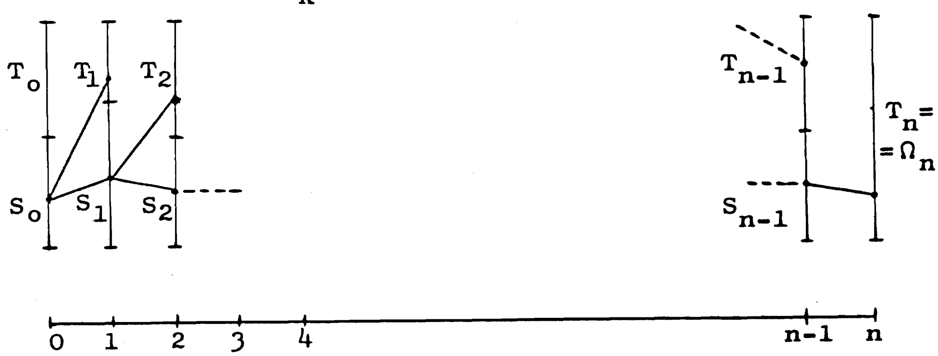


Figura 4.6.1

O experimento se realiza da seguinte forma (Ver figura 4.6.1).

1. Escolhe-se um ponto ao acaso em Ω_o, usando μ. Seja i_o êsse ponto.
2. Se $i_o \in T_o$, o experimento termina. Se $i_o \in S_o$, se escolhe um ponto em Ω_1, usando p_{i_o}. Seja i_1 o ponto agora escolhido.
3. Se $i_1 \in T_1$, o experimento termina. Se $i_1 \in S_1$, se escolhe um ponto i_2 em Ω_2, usando $p_{i_o i_1}$.

----

n+1. Se $i_{n-1} \in T_{n-1}$, o experimento termina. Se $i_{n-1} \in$ $\in S_{n-1}$, se escolhe um ponto em $\Omega_n = T_n$, usando $p_{i_o \ldots i_{n-1}}$.

O problema é agora construir um espaço de probabilidades para êste experimento composto. É natural tomar como espaço amostral o conjunto de <u>trajetórias</u> possíveis, isto é, o conjunto

$$\Omega = T_o+(S_o \times T_1) + (S_o \times S_1 \times T_2) +\ldots+ (S_o \times S_1 \times \ldots \times S_{n-1} \times T_n)$$

(motivado pela Figura 4.6.1 é também natural chamar Ω de árvore a cada trajetória de ramo).

Tendo em conta a definição de probabilidade condicional, definimos

$$p(i_o) = \mu(i_o) \qquad i_o \in T_o$$

$$p(i_o,i_1,\ldots,i_k) = \mu(i_o)p_{i_o}(i_1)\ldots p_{i_oi_1\ldots i_{k-1}}(i_k)$$
$$\text{para } (i_o,\ldots,i_k) \in S_o\times\ldots\times S_{k-1}\times T_k$$

Temos a seguinte

PROPOSIÇÃO 4.6.1 - $\sum_{w\in\Omega} p(w) = 1$ (e portanto p induz uma probabilidade P sôbre Ω, pela Proposição 3.4.1).

Demonstração: Vamos escrever a demonstração no caso $n=2$. O caso geral é idêntico. Êste caso particular dá uma idéia precisa do que acontece. O caso geral é deixado como exercício (Ver exercício 10).

Temos que provar que

$$\sum_{i_o\in T_o} \mu(i_o) + \sum_{i_o\in S_o} \sum_{i_1\in T_1} \mu(i_o)\, p_{i_o}(i_1) +$$
$$+ \sum_{i_o\in S_o} \mu(i_o) \sum_{i_1\in S_1} p_{i_o}(i_1) \sum_{i_2\in T_2} p_{i_oi_1}(i_2) = 1 .$$

Agora o membro esquerdo é igual a

$$\mu(T_o) + \sum_{i_o\in S_o} \mu(i_o)p_{i_o}(T_1) + \sum_{i_o\in S_o} \mu(i_o) \sum_{i_1\in S_1} p_{i_o}(i_1)$$

porque $\sum_{i_2\in T_2} p_{i_oi_1}(i_2) = 1$. Como $\sum_{i_1\in S_1} p_{i_o}(i_1) = p_{i_o}(S_1)$

temos que esta expressão é igual a

$$\mu(T_o) + \sum_{i_o \in S_o} \mu(i_o)p_{i_o}(T_1) + \sum_{i_o \in S_o} \mu(i_o)p_{i_o}(S_1) =$$

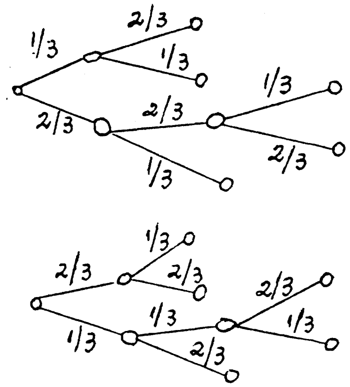
$$\mu(T_o) + \sum_{i_o \in S_o} \mu(i_o)[p_{i_o}(T_1)+p_{i_o}(S_1)] = \mu(T_o) + \sum_{i_o \in S_o} \mu(i_o)$$

$$= \mu(T_o) + \mu(S_o) = \mu(\Omega_o) = 1$$

A proposição anterior permite formalizar uma série de problemas nos quais probabilidades e probabilidades condicionais de certos eventos são dadas e se requer o cálculo de outras com êsses dados. A proposição 4.6.1 permite a construção de um espaço de probabilidades no qual tôdas as manipulações têm sentido. Os seguintes exemplos são uma ilustração dêsse fato.

EXEMPLO 4.6.1 - Um jogador ganha em um torneio se consegue ganhar dois jogos sucessivamente, de uma série de três. Seus oponentes são os jogadores A e B. A probabilidade de ganhar de A é $\frac{1}{3}$. De ganhar de B é $\frac{2}{3}$. Que série lhe convém escolher? (Tente adivinhar a resposta antes de ler a solução).

Solução: As duas séries possíveis são ABA e BAB. As árvores para estas duas séries são



(série ABA)

(série BAB)

Figura 4.6.2

Os conjuntos utilizados na Proposição 4.6.1 neste exemplo são os seguintes (para a primeira série):

$\Omega_o = S_o = \{A\}$

$\Omega_1 = S_1 = \{G,P\}$

$S_2 = \{(P,G)\}$

$T_2 = \{(G,G),(G,P),(P,P)\}$

$\Omega_2 = S_2 + T_2$

$\Omega_3 = T_3 = \{(P,G,G),(P,G,P)\}$

As probabilidades de transição $p_{i_o \dots i_k}$ são as indicadas na árvore. O leitor pode fazer a mesma descrição formal para a série BAB.

A probabilidade de ganhar, se escolhe a primeira série, é $\frac{1}{3}\,\frac{2}{3} + \frac{2}{3}\,\frac{2}{3}\,\frac{1}{3} = \frac{6+4}{27} = \frac{10}{27}$.

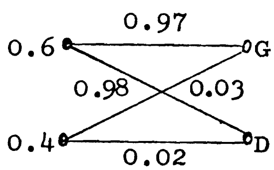
A probabilidade de ganhar, se escolhe a segunda série, é $\frac{2}{3}\,\frac{1}{3} + \frac{1}{3}\,\frac{1}{3}\,\frac{2}{3} = \frac{6+2}{27} = \frac{8}{27}$.

Ou seja, a primeira série (que tem duas vêzes A) é a mais conveniente (procure dar uma resposta intuitiva a êste resultado).

Nota: os conjuntos definidos neste exemplo para poder utilizar a Proposição 4.6.1 não são necessàriamente os que usamos acima, evidentemente.

EXEMPLO 4.6.2 - Da produção total de duas máquinas A e B se sabe que 60% são produzidos por A e 40% por B. Dos artigos produzìdos por A, 3% são defeituosos; dos produzidos por B, só 2% são defeituosos. Da produção total um artigo é escolhido e se verifica que é defeituoso. Qual é a probabilidade de que tenha sido produzido por A?

Solução: A árvore para êste experimento é



$$S_o = \Omega_o = \{A,B\}$$

$$T_1 = \Omega_1 = \{G,D\}$$

Figura 4.6.3

As probabilidades de transição $p_{i_o \ldots i_k}$ estão dadas na figura.

Sejam

$$A' = \{(A,G),(A,D)\}$$
$$B' = \{(B,G),(B,D)\}$$
$$D' = \{(A,D),(B,D)\}$$

Queremos calcular $P(A'/D')$. Usando a Fórmula de Payes temos:

$$P(A'/D') = \frac{P(A')\ P(D'/A')}{P(A')\ P(D'/A')+P(B')\ P(D'/B')} =$$

$$= \frac{0.6 \quad 0.03}{0.6 \quad 0.03 + 0.4 \quad 0.02} = \frac{9}{13}$$

EXEMPLO 4.6.3 - Três chaves estão ligadas como na figura 4.6.4

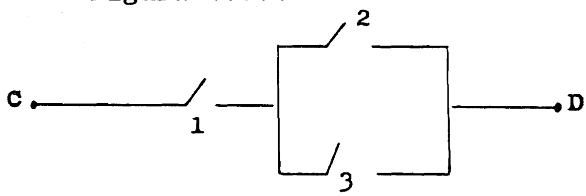


Figura 4.6.4

Seja B_i, $i=1,2,3$, o evento que o relé i esteja aberto. Suponhamos que B_1 e B_2 são independentes e as probabilidades condicionais para B_3 estão dadas por

$$P(B_1) = \frac{1}{4} \qquad P(B_2) = \frac{1}{2}$$

$$P(B_3|B_1B_2) = \frac{1}{4} \qquad P(B_3|B_1^cB_2) = \frac{3}{4}$$

$$P(B_3|B_1B_2^c) = \frac{1}{5} \qquad P(B_3|B_1^cB_2^c) = \frac{4}{5}$$

Calcular a probabilidade de que passe corrente entre os terminais C e D.

<u>Solução</u>: A probabilidade buscada é

$$P[(B_1 \cap B_2) \cup (B_1 \cap B_3)] = P(B_1 \cap B_2) + P(B_1 \cap B_3) - P(B_1B_2B_3) =$$

$$= P(B_1)P(B_2) + P(B_1B_2B_3) + P(B_1B_2^cB_3) - P(B_1B_2B_3) =$$

$$= P(B_1)P(B_2) + P(B_1)P(B_2^c)\ P(B_3|B_1B_2^c) =$$

$$= \frac{1}{2}\,\frac{1}{2} + \frac{1}{2}\,\frac{1}{2}\,\frac{1}{5} = \frac{5+1}{20} = \frac{6}{20} = \frac{3}{10}.$$

EXERCÍCIOS

1. Se $\{A_1,\ldots,A_n\}$ são tais que $P(\bigcap_{i=1}^{n} B_i) = \prod_{i=1}^{n} P(B_i)$ onde B_i é igual a A_i o A_i^c para todo i, então os A_i são independentes.

2. a) Se A é independente de si mesmo, então $P(A)=0$ ou 1.
   b) Se $P(A) = 0$ ou 1, então A é independente de qualquer outro evento B.

3. Jogue um dado duas vêzes. Calcule a probabilidade condicional de obter 2 a primeira jogada, dado que a soma dos resultados foi 7.

4. Pedro quer enviar uma carta a Marina. A probabilidade de que Pedro escreve a carta é 0.80. A probabilidade de que o correio não a perde é 0.9. A probabilidade de que o carteiro entregue é 0.9. Dado que Marina não recebeu a carta, qual é a probabilidade condicional de que Pedro não a tenha escrito.

5. No espaço do Exemplo 4.6.3 calcular:
   a) A probabilidade condicional de B_2 dado que há cor-

rente entre as terminais C e D.

b) São B_1, B_2, B_3 independentes?

6. Durante o mês de Novembro a probabilidade de chuva é 0.3. Fluminense ganha um jògo em um dia com chuva com probabilidade 0.4, em um dia sem chuva com probabilidade 0.6. Se ganhou um jôgo em Novembro, qual é a probabilidade de que êsse dia choveu?

7. Num exame há 3 resposta para cada pergunta. Portanto um aluno tem probabilidade $\frac{1}{3}$ de escolher corretamente se êle está adivinhando e 1 se sabe a resposta. Suponhamos que um bom estudante sabe 95% das respostas e um mal estudante só 30%. Se um mal estudante tem a resposta correta, qual é a probabilidade de que advinhou? E para um bom estudante?

8. Se em lugar das famílias finitas de conjuntos $\{T_i\}$ e $\{S_i\}$ na Proposição 4.6.1 teríamos famílias numeráveis $\{T_i\}_{i=0,1,\ldots}$ e $\{S_i\}_{i=0,1,2,\ldots}$ um poderia, usando o mesmo procedimento, definir uma função p sôbre

$$\Omega = T_o + (S_o \times T_1) + (S_o \times S_1 \times T_2) + \ldots$$

Dê um exemplo no qual p não induza uma probabilidade sôbre Ω; e dizer $\sum_{w\in\Omega} p(\omega) < 1$. Sugestão: $S_i = \{0\}$, $T_i = \{1\}$ $\mu\{0\} = 1$, $\{\alpha_i\}_{i=1,2,\ldots}$ é uma sucessão de números >0

tal que $\sum_{i=1}^{\infty} \alpha_i < \infty$, $\beta_n = e^{-\alpha_n}$ e $p_{\underbrace{0,\ldots,0}_{n}}(0) = \beta_n$, $p_{\underbrace{0,\ldots,0}_{n}}(1) = 1 - \beta_n$.

Note que é possível agregar um ponto a Ω e converter ês te nôvo conjunto num espaço de probabilidades que represen ta ao experimento naturalmente.

9. Provar a Proposição 4.1.1.

10. Provar a Proposição 4.6.1 em geral.

11. Consideremos o seguinte espaço de probabilidades $\Omega = \{1,2,\ldots\}$. Se $C_\gamma = \sum_{k=1}^{\infty} \frac{1}{k^{1+\gamma}}$ $\gamma > 0$

$p_\gamma(k) = \frac{1}{C_\gamma} \frac{1}{k^{1+\gamma}}$.

Seja $\{p_n\}_{n=1,2,\ldots}$ a sucessão dos números primos $p_1 = 2$, $p_2 = 3$, $p_3 = 5,\ldots$

Seja $A_j = \{x\colon x$ é divisível por $p_j\}$

a) $A_1, A_2, \ldots$ são independentes, (Ajuda: Usar a 3) e a seguinte observação: se $A = \{x\colon x = js,\ j=1,2,\ldots,s$ inteiro positivo$\}$ então $P(A) = \frac{1}{s^{1+\gamma}}$.

b) Provar a seguinte fórmula (Euler)

$$C_\gamma = \sum_{k=1}^{\infty} \frac{1}{k^{1+\gamma}} = \prod_{j=1}^{\infty} \left(1 - \frac{1}{p_j^{1+\gamma}}\right)^{-1}$$

(Sugestão: $\{1\} = A_1^c \cap A_2^c \cap \ldots$)

12. Há duas urnas A e B. Urna A contém uma bolinha

branca e outra preta. Urna B contém duas brancas e três prestas. Uma bolinha é escolhida ao acaso da Urna A e introduzida na Urna B. Uma bolinha é extraída então da urna B.

a) Qual é a probabilidade de que ambas bolinhas sejam da mesma côr.

b) Qual é a probabilidade de que a primeira bolinha foi preta, dado que a segunda foi branca.

CAPÍTULO 5

VARIÁVEIS ALEATÓRIAS

5.1 - Introdução

Quando se joga uma moeda um certo número de vêzes, pode-se estar interessado, não na particular sucessão obtida, mas, sòmente, no número de caras. Do mesmo modo, se se joga uma moeda até obter uma cara, pode-se estar interessado ùnicamente no número de jogadas necessário para obtê-la. Estas são características numéricas associadas com o experimento. Se Ω é um espaço amostral definimos

DEFINIÇÃO 5.1.1 - Chamaremos <u>variável aleatória</u> a tôda função real definida sôbre Ω.

As variáveis aleatórias serão denotadas usualmente com as letras X, Y, Z, etc. Se X é uma variável aleatória, como Ω é enumerável, $X(\Omega)$ também é enumerável; digamos $X(\Omega) = \{x_1, x_2, \ldots\}$. Os números x_i são os <u>valores da variável aleatória</u>.

Seja $\mathbb{R}$ a reta real.

DEFINIÇÃO 5.1.2 - A função $A \mapsto P[X \in A]$ definida sôbre $\mathbb{P}(R)$ com valores em $[0,1]$ é chamada

<u>distribuição da variável aleatória X.</u>

É claro que os valores desta função estão determinados só pelos valores $P[X=x_i$, $i=1,2,\ldots$ já que

$$P[X\in A] = \sum_{i=1}^{\infty} I_A(x_i)P[X=x_i]$$

É também fácil ver que a distribuição de uma variável aleatória é uma probabilidade sôbre $\mathbb{P}(\mathbb{R})$.

Vamos denotar a distribuição da variável aleatória X por $\mathcal{L}(X)$ (lei de X).

DEFINIÇÃO 5.1.3 - A função $F_X: R \to [0,1]$ definida por $F_X(x) = P[X \in (-\infty,x]] = P[X\leq x]$ é chamada da <u>função de distribuição</u> da variável aleatória X.

Note que $F_X(x) = \mathcal{L}(X)((-\infty,x])$.

As propriedades mais importantes das funções de distribuição estão contidas na seguinte

PROPOSIÇÃO 5.1.1 - Se F_X é a função de distribuição de uma variável aleatória então

1. $\lim\limits_{x\to -\infty} F_X(x) = 0 \qquad \lim\limits_{x\to +\infty} F_X(x) = 1$
2. F_X é não decrescente.
3. F_X é contínua à direita.

<u>Demonstração</u>: Seja $S_x = (-\infty,x]$. Como

$$\bigcap_{x\in\mathbb{R}} S_x = \emptyset, \quad \bigcup_{x\in\mathbb{R}} S_x = \mathbb{R}, \quad \bigcap_{y>x} S_y = S_x$$

1 e 3 seguem da continuidade da probabilidade. 2 se deduz de monotonia da probabilidade.

A função de distribuição, por ser uma função real de variável real, é em princípio mais simples de manejar que a distribuição $\mathcal{L}(X)$. Mas temos o seguinte resultado.

PROPOSIÇÃO 5.1.2 - F_X determina $\mathcal{L}(X)$.

<u>Demonstração</u>: Seja $X(\Omega) = \{x_1, x_2, \ldots\}$. Então

$$P[X=x_i] = F_X(x_i) - \lim_{\substack{y \uparrow x_i \\ y < x_i}} F_X(y) .$$

Agora, $\mathcal{L}(X)$ fica determinada pela fórmula

$$\mathcal{L}(X)(A) = \sum_{i=1}^{\infty} I_A(x_i) P[X=x_i] .$$

É conveniente ter uma noção de função de distribuição independente da noção de variável aleatória. Vamos chamar então <u>função de distribuição</u> a qualquer função que satisfaz as condições 1, 2 e 3 da Proposição 5.1.1. É claro que agora nem tôda função de distribuição é a função de distribuição de uma variável aleatória definida num espaço de probabilidades discreto.

5.2 - Exemplos

No Capítulo 3 já vimos as distribuições de Poisson e hipergeométrica. Vamos agora dar alguns outros exemplos de distribuições, descrevendo ao mesmo tempo variáveis aleatórias dos quais provêm.

5.2.1. <u>Distribuição binomial</u>. Sejam $\Omega = \{0,1\}^{I_n}$ e $P(x_1,\ldots,x_n) = $
$= p^{\sum_{i=1}^{n} x_i} (1-p)^{n-\sum_{i=1}^{n} x_i}$. Seja X a variável aleatória definida por $X(x_1,\ldots,x_n) = \sum_{i=1}^{n} x_i$ (o número de caras na sucessão). Os possíveis valores de X são $0,1,\ldots,n$.

A sucessão $\{P[X=k]\}_{k=0,1,\ldots,n}$ é chamada distribuição binomial. Vamos calcular agora $P[X=k]$. Seja

$$A_k = \{(x_1,\ldots,x_n): \sum_{i=1}^{n} x_i = k\}$$

$$P[X=k] = P(A_k) = \#(A_k)\ p^k(1-p)^{n-k} =$$
$$= \binom{n}{k} p^k(1-p)^{n-k} .$$

Como claramente P depende de n e de p, vamos denotar êstes números por

$$b_{n,p}(k) = \binom{n}{k} p^k(1-p)^{n-k} = \binom{n}{k} p^k q^{n-k},\ k=0,1,\ldots,n,\ q=1-p.$$

(Nota: valores de $b_{n,p}(k)$ estão contidos na Tabela 3).

Vamos agora estudar o comportamento desta distribuição.

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

$$= \frac{(n+1)p-kp}{k(1-p)} = \frac{(n+1)p-k+kq}{kq} = 1 + \frac{(n+1)p-k}{kq} \qquad (5.2.1)$$

Disto resulta que se $(n+1)p$ nâo é inteiro, há só um máximo e êsse ocorre para k igual ao maior inteiro menor que $(n+1)p$. Se $(n+1)p$ é um inteiro, há dois máximos nos pontos $(n+1)p$ e $(n+1)p-1$. A função cresce até o ponto em que o máximo é alcançado e depois decresce. A seguinte figura dá uma idéia do que acontece. (Note que sempre há um máximo no ponto $[(n+1)p]$).

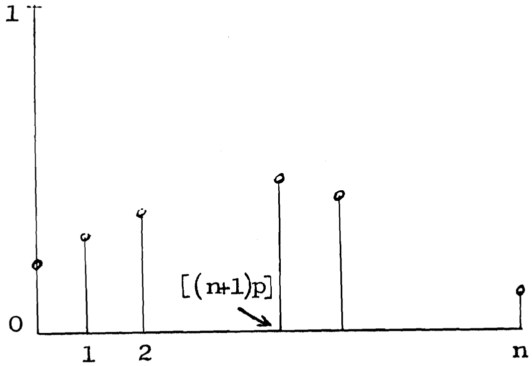


Figura 5.2.1

Tomemos agora $\ell > [(n+1)p]$.

$$P[S_n \geq \ell] = \sum_{k=\ell}^{n} b_{n,p}(k)$$

Pela fórmula (5.2.1) resulta que

$$b_{n,p}(\ell+j) \leq b_{n,p}(\ell) \left[1+\frac{(n+1)p-\ell}{\ell q}\right]^j$$

Portanto

$$\sum_{k=\ell}^{n} b_{n,p}(k) \leq \sum_{j=0}^{\infty} b_{n,p}(\ell) \left[1+\frac{(n+1)p-\ell}{\ell q}\right]^j =$$

$$= b_{n,p}(\ell) \frac{1}{1-[1+\frac{(n+1)p-\ell}{\ell q}]} = b_{n,p}(\ell) \frac{\ell q}{\ell-(n+1)p} .$$

Como também

$$b_{n,p}(\ell) \leq \frac{\sum_{k=[(n+1)p]}^{\ell} b_{n,p}(k)}{\ell-[(n+1)p]+1} \leq \frac{1}{\ell-(n+1)p}$$

e portanto $P[S_n \geq \ell] \leq \dfrac{\ell q}{[\ell-(n+1)p]^2}$ (5.2.2)

5.5.2. Distribuição de Poisson e distribuição binomial.

Seja $\{b_{n,p}(k)\}_{k=0,1,\ldots,n}$ a distribuição binomial com parâmetros n e p e $\{p_\lambda(k)\}_{k=0,1,\ldots}$ a distribuição de Poisson com parâmetro $\lambda > 0$. Temos o seguinte resultado

PROPOSIÇÃO 5.2.1 - Se $n \to \infty$ e $np \to \lambda > 0$ (portanto $p \to 0$) então $b_{n,p}(k) \to p_\lambda(k)$, $k=0,1,\ldots$

Demonstração: Seja $\lambda' = np$.

$$\frac{n(n-1)\ldots(n-k+1)}{k!} \left(\frac{\lambda'}{n}\right)^k \left(1-\frac{\lambda'}{n}\right)^{n-k} =$$

$$= \frac{n(n-1)\ldots(n-k+1)}{n^k} \frac{(\lambda')^k}{k!} \left(1-\frac{\lambda'}{n}\right)^n \left(1-\frac{\lambda'}{n}\right)^{-k} =$$

$$= 1\left(1-\frac{1}{n}\right) \ldots \left(1-\frac{(k-1)}{n}\right) \frac{1}{\left(1-\frac{\lambda'}{n}\right)^k} \frac{(\lambda')^k}{k!} \left(1-\frac{\lambda'}{n}\right)^n .$$

Se agora $n \to \infty$, de forma tal que $\lambda' = np \to \lambda > 0$, temos que o membro direito converge a $\frac{\lambda^k}{k!} e^{-\lambda}$, que é igual a $p_\lambda(k)$.

(Nota: Se $x_n \to x$ então $\lim_{n\to\infty} (1 + \frac{x_n}{n})^n = e^x$).

NOTA: Um resu tado mais forte que o co tido na Proposição 5.2.1 vai ser provado no Capítulo 7. (Ver Proposição 7.7.3).

5.2.3. Distribuição de Pascal. Seja $0 < p < 1$, r um inteiro positivo e consideremos o experimento consistente em jogar uma moeda até observar r caras. O espaço de probabilidades para êste experimento é

$$\Omega = \{ x_1,\ldots,x_n): x_i = 0 \text{ ou } 1,\ x_n = 1,\ \sum_{i=1}^{n} x_i = r\}$$

$$P((x_1,\ldots,x_n)) = p^r(1-p)^{n-r} .$$

Antes de verificar que (Ω, P) é um espaço de probabilidades vamos definir uma variável aleatória $X:\Omega \to R$ da seguinte forma

$$X\ ((x_1,\ldots,x_n)) = n$$

ou seja X indica o tempo no qual observamos a r-ésima cara. Os valores de X são $\{r, r+1,\ldots\}$ e

$$P[X=n] = \binom{n-1}{r-1} p^r q^{n-r} \quad n=r,r+1,\ldots,\ q=1-p$$

Verificar que $\sum_{n=r}^{\infty} P[X=n]$ é equivalente a verificar que P é uma probabilidade. Temos

$$\sum_{n=r}^{\infty} P[X=n] = \sum_{n=r}^{\infty} \binom{n-1}{r-1} p^r q^{n-r} =$$

$$= p^r \sum_{i=0}^{\infty} \binom{r+i-1}{r-1} q^i = \frac{p^r}{(r-1)!} \sum_{i=0}^{\infty} \frac{(r+i-1)!}{i!} q^i =$$

$$= \frac{p^r}{(r-1)!} \frac{(r-1)!}{(1-q)^r} = 1 .$$

A segunda igualdade se obtém definindo $i = n-r$, a quarta pelo exercício 2 do Anexo 1.

A distribuição da variável aleatória X chama-se <u>distribuição de Pascal</u>. Quando $r=1$, a distribuição obtida chama-se <u>distribuição geométrica</u>.

5.2.4. Distribuição uniforme. Esta é a distribuição concentrada nos inteiros $\{0,1,\ldots,n\}$ com probabilidade $\frac{1}{n+1}$ para cada um dêles.

5.3 - Variáveis Aleatórias Simples

DEFINIÇÃO 5.3.1 - Uma variável aleatória X é chamada simples se e só se existem conjuntos disjuntos $\{A_i\}_{i=1,\ldots,n}$ tais que $\sum_{i=1}^{n} A_i = \Omega$ e números $\{a_i\}_{i=1,\ldots,n}$ tais que $X = \sum_{i=1}^{n} a_i I_{A_i}$.

Nota: a representação de X não é única.

PROPOSIÇÃO 5.3.1 - Seja $X \geq 0$ uma variável aleatória.

Existe uma sucessão de variáveis aleatórias simples $\{X_n\}_{n=1,2,\ldots}$ tal que $0 \leq X_n \uparrow X$.

Demonstração: Tomar

$$X_n = \sum_{i=1}^{n} x_i \, I_{[X=x_i]} \quad \text{onde} \quad X(\Omega) = \{x_1, x_2, \ldots\}$$

Nota: também

$$Y_n = \sum_{k=0}^{n2^n} \frac{k}{2^n} I_{[\frac{k}{2^n} \leq X < \frac{k+1}{2^n}]} + n \, I_{[X>n]}$$

tem a propriedade $0 \leq Y_n \uparrow X$.

5.4 - Variáveis Aleatórias Independentes

Sejam $X_1, X_2, \ldots, X_n$ n variáveis aleatórias sôbre (Ω, P).

DEFINIÇÃO 5.4.1 - As variáveis aleatórias $X_1, \ldots, X_n$ dizem-se <u>independentes</u> se e só se

$$\forall \; A_1, A_2, \ldots, A_n \qquad A_i \subseteq \mathbb{R} \qquad i=1,2,\ldots,n$$

$$P\left(\bigcap_{i=1}^{n} [X_i \in A_i]\right) = \prod_{i=1}^{n} P[X_i \in A_i]$$

Sejam agora $\{x_{i1}, x_{i2}, \ldots\}$ os valores de variável aleatória X_i, $i=1,2,\ldots,n$.

PROPOSIÇÃO 5.4.1 - $X_1, X_2, \ldots, X_n$ são independentes se e só se

$$\forall j_1, j_2, \ldots, j_n \quad P\left(\bigcap_{i=1}^{n} [X_i = x_{i j_i}]\right) = \prod_{i=1}^{n} P[X_i = x_{i j_i}] \; .$$

<u>Demonstração</u>: Ver exercício 1.

O leitor deve convencer-se por si mesmo de que esta definição corresponde a sua intuição verificando em exemplos concretos a independência de diferentes variáveis aleatórias. Nós vamos dar um exemplo de variáveis aleatórias independentes um pouco mais complicado.

EXEMPLO 5.4.1 - No espaço da seção 5.2.3, definamos

$$\tau_i((x_1, \ldots, x_n)) = k \qquad \text{se} \qquad x_k = 1 \quad \text{e}$$
$$\sum_{j=1}^{k} x_j = i \qquad\qquad i=1, \ldots, r$$

τ_i indica o tempo no qual se observa a i-ésima cara. Daí $\tau_r = X$. Sejam

$$X_1 = \tau_1 \quad \text{e} \quad X_i = \tau_i - \tau_{i-1} \qquad i=2, \ldots, r$$

PROPOSIÇÃO 5.4.2 - $X_1, \ldots, X_r$ são independentes e têm a mesma distribuição (ou, como se diz usualmente, são <u>idênticamente distribuidas</u>). A distribuição de X_i é geométrica com parâmetro p.

Demonstração:

$$P[X_i=s] = \sum_{k_1,k_2,\dots,k_{i-1}} P[X_1=k_1,\dots,X_{i-1}=k_{i-1},\ X_i=s] =$$

$$= \sum_{k_1,k_2,\dots,k_{i-1}} P[\tau_1=k_1,\dots,\tau_{i-1}=\sum_{j=1}^{i-1} k_j,\ \tau_i = \sum_{j=1}^{i-1} k_j+s] =$$

$$= \sum_{k_1,k_2,\dots,k_{i-1}} p^i\, q^{\sum_{j=1}^{i-1} k_j+s-i} = p^i\, q^{s-i}\Big(\sum_{k_1,k_2,\dots,k_{i-1}} q^{\sum_{j=1}^{i-1} k_j}\Big)=$$

$$= p^i\, q^{s-i}\Big(\sum_{\ell=1}^{\infty} q^{\ell}\Big)^{i-1} = p^i\, q^{s-i}\, \frac{q^{i-1}}{p^{i-1}} = p\, q^{s-1} .$$

Mas, $p\,q^{s-1} = P[X_1=s]$ e portanto as variáveis aleatórias $X_1,\dots,X_r$ são idênticamente distribuídas com lei comum geométrica com parâmetro p.

Agora

$$P[X_1=s_1,\dots,X_r=s_r] = P[\tau_1=s_1,\dots,\tau_r=s_1+s_2+\dots+s_r] =$$

$$= p^r\, q^{\sum_{i=1}^{r} s_i-r}$$

e

$$\prod_{i=1}^{r} P[X_i=s_i] = \prod_{i=1}^{r} P[X_1=s_i] = \prod_{i=1}^{r} p\, q^{s_i-1} =$$

$$= p^r\, q^{\sum_{i=1}^{r}(s_i-1)} = p^r\, q^{\sum_{i=1}^{r} s_i-r} .$$

Portanto $P[X_1=s_1,\dots,X_r=s_r] = \prod_{i=1}^{r} p[X_i=s_i]$ e as variáveis aleatórias $X_1,\dots,X_r$ são independentes.

Para terminar esta seção enunciamos a seguinte proposição cuja prova elementar deixamos como exercício (Ver exercício 2).

PROPOSIÇÃO 5.4.3 - Sejam $X_1, X_2, \ldots, X_n$ n variáveis aleatórias independentes, e $f_1, f_2, \ldots, f_n$ n funções de $\mathbb{R}$ em $\mathbb{R}$. Então $f_1(X_1), f_2(X_2), \ldots, f_n(X_n)$ são variáveis aleatórias independentes.

5.5 - Convergência de variáveis aleatórias

DEFINIÇÃO 5.5.1 - Seja $\{X_n\}_{n=1,2,\ldots}$ uma sucessão de variáveis aleatórias e X outra variável aleatória. Dizemos que X_n <u>converge em probabilidade a</u> X (e escrevemos $X_n \xrightarrow{P} X$) se e sòmente se

$$\forall \epsilon > 0 \qquad P[|X_n - X| \geq \epsilon] \underset{n\to\infty}{\to} 0$$

DEFINIÇÃO 5.5.2 - Dizemos que uma propriedade π vale quase certamente se

$$P\{\omega : \omega \text{ não tem a propriedade } \pi\} = 0$$

Introduzimos finalmente uma definição que dá uma noção mais forte de convergência que a descrita na Definição 5.5.1 mas que para espaços de probabilidade discretos é equivalente. (Ver exercício 6).

DEFINIÇÃO - Dizemos que a sucessão X_n <u>converge quase certamente a</u> X se e sòmente se

$$P\{\omega: X_n(\omega) \not\to X(\omega)\} = 0$$

(o símbolo $\not\to$ indica que a sucessão $X_n(\omega)$ não converge a $X(\omega)$).

EXERCÍCIOS

1. Provar a Proposição 5.4.1.
2. Provar a Proposição 5.4.3.
3. Provar que o conjunto das variáveis aleatórias simples é um espaço vetorial. Provar também que se X e Y são variáveis aleatórias simples então XY, $X \vee Y$ e $X \wedge Y$ são variáveis aleatórias simples. Isto equivale a dizer que o conjunto das variáveis aleatórias simples é uma álgebra reticulada.
4. <u>Falta de desgaste da distribuição geométrica</u>.

Seja P uma probabilidade geométrica (com parâmetro p, $0 < p < 1$) sôbre o conjunto $\Omega = \{1,2,\ldots\}$ ($P(k) = p\, q^{k-1}$).
Seja $B_r = \{x: x \in \Omega \quad x > r\}$.
Provar que $P(B_{r+s}/B_r) = P(B_s) \quad r = 1,2,\ldots \quad s = 1,2,\ldots$
Resulta então que $P(B_{r+1}) = P(B_r)\, P(B_1) \quad r = 1,2,\ldots$
Provar que se Q é outra probabilidade sôbre o mesmo es-

paço Ω que satisfaz esta última equação então Q é uma probabilidade geométrica com parâmetro $Q(\{1\})$.

5. <u>Desgaste da distribuição de Poisson</u>.

Provar que se P é uma distribuição de Poisson sôbre $\Omega = \{0,1,\ldots\}$ então $P[\{k\}|\{x:x \geq k\}] \to 1$ quando $k \to \infty$.

6. Provar que nos espaços de probabilidades que usamos neste livro (espaços discretos) $X_n \xrightarrow{P} X$ se e sòmente se $X_n \to X$ quase certamente.

7. Provar que $A_1, A_2, \ldots, A_n$ são independentes se e sòmente se $I_{A_1}, \ldots, I_{A_n}$ são variáveis aleatórias independentes.

8. Seja $\{x_1, x_2, \ldots\} \subseteq \mathbb{R}$ e $\{p_i\}_{i=1,2,\ldots}$ uma sucessão tal que $p_i \geq 0$ e $\sum_{i=1}^{\infty} p_i = 1$. Construir um espaço (Ω, P) e uma variável aleatória X sôbre Ω tal que $P[X=x_i] = p_i$.

9. Se F é a função de distribuição de uma variável aleatória X com valores $x_1, x_2, \ldots$ Provar que se $x \neq x_i$ então F é contínua em x.

10. (Sôbre a distribuição de Poisson e a distribuição binomial). $(\lambda = np)$

a) Se

$$h(k) = \begin{cases} b_{n,p}(k)/p_{np}(k) & 0 \leq k \leq n \\ 0 & k>n \end{cases}$$

então h cresce, atinge um máximo no ponto $k = [np] + 1$ primeiro e depois decresce.

b) Provar que

$$\frac{\lambda^k}{k!}\left(1-\frac{k}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^n \le b_{n,p}(k) \le \frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{n-k}$$

c) Usando $e^{-\frac{t}{1-t}} \le 1-t \le e^{-t} \qquad 0 < t < 1$

provar que

$$p_\lambda(k)e^{-\frac{k^2}{n-k}-\frac{\lambda^2}{n-\lambda}} < b_{n,p}(k) < p_\lambda(k)\, e^{-\frac{k\lambda}{n}} \le e^{\lambda\left(p+\frac{1}{n}\right)} .$$

CAPÍTULO 6

ESPERANÇA DE VARIÁVEIS ALEATÓRIAS REAIS

6.1 - Definição e propriedades básicas

Seja $X = \sum_{i=1}^{n} a_i I_{A_i}$ uma variável aleatória simples.

DEFINIÇÃO 6.1.1 - O número $EX = \sum_{i=1}^{n} a_i P(A_i)$ é chamado <u>esperança</u> da variável aleatória X.

Êste número será também denotado por $E(X) = \int X(w)P(dw) = \int XdP = \int X$.

PROPOSIÇÃO 6.1.1 - EX não depende da particular representação de X.

<u>Demonstração</u>: Se $X = \sum_{i=1}^{n} a_i I_{A_i} = \sum_{j=1}^{m} b_j I_{B_j}$, então

$$\sum_{ij} a_i I_{A_i \cap B_j} = \sum_{ij} b_j I_{A_i \cap B_j} \quad \text{e portanto}$$

$\sum_{ij} a_i P(A_i \cap B_j) = \sum_{ij} b_j P(A_i \cap B_j)$, donde resulta imediatamente que $\sum_i a_i P(A_i) = \sum_j b_j P(B_j)$.

PROPOSIÇÃO 6.1.2 - a) Se $X \geq 0$, então $EX \geq 0$

b) $E(aX+bY) = a\,EX + b\,EY$, onde a e b são números reais.

c) $E(I_A) = P(A)$.

Demonstração: a) e c) são imediatas.

Para provar b), escrevamos:

$$X = \sum_{i=1}^{n} a_i I_{A_i} \quad , \quad Y = \sum_{j=1}^{m} b_j I_{B_j}$$

Então

$$aX+bY = \sum_{i=1}^{n} \sum_{j=1}^{m} (a\, a_i + b\, b_j) I_{A_i B_j}$$

e

$$E(aX+bY) = \sum_{i=1}^{n} \sum_{j=1}^{m} (a\, a_i + b\, b_j)\, P(A_i B_j) =$$

$$= \sum_{i=1}^{n} a a_i \sum_{j=1}^{m} P(A_i B_j) + \sum_{j=1}^{m} b b_j \sum_{i=1}^{n} P(A_i B_j) =$$

$$= a \sum_{i=1}^{n} a_i\, P(A_i) + b \sum_{j=1}^{m} b_j\, P(B_j) = a\, EX + b\, EY\ .$$

Como $EX \geq 0$ se $X \geq 0$, temos $EX_1 \leq EX_2$ se $X_2 \geq X_1$.

Se A é um evento, $EXI_A = \int_\Omega I_A X dP$ será denotado por $\int_A X dP$.

LEMA 6.1.1 - Se $\{X_n\}_{n=1,2,\ldots}$ é uma sucessão de funções simples e $X_n \downarrow 0$, então $E\, X_n \downarrow 0$.

Demonstração: Dado $\varepsilon > 0$. Para cada $w \in \Omega$, definimos $N(\omega) = \inf\{n : X_n(\omega) < \frac{\varepsilon}{2}\}$. Determinamos $a > 0$ tal que $P[N>a] < \dfrac{\varepsilon}{2 \sup(X_1)}$. Seja agora $n \geq a$. Temos

$EX_n = \int_{[N\leq a]} X_n \, dP + \int_{[N>a]} X_n \, dP$. Mas, sôbre $[N\leq a]$, X_n é menos que $\frac{\varepsilon}{2}$ e portanto

$$EX_n \leq \frac{\varepsilon}{2} P[N\leq a] + \sup(X_1) \quad P[N>a]$$

$$\leq \frac{\varepsilon}{2} + \frac{\sup(X_1)\varepsilon}{2 \sup(X_1)} = \varepsilon$$

É dizer $EX_n < \varepsilon$ se $n \geq a$ e portanto $\lim\limits_{n\to\infty} EX_n = 0$.

Seja $X \geq 0$ e $0 \leq X_n \uparrow X$ uma sucessão de funções simples que convergem a X.

DEFINIÇÃO 6.1.2 - $EX = \lim\limits_{n\to\infty} EX_n$.

PROPOSIÇÃO 6.1.3 - Se $0 \leq Y_n \uparrow X$, então $\lim\limits_{n} EX_n = \lim\limits_{n} EY_n$ (ou seja, EX está bem definida; não depende da sucessão escolhida).

Demonstração: Para todo n, temos

$$0 \leq X_n - X_n \wedge Y_m \downarrow 0 \quad (\text{quando } m \to \infty).$$

Pel Proposição 6.1.2 e o Lema 6.1.1:

$$EX_n - E\ (X_n \wedge Y_m) \downarrow 0, \text{ isto é, } 0 \leq E(X_n \wedge Y_m) \uparrow EX_n$$

e portanto, $\forall n$, $EX_n = \lim\limits_{m\to\infty} (EX_n \wedge Y_m) \leq \lim\limits_{m\to\infty} EY_m$

donde $\lim\limits_{n\to\infty} EX_n \leq \lim\limits_{m\to\infty} EY_m$.

Como os papéis de X_n e Y_m podem ser trocados, temos também $\lim_{m\to\infty} EY_m \leq \lim_{n\to\infty} EX_n$ e, portanto, a igualdade desejada.

PROPOSIÇÃO 6.1.4

a) Se $X \geq 0$, então $0 \leq EX \leq +\infty$

b) Se $0 \leq X_1 \leq X_2$, então $EX_1 \leq EX_2$

c) Se $X \geq 0$, $c \geq 0$, então $E(cX) = cEX$

d) Se $X_1 \geq 0$, $X_2 \geq 0$, então $E(X_1+X_2) = EX_1 + EX_2$

Demonstração: Vamos provar só d). Sejam $0 \leq Y_n \uparrow X_1$, $0 \leq Z_n \uparrow X_2$. Então $0 \leq Y_n + Z_n \uparrow X_1 + X_2$ e portanto $E(X_1+X_2) = \lim_n E(Y_n+Z_n) = \lim_n (EY_n + EZ_n) =$ $= \lim_n EY_n + \lim_n EZ_n = EX_1 + EX_2$.

A segunda igualdade segue da linearidade de E sôbre o conjunto das variáveis aleatórias simples (Proposição 6.1.2).

Seja agora X uma variável aleatória qualquer.

DEFINIÇÃO 6.1.3 - Se $EX^+ < \infty$ e $EX^- < \infty$, diz-se que X é integrável e se define $EX = EX^+ - EX^-$.

PROPOSIÇÃO 6.1.5 - a) $\forall c$, $E(cX) = cEX$

b) Se X e Y são integráveis, então $X+Y$ é integrável e $E(X+Y) = EX + EY$.

c) Se $X \leq Y$, então $EX \leq EY$.

<u>Demonstração</u>: a) é imediata. Para provar b), observamos primeiro que, se X_1 e X_2 são ≥ 0 e integráveis, então a variável $X = X_1 - X_2$ é integrável e $EX = EX_1 - EX_2$. De fato, porque $X^+ \leq X_1$ e $X^- \leq X_2$, X é integrável e, como $X = X^+ - X^- = X_1 - X_2$ temos $X^+ + X_2 = X_1 + X^-$ donde $EX^+ + EX_2 = EX_1 + EX^-$ (Proposição 6.1.4 d)), isto é, $EX = EX^+ - EX^- = EX_1 - EX_2$.

Agora $X+Y = (X^+-X^-) + (Y^+-Y^-) = X^++Y^+-(X^-+Y^-)$. Usando a observação anterior, $X+Y$ é integrável e $E(X+Y) = E(X^++Y^+) - E(X^-+Y^-) = EX^+ - EX^- + EY^+ - EY^- =$ $= EX + EY$.

Seja agora X uma variável aleatória ≥ 0 e sejam $x_1, x_2, \ldots$ os valores de X (em alguma ordem). Se definimos $X_n = \sum_{i=1}^{n} x_i \, I_{[X=x_i]}$, resulta que $0 \leq X_n \uparrow X$ e, portanto, $EX = \lim_n EX_n = \lim_{n\to\infty} \sum_{i=1}^{n} x_i \, P[X=x_i] = \sum_{i=1}^{\infty} x_i \, P[X=x_i]$

Se X é agora uma variável aleatória qualquer com valores $x_1, x_2, \ldots$, temos o seguinte resultado que é uma simples generalização desta observação e que deixamos como exercício. (Ver exercício 1).

PROPOSIÇÃO 6.1.6 - X é integrável se e só se

$$\sum_{i=1}^{\infty} |x_i| P[X=x_i] < \infty \text{ e nesse caso } EX = \sum_{i=1}^{\infty} x_i \, P[X=x_i].$$

Vamos agora calcular expectâncias de algumas distribuições conhecidas. As palavras expectância, média e primeiro momento serão usadas como sinônimos de esperança.

6.2 - Exemplos

6.2.1. Distribuição Binomial. Temos que

$$\sum_{k=0}^{n} \binom{n}{k} p^k\, q^{n-k}\, k = \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k\, q^{n-k} =$$

$$= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!}\, p^{k-1}\, q^{n-k} = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j} = np.$$

Isto é, a esperança de uma variável aleatória binomial é np. Outra forma de calcular êste número é a seguinte. Se $X_1, X_2, \ldots, X_n$ são variáveis aleatórias independentes tais que $P[X_i=1] = p$ e $P[X_i=0] = 1-p = q$, $i=1,\ldots,n$, então é fácil ver que $\mathcal{L}(\sum_{i=1}^{n} X_i)$ é binomial com parâmetros n e p. Portanto, para calcular a esperança calculamos primeiro $EX_1 = 1p + 0q = p$. Como as X_i são idênticamente distribuidas $E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i) = nEX_1 = np$.

Esta observação de que $\mathcal{L}(\sum_{i=1}^{n} X_i) = B_{n,p}$ (binomial com parâmetros n e p) vai ser usada novamente no futuro e deve ser lembrada.

6.2.2. Distribuição de Pascal. Seja X a variável aleatória do Exemplo 5.4.1 e escrevemos X como $X = \sum_{i=1}^{r} X_i$ onde as X_i são independentes com distribuição geométrica de parâmetro p. Resulta então que $EX = \sum_{i=1}^{r} EX_i = r\,EX_1$. Agora

$$EX_1 = \sum_{k=1}^{\infty} kpq^{k-1} = \frac{p}{q}\sum_{k=1}^{\infty} kq^k = \frac{p}{q}\,\frac{q}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}\,.$$

(A segunda igualdade segue do primeiro exercício do Anexo 1). Portanto, se X tem uma distribuição de Pascal, $EX = \frac{r}{p}$.

6.2.3. Distribuição de Poisson. Temos que calcular

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

$$= e^{-\lambda}\lambda\sum_{j=0}^{\infty}\frac{\lambda^j}{j!} = e^{-\lambda}\lambda\, e^{\lambda} = \lambda\,.$$

6.2.4. (Ver Proposição 3.3.6)

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{k=1}^{n}(-1)^{k-1}\sum_{i\le i_1<i_2<\dots<i_k\le n} P(A_{i_1}\cap A_{i_2}\cap\dots\cap A_{i_k})$$

Aplicar esperanças a igualdade 1.3.4.

6.3 - Funções de Variáveis Aleatórias. Momentos.

Seja (Ω,P) um espaço de probabilidades, Ω' um conjunto enumerável, $T:\Omega\to\Omega'$ e $g:\Omega'\to\mathbb{R}$ duas funções. Para cada $A'\subseteq\Omega'$, definimos $P'(A) = P(T^{-1}(A'))$. Vê-se fàcilmente que P' é uma probabilidade sôbre Ω'. A relação entre as expectâncias de $g\circ T$ com respeito a P e de g' com respeito a P' está dada pela seguinte propo sição.

PROPOSIÇÃO 6.3.1 - Se $g \circ T$ ou g é integrável, então $\int_\Omega g \circ T \, dP = \int_{\Omega'} g dP'$ (a igualdade é sempre válida se $g \geq 0$).

Demonstração: Vamos sòmente esboçar a prova.

Primeiro, se $g = I_A$, a igualdade é trivial. Se g é simples e ≥ 0, a igualdade é consequência da linearidade da expectância. Para $g \geq 0$ qualquer, se toma $0 \leq g_n \uparrow g$, g_n variáveis aleatórias simples; como para todo n a igualdade é válida, o é também para o limite. Finalmente, para g qualquer, se decompõe g em g^+ e g^- e se utiliza o resultado do caso anterior.

Nas condições da Proposição 6.3.1, temos o seguinte corolário.

COROLÁRIO 6.3.1 - Se X é uma variável aleatória sôbre (Ω, P) com valores $\{x_1, x_o, \ldots\}$ e $g: \mathbb{R} \to \mathbb{R}$, então $Eg \circ X = \sum_{i=1}^{\infty} g(x_i) P[X=x_i]$.

$$(E|g \circ X| < \infty, \quad \text{ou} \quad \sum_{i=1}^{\infty} |g(x_i)| P[X=x_i] < \infty)$$

Uma classe de funções g que é de grande interêsse para nós é a que se obtém tomando $g(x) = |x|^r$, $r \geq 1$, real ou $g(x) = x^n$, $n = 1, 2, \ldots$

DEFINIÇÃO 6.3.1 - O número $E|X|^r$ se denomina <u>momento absoluto de ordem</u> r. O número EX^n, <u>momento de ordem</u> n.

Existem várias importantes propriedades dos momentos absolutos de uma variável aleatória. Algumas dessas propriedades estão contidas no Anexo 2. Nós sòmente vamos usar as duas proposições seguintes.

PROPOSIÇÃO 6.3.2 - Se $E|X|^r < \infty$, $r \geq 1$, então $E|X|^{r'} < \infty$, $\forall\, 1 \leq r' \leq r$.

<u>Demonstração</u>: $\forall x \in \mathbb{R}$, $|x|^{r'} \leq 1 + |x|^r$ e portanto $E|X|^{r'} \leq 1 + E|X|^r < \infty$.

PROPOSIÇÃO 6.3.3 - $E|X+Y|^r \leq 2^{r-1}\{E|X|^r + E|Y|^r\}$, para $r \geq 1$.

<u>Demonstração</u>: A função $x \to |x|^r$ é convexa e portanto $\left|\frac{x+y}{2}\right|^r \leq 2^{r-1}\,\{|x|^r + |y|^r\}$ e calculando agora as expectâncias de ambos os membros da desigualdade obtemos o resultado.

COROLÁRIO 6.3.2 - Se $E|X|^r < \infty$ e $E|Y|^r < \infty$ $(r \geq 1)$, então $E|X+Y|^r < \infty$.

Seja agora X uma variável aleatória tal que $EX^2 < \infty$ (e, pela Proposição 6.3.2, $E|X| < \infty$).

DEFINIÇÃO 6.3.2 - O número $E(X-EX)^2$ é chamado <u>variância</u> de X. Usaremos as notações σ_X^2 e var(X) para indicá-lo.

PROPOSIÇÃO 6.3.4 - Se $EX^2 < \infty$,

a) $\operatorname{var}(X+c) = \operatorname{var}(X), \forall c \in \mathbb{R}$

b) $\operatorname{var}(cX) = c^2 \operatorname{var}(X), \ \forall c \in \mathbb{R}$

c) $\operatorname{var}(X) = EX^2 - (EX)^2$

<u>Demonstração</u>: a) $\operatorname{var}(X+c) = E[X+c - E(X+c)]^2 =$

$$= E(X-EX)^2 = \operatorname{var}(X).$$

b) $\operatorname{var}(cX) = E(cX - EcX)^2 = E[c(X-EX)]^2 = c^2 E(X-EX)^2 =$

$$= c^2 \operatorname{var}(X).$$

c) $\operatorname{var}(X) = E(X-EX)^2 = E[X^2 \cdot 2(EX)X + (EX)^2] =$

$$= EX^2 - 2(EX)(EX) + (EX)^2 = EX^2 \quad (EX)^2$$

A seguinte proposição é muito importante.

PROPOSIÇÃO 6.3.5 - a) Se X e Y são duas variáveis aleatórias independentes ≥ 0, então $EXY = EXEY$.

b) Se X e Y são integráveis e independentes, então XY é integrável e $EXY = EXEY$.

<u>Demonstração</u>: a) Se $X = I_A$ e $Y = I_B$, então X e Y são independentes se e sòmente se A e B são independentes e portanto

$$EXY = EY_A X_B = EI_{AB} = P(A \cap B) = P(A)P(B) = E(I_A)\, E(I_B) .$$

Se X e Y são variáveis aleatórias simples ≥ 0 e independentes a fórmula decorre imediatamente da linearidade de E e da validade da igualdade para indicadores.

Se $X \geq 0$ e $Y \geq 0$ são independentes, se escolhem $0 \leq X_n \uparrow X$ e $0 \leq Y_n \uparrow Y$, simples, tais que X_n é independente de Y_m, $\forall (n,m)$ (as variáveis definidas na seção 5.3, têm esta propriedade).

Temos

$$EXY = \lim_n E(X_n Y_n) = \lim_n (EX_n)(EY_n) =$$
$$= (\lim_n EX_n)(\lim_n EY_n) = (EX)(EY).$$

b) Sejam agora X e Y arbitrárias, Escrevemos

$$X = X^+ - X^- \quad \text{e} \quad Y = Y^+ - Y^-$$

e portanto

$$XY = (X^+-X^-)(Y^+-Y^-) = (X^+Y^+ + X^-Y^-) - (X^+Y^- + X^-Y^+)$$

Como X^+, X^-, Y^+ e Y^- são integráveis e X^+Y^+, X^-Y^-, X^+Y^-, X^-Y^+ são pares racionais independentes, resulta pela primeira parte, que $X^+Y^+ + X^-Y^-$ e X^-Y^+ são integráveis. Portanto a diferença (quer dizer XY) é integrável e temos

$$EXY = E(X^+Y^+ + X^-Y^-) - E(X^+Y^- + X^-Y^+) =$$
$$= (EX^+)(EY^+) + (EX^-)(EY^-) - (EX^+)(EY^-) - (EX^-)(EY^+) =$$
$$= (EX^+)[EY^+-EY^-] - EX^-[EY^+-EY^-] =$$
$$= [EY^+-EY^-][EX^+-EX^-] = (EX)(EY)$$

Sejam agora X e Y duas variáveis aleatórias tais que $EX^2 < \infty$ e $EY^2 < \infty$.

COROLÁRIO 6.3.3 - Se X e Y são independentes, então

$$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y).$$

<u>Demonstração</u>: Pelo Corolário 6.3.1 $\text{var}(X+Y)$ é finita. Pela Proposição 6.3.4 c) temos

$$\text{var}(X+Y) = E(X+Y)^2 - (EX+EY)^2 = EX^2+EY^2+2EXY -$$
$$- (EX)^2 - (EY)^2 - 2(EX)(EY).$$

Como $EXY = EXEY$, a última expressão é igual a

$$EX^2 + EY^2 - (EX)^2 - (EY)^2 = \text{var}(X) + \text{var}(Y)$$

novamente pela Proposição 6.3.4 c).

Por indução, êste corolário é generalizado ao seguinte

COROLÁRIO 6.3.4 - Se $X_1,\ldots,X_n$ são variáveis aleatórias independentes tais que $E\,X_i^2 < \infty$, $i=1,2,\ldots,n$, então $\text{var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \text{var}(X_i)$.

Vamos agora calcular as variâncias de algumas distribuições conhecidas.

6.4 - Exemplos

EXEMPLO 6.4.1 (Distribuição de Poisson).

Calculamos primeiro

$$EX^2 = \sum_{k=0}^{\infty} k^2 \frac{e^{-\lambda}\lambda^k}{k!} = \lambda \sum_{k=1}^{\infty} k \frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!} =$$

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

$$= \lambda\{\sum_{j=0}^{\infty} j\frac{e^{-\lambda}\lambda^{j}}{j!} + \sum_{j=0}^{\infty}\frac{e^{-\lambda}\lambda^{j}}{j!}\} =$$

$$= \lambda\{\lambda + 1\} = \lambda^2 + \lambda.$$

Portanto

$$\text{var}(X) = EX^2 - (EX)^2 = \lambda^2+\lambda - \lambda^2 = \lambda$$

EXEMPLO 6.4.2 - (Distribuição geométrica)

$$EX^2 = \sum_{k=1}^{\infty} k^2\, p\, q^{k-1} = p \sum_{k=1}^{\infty} k^2(1-p)^{k-1}$$

Consideremos

$$\frac{d^2}{dp^2}(\sum_{k=1}^{\infty}(1-p)^{k+1}) = \sum_{k=1}^{\infty}(k+1)k(1-p)^{k-1} =$$

$$= \sum_{k=1}^{\infty} k^2(1-p)^{k-1} + \sum_{k=1}^{\infty} k(1-p)^{k-1}$$

Multiplicando por p temos

$$p\frac{d^2}{dp^2}(\sum_{k=1}^{\infty}(1-p)^{k+1}) = \sum_{k=1}^{\infty} k^2\, p(1-p)^{k-1} + \underbrace{\sum_{k=1}^{\infty} k\, p(1-p)^{k-1}}_{=\frac{1}{p}}$$

Portanto

$$EX^2 = \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} = p\underbrace{\frac{d^2}{dp^2}\left(\frac{(1-p)^2}{p}\right)}_{\frac{2}{p^3}} - \frac{1}{p} =$$

$$= p\frac{2}{p^3} - \frac{1}{p} = \frac{2}{p^2} - \frac{1}{p}.$$

Resulta então que

$$\text{var}(X) = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2} = \frac{q}{p^2} \quad .$$

EXEMPLO 6.4.3 - (Distribuição binomial)

Seja $X = \sum_{i=1}^{n} X_i$ onde as X_i são independentes e idênticamente distribuidas $P[X_i=1] = p$ $P[X_i=0] = 1-p=q$. Temos pelo Corolário 6.3.4

$$\text{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{var}(X_i) = n\,\text{var}(X_1)$$

agora como $X_1^2 = X_1$ resulta que

$$\text{var}\,X_1 = EX_1^2 - (EX_1)^2 = EX_1 - (EX_1)^2 = p-p^2 = pq$$

Portanto

$$\text{var}\left(\sum_{i=1}^{n} X_i\right) = npq \; .$$

EXEMPLO 6.4.4 - (Distribuição de Pascal)

Seja $X = X_1 + \ldots + X_r$ onde as X_i são independentes, idênticamente distribuidas em lei comum geométrica com parâmetro p. Temos pelo Corolário 6.3.4

$$\mathrm{var}(X) = \mathrm{var}(\sum_{i=1}^{r} X_i) = r\ \mathrm{var}(X_1) = \frac{r\,q}{p^2}\ .$$

A última igualdade segue do Exemplo 6.4.2.

6.5 - A Desigualdade de Markov

Seja $A \subseteq \mathbb{R}$, X uma variável aleatória sôbre (Ω,P), $f:\mathbb{R} \to \mathbb{R}$, $f \geq 0$ e $f \geq I_A$. Temos o seguinte lema.

LEMA 6.5.1 - $P[X \in A] \leq Ef(X)$.

Demonstração:

$$P[X\in A] = E(I_A \circ X) \leq E(f \circ X) = Ef(X)$$

Dêste inocente lema se deduzem várias desigualdades muito úteis na Teoria de Probabilidades; uma das mais usuais está contida na seguinte proposição.

PROPOSIÇÃO 6.5.1 - (Desigualdade de Markov)

Seja X uma variável aleatória sôbre (Ω,P).

a) $g:\mathbb{R} \to \mathbb{R}$, $g \geq 0$, $g(x) > 0$ se $x > 0$, g é par e não decrescente em $[0,+\infty)$. Então $\forall e > 0$, $P[|X| \geq e] \leq$ $\leq \frac{Eg(X)}{g(e)}$.

b) $g:\mathbb{R}^+ \to \mathbb{R}^+$, $g(x) > 0$ se $x > 0$, g não decrescente em $[0,+\infty)$. Então $P[X \geq e] \leq \frac{Eg(x)}{g(e)}$.

Demonstração: a) $I_{\{x:|x|\geq e\}} \leq \frac{g}{g(e)}$ (porque se $|x| \geq e$,

$g(X) = g(|X|) \geq g(\epsilon))$ e portanto pelo Lema 6.5.1

$$P[|X| \geq \epsilon] \leq E[\frac{g(X)}{g(\epsilon)}] = \frac{Eg(X)}{g(\epsilon)} \quad .$$

b) $I_{\{X:X\geq\epsilon\}} \leq \frac{g}{g(\epsilon)}$ e a desigualdade segúe como em a).

COROLÁRIO 6.5.1 - $\forall r > 0$, $\forall \epsilon > 0$, $P[|X|\geq\epsilon] \leq \frac{E|X|^r}{\epsilon^r}$.

Demonstração: tome $g(X) = |X|^r$.

COROLÁRIO 6.5.2 - (Desigualdade de Chebychev)

Seja X uma variável aleatróia com média μ.
Então $P[|X-\mu|\geq\epsilon] \leq \frac{\mathrm{var}(X)}{\epsilon^2}$.

Demonstração: Seja $Y = X-\mu$ e $g(y) = y^2$. Então

$$P[|X-\mu|\geq\epsilon] \leq \frac{EY^2}{\epsilon^2} = \frac{E(X-\mu)^2}{\epsilon^2} = \frac{\mathrm{var}(X)}{\epsilon^2}$$

COROLÁRIO 6.5.2 - Se X é uma variável aleotória com variância finita $(\mu = EX)$, então

$$\forall \epsilon > 0, \quad P[X \geq \mu+\epsilon] \leq \frac{\mathrm{var}(X)}{\epsilon^2 + \mathrm{var}(X)}$$

e

$$P[X \leq \mu-\epsilon] \leq \frac{\mathrm{var}(X)}{\epsilon^2 + \mathrm{var}(X)}$$

Demonstração: Temos

$$\forall b < \mu+\epsilon \ , \quad I_{\{X:X\geq\mu+\epsilon\}} \leq \frac{(X-b)^2}{(\mu+\epsilon-b)^2}$$

(Ver figura 6.5.1)

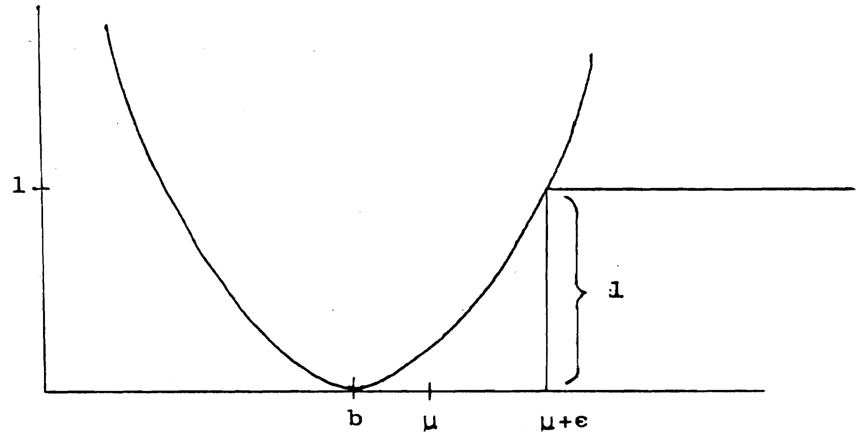


Figura 6.5.1

Portanto

$$P[X \geq \mu+\varepsilon] \leq \frac{E(X-b)^2}{(\mu+\varepsilon-b)^2} = \frac{E(X-\mu+\mu-b)^2}{(\mu+\varepsilon-b)^2} =$$

$$= \frac{\mathrm{var}(X) + (\mu-b)^2}{(\mu+\varepsilon-b)^2} \qquad \forall\ b < \mu+\varepsilon$$

queremos agora escolher b que torne mínima a última expressão. Derivando com respeito a b, obtemos que o mínimo ocorre no ponto $b_o = \mu - \frac{\mathrm{var}(X)}{\varepsilon}$.

Substituindo êste valor na última expressão obtemos

$$P[X \geq \mu+\varepsilon] \leq \frac{\mathrm{var}(X) + [\mu - (\mu-\frac{\mathrm{var}(X)}{\varepsilon})]^2}{(\mu+\varepsilon - \mu+\frac{\mathrm{var}(X)}{\varepsilon})^2} =$$

$$= \frac{\mathrm{var}(X)}{\varepsilon^2+\mathrm{var}(X)}$$

o que prova a primeira desigualdade. A outra se deduz da mesma forma.

6.6 - Aplicação ao Teorema de Wierstrass

Sejam $b_{n,p}(k) = \binom{n}{k} p^k(1-p)^{n-k}$ os coeficientes binomiais sôbre $\{0,1,\dots,n\}$ e $B_{n,p}$ a distribuição binomial induzida. Seja $\mu_{n,p}$ uma distribuição concentrada em $\{0, \frac{1}{n}, \frac{2}{n}, \dots, 1\}$ tal que $\mu_{n,p}(\frac{k}{n}) = b_{n,p}(k)$. Se $f:\Omega \to \mathbb{R}$ e $A \subseteq \Omega$, definimos $(\operatorname{os} f)(A) = \sup\{|f(x)-f(y)| \ {}^{x\in A}_{y\in A}\}$. Se $f:[0,1] \to \mathbb{R}$, então $\int f d\mu_{n,p} = \sum_{k=0}^{n} f(_n) b_{n,p}(k) =$
$= \sum_{k=0}^{n} f(\frac{k}{n}) \binom{n}{k} p^k (1-p)^{n-k}$ é um polinômio em p de grau n onde os coeficientes dependem de k,n e $f(0), f(\frac{1}{r}), \dots,$ $f(1)$.

TEOREMA 6.6.1 (Wierstrass) - Seja $f:[0,1] \to \mathbb{R}$ contínua. Então, $\forall e > 0$, $\exists P$, polinômio tal que

$$\sup_{0\le x\le 1} |f(x)-P(x)| \le e.$$

<u>Nota</u>: Dêste teorema se deduz imediatamente que a aproximação uniforme por polinômios é válida em qualquer intervalo $[a,b]$.

<u>Demonstração</u>: Vamos demonstrar que $\lim_{n\to\infty} \sup_{0\le p\le 1} \left|\int f d\mu_{n,p} - f(p)\right| =$ $= 0$ o que provará o resultado.

$$\left|\int f d\mu_{n,p} - f(p)\right| = \left|\int [f-f(p)] d\mu_{n,p}\right|$$

$$\leq \int |f-f(p)| d\mu_{n,p} = \int_{\{x:|x-p|\leq \epsilon\}} |f-f(p)| d\mu_{n,p} + \int_{\{x:|x-p|>\epsilon\}} |f-f(p)| d\mu_{n,p}$$

$$\leq (\text{os } f)(\{x:|x-p|\leq \epsilon\}) + 2 \sup_{0\leq x\leq 1} |f(x)|\ \mu_{n,p}\{x:|x-p|>\epsilon\}$$

$$= (\text{os } f)(\{x:|x-p|\leq \epsilon\}) + 2 \sup_{0\leq x\leq 1} |f(x)|\ B_{n,p}\{k:|k-np|>\epsilon n\}$$

Chebychev

$$\leq (\text{os } f)(\{x:|x-p|\leq \epsilon\}) + 2 \sup_{0\leq x\leq 1} |f(x)| \frac{1}{4\epsilon^2 n} .$$

Portanto, temos para todo $\epsilon > 0$

$$\lim_n \sup_{0\leq p\leq 1} \left|\int f d\mu_{n,p} - f(p)\right| \leq \sup_{0\leq p\leq 1} (\text{os } f)(\{x.|x-p|\leq \epsilon\}) .$$

Como f é contínua e uniformemente contínua

$$\sup_{0\leq p\leq 1} (\text{os } f)(\{x:|x-p|\leq \epsilon\}) \xrightarrow[\epsilon\to 0]{} 0$$

(Esta última expressão é equivalente à continuidade uniforme).

EXERCÍCIOS

1. Provar a Proposição 6.1.6
2. Provar que se $E|Y| < \infty$ e $A_n \downarrow \emptyset$ então $\int_{A_n} |Y| dP = E|Y|I_{A_n} \downarrow 0$.

3. Calcular a média e a variância da distribuição uniforme. (Sugestão: $1^2+2^2+\ldots+n^2 = \frac{1}{6}\{n(n+1)(2n+1)\}$).

4. Sejam $X_1 \ldots X_n$ variáveis aleatórias independentes identicamente distribuidas com lei comum igual a Poisson com parâmetro 1.

Seja $Y_n = \prod_{i=1}^{n} X_i$.

Calcular:

a) $E\, Y_n$

b) $\operatorname{var}(Y_n)$

c) $P[Y_n=0]$.

5. Se X é uma variável aleatória tal que $|X| \le 1$ então $\forall\, \epsilon > 0$ $P[|X| \ge \epsilon] \ge EX^2 - \epsilon^2$ (Kolmogorov).

6. Se $\{A_n\}_{n=1,2,\ldots}$ é uma sucessão de conjuntos, definimos

$$S_n = \sum_{i=1}^{n} I_{A_i} \qquad e \qquad Y_n = \frac{S_n}{n}$$

Então $\forall\, e > 0$

$$P[|Y_n - EY_n| \ge e] \to 0 \quad \text{se e sòmente se} \quad \operatorname{var}(Y_n) \to 0.$$

7. Provar que a média da distribuição hipergeométrica (com parâmetros n_1, n_2, r; $r \le n_1$, $r \le n_2$) é $\frac{n_1 r}{(n_1+n_2)}$.

8. (Entropia). Seja P uma probabilidade sôbre $\Omega_n = \{\omega_1, \ldots, \omega_n\}$ tal que $p_i = P(\{\omega_i\}) > 0$ $i=1,..,n$.

Definimos

$$H(P) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i} .$$

O número $H(P)$ é chamado a <u>entropia</u> de P.

Se Q é outra probabilidade sôbre Ω_n (com $q_i = Q(\{\omega_i\}) > 0$) definimos

$$D(P,Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} .$$

Provar

a) $D(P,Q) \geq 0$ e $D(P,Q) \doteq 0$
se e sòmente se $P = Q$.

b) Se $p_i = \frac{1}{n}$ $i=1,..,n$ então
$\forall Q \quad H(Q) \leq H(P) = \log(N)$

c) Se $\Omega_r = \{u_1,...,u_r\}$ e $f:\Omega_n \to \Omega_r$
e sôbre então

$$H(Pf^{-1}) \leq H(P)$$

CAPÍTULO 7

VETORES ALEATÓRIOS

7.1 - Propriedades básicas

DEFINIÇÃO 7.1.1 - Seja (Ω,P) um espaço de probabilidades e $\mathbb{R}^n = \{(x_1,\ldots,x_n): x_i \in \mathbb{R}\}$. Tôda função $X:\Omega \to \mathbb{R}^n$ é chamada um <u>vetor aleatório</u> (n-dimensional).

Seja $\pi_i:\mathbb{R}^n \to \mathbb{R}$ a função $(x_1,\ldots,x_n) \mapsto x_i$, $1 \le i \le n$, e $X_i = \pi_i \circ X$, $1 \le i \le n$.

DEFINIÇÃO 7.1.2 - A função $A \longmapsto P[X \in A]$ definida para todo $A \subseteq \mathbb{R}^n$ é chamada distribuição do vetor aleatório X (também <u>distribuição conjunta</u>). A distribuição do vetor aleatório X será denotada com $\mathcal{L}(X)$.

DEFINIÇÃO 7.1.3 - As distribuições das X_i, $1 \le i \le n$, são chamadas <u>distribuições marginais</u>. Sejam $\{x_{i1}, x_{i2}, \ldots\}$ os valores da variável aleatória X_i. Os possíveis valores de $X = (X_1,\ldots,X_n)$ estão contidos no conjunto $\{(x_{1j_1},\ldots,x_{nj_n}): 1 \le j_i < \infty,\ 1 \le i \le n\}$. Como no caso de variáveis aleatórias, a função

$(x_{1j_1},\ldots,x_{nj_n}) \longmapsto P[X = (x_{1j_1},\ldots,x_{nj_n})]$ determina $\mathcal{L}(X)$.

Vamos denotar essa função por

$$p_X(x_{1j_1},\ldots,x_{nj_n})$$

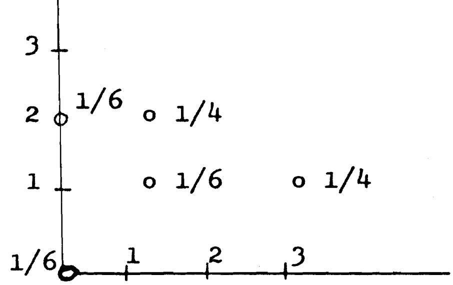
(Note que o valor de p_X vai ser muito freqüentemente igual a 0 em geral). Também determina as leis $\mathcal{L}(X_i)$, $1 \le i \le n$, porque

$$P[X_i = x_{ij}] = \sum_{j_1,\ldots,j_{i-1},j_{i+1},\ldots,j_n} P[X = (x_{1j_1},\ldots,x_{i-1,j-1},x_{ij},x_{i+1,j_{i+1}},\ldots,x_{nj_n})]$$

$$= \sum_{j_1,\ldots,j_{i-1},j_{i+1},\ldots,j_n} p_X(x_{1j_1},\ldots,x_{i-1,j-1},x_{ij},x_{i+1,j+1},\ldots,x_{nj_n})$$

O conhecimento da distribuição conjunta de n variáveis aleatórias e o instrumento necessário para calcular a distribuição de funções do vetor aleatório $(X_1,\ldots,X_n)$, como, por exemplo, $\sum_{i=1}^{n} X_i$, $\max_{1\le i\le n} X_i$, $\min_{1\le i\le n} X_i$, $\sum_{i=1}^{n} |X_i|$, etc.

EXEMPLO 7.1.1 - Consideremos a distribuição conjunta da figura correspondente ao vetor (X,Y).



(Figura 7.1.1)

As seguintes tabelas dão a distribuição de algumas funções de (X,Y).

$X+Y$

valores	distribuição
0	1/6
2	2/6
3	1/4
4	1/4

$X \vee Y$

valores	distribuição
0	1/6
1	1/6
2	5/12
3	1/4

$X \wedge Y$

valores	distribuição
0	1/3
1	2/3

X

valores	distribuição
0	1/3
1	5/12
3	1/4

Y

valores	distribuição
0	1/6
1	5/12
2	5/12

7.2 - Distribuição conjunta de variáveis aleatórias independentes.

Sejam $X_1,\dots,X_n$ n variáveis aleatórias independentes, os valores de X_i são o conjunto $\{x_{i_1}, x_{i_2}, \dots\}$. A relação entre a distribuição conjunta e as distribuições marginais está dada pela Proposição 5.4.1.

Reescreveremos essa proposição da seguinte forma:

PROPOSIÇÃO 7.2.1 - As variáveis $X_1,\dots,X_n$ são independentes se e sòmente se

$$p_X(x_{1j_1},\dots,x_{nj_n}) = \prod_{i=1}^{n} p_{X_i}(x_{ij_i})$$

Dizemos então que a distribuição conjunta é o "produto" das distribuições marginais.

7.3 - Distribuições condicionais

Vamos considerar um vetor bidimensional (X,Y).

Sejam os valores possíveis de X e Y os conjuntos $\{x_1,x_2,\dots\}$, $\{y_1,y_2,\dots\}$.

DEFINIÇÃO 7.3.1 - A função $(x_i,y_j) \longmapsto P[Y=y_j | X=x_i]$ se $P[X=x_i] > 0$ e 0 se $P[X=x_i] = 0$ é chamada <u>distribuição condicional de Y dado X.</u>

O valor desta função no ponto (x_i, y_j) será denotado por $\mu(x_i;\{y_j\})$. Note que para cada x_i tal que $P[X=x_i] > 0$, $\sum_j \mu(x_i,\{y_j\}) - 1$, e portanto é uma probabilidade sôbre $\{y_1,y_2,\ldots\}$ tal que no conjunto A vale

$$\mu(x_i;A) \overset{(\text{def.})}{=} \sum_{\{j:y_j\in A\}} \mu(x_i;\{y_j\}) = P[Y \in A | X = x_i] .$$

É conveniente às vêzes definir a noção de distribuição condicional de uma maneira um pouco diferente. Seja Q uma probabilidade sôbre $\{y_1,y_2,\ldots\}$ determinada pela sucessão $q_i = Q(\{y_i\})$. Definimos

$$\bar{\mu}(x_i;\{y_j\}) = \begin{cases} P[Y=y_j | X=x_i] & \text{se} \quad P[X=x_i] > 0 \\ Q(\{\bar{y}_j\}) & \text{se} \quad P[X=x_i] = 0 \end{cases}$$

e agora $\bar{\mu}(x_i;A) = \sum_{\{j:Y_j\in A\}} \bar{\mu}(x_i;\{Y_j\})$ é uma probabilidade sôbre $\{y_1,y_2,\ldots\}$ <u>para todo</u> X_i $i=1,2,\ldots$.

Dada a distribuição de Y, dado X e a distribuição de X, a distribuição de Y está determinada pela fórmula contida na seguinte

PROPOSIÇÃO 7.3.1:

$$P[Y\in A] = \sum_{i=1}^{\infty} \mu(x_i;A)p_X(x_i)$$

<u>Nota</u>: a mesma fórmula é válida com μ substituida por $\bar{\mu}$.

Demonstração:

$$P[Y\in A] = \sum_{\{i:p_X(x_i)>0\}} P([Y\in A]\cap[X=x_i]) =$$

$$= \sum_{\{i:p_X(x_i)>0\}} P[Y\in A|X=x_i]p_X(x_i) =$$

$$= \sum_{\{i:p_X(x_i)>0\}} \mu(x_i;A)p_X(x_i) = \sum_{i=1}^{\infty} \mu(x_i;A)p_X(x_i) \quad .$$

7.4 - Expectância condicional

Seja Y uma variável aleatória integrável e $A \subseteq \Omega$ com $P(A) > 0$.

DEFINIÇÃO 7.4.1 - O número $E(Y|A) = \frac{1}{P(A)} \int_A Y\, dP$ é chamado <u>expectância condicional de Y dado A</u> (como Y é integrável, êste número está bem definido).

Intuitivamente a $E(Y|A)$ é o valor médio esperado de Y se A ocorreu.

Seja agora X outra variável aleatória com valores $\{x_1,x_2,\ldots\}$.

DEFINIÇÃO 7.4.2 - A variável aleatória

$$E(Y|X) = \sum_{\{i:p_X(x_i)>0\}} E(Y|[X=x_i])I_{[X=x_i]}$$

é chamada <u>expectância condicional de Y dado X</u>.

Intuitivamente ela dá para cada valor de X o valor médio esperado de Y.

Sejam $\{y_1, y_2, \ldots\}$ os valores de Y.

DEFINIÇÃO 7.4.3 - Afunção $x_i \longmapsto \sum_j y_j\, \mu(x_i, \{y_j\})$ é chamada <u>função de regressão de Y sôbre X</u>.

A seguinte proposição contém uma lista das principais propriedades da expectância condicional.

<u>Nota</u>: Se escolhemos Q tal que $\sum_j q_j |y_j| < \infty$ poderíamos usar $\bar{\mu}$ em lugar de μ na Definição 7.4.3. Se chamamos $\bar{g}$ a função obtida usando $\bar{\mu}$ a parte a) da Proposição 7.4.1 é válida quase certamente.

PROPOSIÇÃO 7.4.1 - Seja Y integrável

a) Se g é a função de regressão de Y sôbre X, então $E(Y/X) = g \circ X$.

b) $E(s|X) = 1$

c) $E(CY|X) = CE(Y|X)$

d) $E(Y+Z|X) = E(Y|X) + E(Z|X)$

e) $E(E(Y|X)) = EY$

f) $E(h(X)Y|X) = h(X)\, E(Y|X)$ (aqui se supõe que $E\, h(X)^2 < \infty$ e que $EY^2 < \infty$, e portanto $h(X).Y$ é integrável. Ver anexo 2).

<u>Demonstração</u>: b), c) e d) decorrem imediatamente das pro-

priedades correspondentes para expectância.

a) $$E(Y|X) = \sum_{\{i:p_X(x_i)>0\}} \frac{1}{P[X-x_i]}\left(\int_{[X=x_i]} YdP\right)I_{[X=x_i]} =$$

$$= \sum_{\{i:p_X(x_i)>0\}} \frac{1}{P[X=x_i]} \int I_{[X=x_i]} \left(\sum_{j=1}^{\infty} y_j\, I_{[Y=\bar{y}_j]}\right)dP\; I_{[X=x_i]} =$$

$$= \sum_{\{i:p_X(x_i)>0\}} \frac{1}{P[X=x_i]} \left(\sum_j Y_j\, P(X=x_i, Y=y_j]\right)I_{[X=x_i]} =$$

$$= \sum_{\{i:p_X(x_i)>0\}} \left(\sum_j y_j\, P[Y=y_j|X=x_i]\right)I_{[X=x_i]} =$$

$$= \sum_{\{i:p_X(x_i)>0\}} \left(\sum_j y_j\, \mu(x_i;\{y_j\})\right)I_{[X=x_i]} =$$

$$= \sum_{\{i:p_X(x_i)>0\}} g(x_i)I_{[X=x_i]} = \sum_{i=1}^{\infty} = g \circ X .$$

e) $$EE(Y|X) = \sum_{\{i:p_X(x_i)>0\}} \frac{1}{P[X=x_i]}\left(\int_{[X=x_i]} ydP\right)P[X=x_i] =$$

$$= \sum_{\{i:p_X(x_i)>0\}} \int_{[X=x_i]} YDP = \int_{\underset{\{i:p_X(x_i)>0\}}{\bigcup}[X=x_i]} =$$

$$= \int_{\Omega} YdP.$$

Esta última igualdade decorre do fato de que $P\{\omega : p_X(X(\omega)) = 0\} = 0$.

f) $E(h(X)Y|X) = \sum_{\{i:p_X(x_i)>0\}} \frac{1}{P[X=x_i]} \left(\int_{[X=x_i]} h(X)YdP\right) I_{[X=x_i]} =$

$= \sum_{\{i:p_X(x_i)>0\}} \frac{h(x_i)}{P[X=x_i]} \left(\int_{[X=x_i]} YdP\right) I_{[X=x_i]} =$

$= \left(\sum_{\{i:p_X(x_i)>0\}} \frac{1}{P[X=x_i]} \left(\int_{[X=x_i]} YdP\right) I_{[X=x_i]}\right)\left(\sum_i h(x_i) I_{[X=x_i]}\right)$

$= E(Y|X)\ h(X).$

A seguinte proposição contém uma das propriedades mais importantes da expectância condicional.

Suponhamos que se quer predizer Y usando X. Mais precisamente, se deseja deterinar uma função h tal que $h(X)$ seja o "melhor" preditor de Y, em algum sentido. Êste sentido vai ser o de distância quadrada média. Significa que queremos determinar h de tal forma que $E(Y-h(X))^2$ seja mínima.

A proposição nos diz que tal função existe e é igual à função de regressão de Y sôbre X.

PROPOSIÇÃO 7.4.2 - Seja Y tal que $EY^2 < \infty$. Então $E(Y|X) = g(X)$ é tal que $\forall$ h.

$$E(Y - h(X))^2 \geq E(Y - E(Y|X))^2 .$$

<u>Demonstração</u>:

$$E(Y - h(X))^2 = E(Y - E(Y|X)+E(Y|X) - h(X))^2 =$$

$$= E(Y - E(Y|X))^2 + [E(Y|X) - h(X)]^2 +$$

$$+ 2 E\{(Y - E(Y|X)) (E(Y|X) - h(X))\}.$$

Como $E(Y|X) - h(X) = (g-h)\circ X$ é uma função de X o terceiro têrmo do membro direito é igual a (Proposição 7.4.1, f))

$$2 [E(Y|X) - h(X)]E(Y - E(Y|X))$$

Mais $E(Y - E(Y|X)) = EY - E\,E(Y|X) = EY - EY = 0$

(a primeira igualdade decorre da Proposição 7.4.1 e) e d) e a segunda da Proposição 7.4.1 e)).

Portanto

$$E(Y-h(X))^2 = E[Y-E(Y|X)]^2+[E(Y|X) - h(X)]^2 \geq E(Y-E(Y|X))^2$$

que é o resultado desejado.

Outra noção que é de utilidade é a de variância condicional.

DEFINIÇÃO 7.4.4 - Seja Y tal que $EY^2 < \infty$. A variável aleatória $var(Y|X) = E(Y^2|X) - (E(Y|X))^2$ é chamada <u>variância condicional de Y dado X</u>.

A seguinte proposição indica as propriedades fundamentais das variâncias condicionais.

PROPOSIÇÃO 7.4.3

a) $\operatorname{var}(Y|X) = \sum [\sum_j Y_j^2 \,\mu(x_i;\{y_j\}) - \sum_j y_j \,\mu(x_i;\{y_j\})] I_{[X=x_i]}$

b) $\operatorname{var}(Y) = E \operatorname{var}(Y|X) + E(E(Y|X) - EY)^2$

Demonstração: a) A demonstração é semelhante a da parte a) da Proposição 7.4.1.

b) $\operatorname{var}(Y) = E(Y^2) - (EY)^2 = E\{E(Y^2|X)-(E(Y|X))^2\} +$
$+ E(E(Y|X))^2 - (EY)^2 = E \operatorname{var}(Y|X)+E[E(Y|X)-E(Y)]^2$

Em continuação exemplificamos as noções introduzidas até agora.

EXEMPLO 7.4.1 - Sejam X e Y duas variáveis aleatórias independentes tais que $\mathcal{L}(X)$ e Poisson com parâmetro $\lambda > 0$ e $\mathcal{L}.Y)$ e Poisson com parâmetro $\mu > 0$. Calcular:

a) a distribuição condicional de X dado $X+Y$,

b) a função de regressão de X dado $X+Y$,

c) $E(X|X+Y)$

d) $\operatorname{var}(Y|X)$

e) verifique a fórmula contida na Proposição 7.4.3 b).

Solução: Vamos determinar primeiro a distribuição conjunta de X e Y e a distribuição de $X+Y$.

$$p_{X,Y}(i,j) = P[X=i,Y=j] = \frac{e^{-\lambda}\lambda^i \; e^{-\mu}\mu^j}{i!\,j!}$$

$$i=0,1,\ldots$$
$$j=0,1,\ldots$$

Agora $\quad P[X+Y=t] = \sum_{i=0}^{t} P[X=i,Y=t-i] =$

$$\sum_{i=0}^{t} \frac{e^{-\lambda}\lambda^i \; e^{-\mu}\mu^{t-i}}{i!(t-i)!} = \frac{e^{-(\lambda+\mu)}}{t!} \sum_{i=0}^{t} \binom{t}{i}\lambda^i \mu^{t-i} =$$

$$= \frac{e^{-(\lambda+\mu)}}{t!}(\lambda+\mu)^t \; .$$

Portanto a distribuição de $X+Y$ e Poisson com parâmetro $\lambda+\mu$

a) $\mu(t;\{i\}) = P[X=i \mid X+Y=t] = \dfrac{P[X=i,Y=t-i]}{P[X+Y=t]} =$

$$= \frac{e^{-\lambda}\lambda^i \; e^{-\mu}\mu^{t-i} \; t!}{i!(t-i)!e^{-(\lambda+\mu)}(\lambda+\mu)^t} = \binom{t}{i} \frac{\lambda^i \mu^{t-i}}{(\lambda+\mu)^i(\lambda+\mu)^{t-i}} =$$

$$= \binom{t}{i} \left(\frac{\lambda}{\lambda+\mu}\right)^i \left(\frac{\mu}{\lambda+\mu}\right)^{t-i}$$

para $0 \le i \le t \quad t=0,1,\ldots$

Portanto a distribuição condicional de X dado $X+Y$ é binomial com parâmetro $\dfrac{\lambda}{\lambda+\mu}$. Isto significa que

$$\mu(t;\{i\}) = \begin{cases} b_{t,\frac{\lambda}{\lambda+\mu}}(i) & 0 \le i \le t \\ 0 & t < i \end{cases}$$

$$t = 0,1,\ldots$$

b) $\quad g(t) = \sum_{i=0}^{t} i \; b_{t,\frac{\lambda}{\lambda+\mu}}(i) = \dfrac{t\lambda}{\lambda+\mu}$

(média de uma distribuição binomial com parâmetros $t, \frac{\lambda}{\lambda+\mu}$)

c) $E(X|X+Y) = g \circ (X+Y) = \dfrac{(X+Y)\lambda}{\lambda+\mu}$.

d) $$\sum_{i=0}^{t} i^2\, \mu(t;\{i\}) - [\sum_{i=0}^{t} i\, \mu(t;\{i\})]^2 = \frac{t\lambda\mu}{(\lambda+\mu)(\lambda+\mu)} =$$

$$= \frac{t\lambda\mu}{(\lambda+\mu)^2} \ .$$

Portanto

$$\operatorname{var}(X|X+Y) = \frac{(X+Y)\lambda\mu}{(\lambda+\mu)^2}$$

e) $\operatorname{var}(X) = \lambda$

$$E\left[\frac{(X+Y)\lambda\mu}{(\lambda+\mu)^2}\right] + E\left[\frac{(X+Y)\lambda}{\lambda+\mu} - \lambda\right]^2 =$$

$$= \frac{(\lambda+\mu)\lambda\mu}{(\lambda+\mu)^2} + E\left[\frac{\lambda[(X+Y)-(\lambda+\mu)]}{\lambda+\mu}\right]^2 =$$

$$= \frac{\lambda\mu}{\lambda+\mu} + \frac{\lambda^2}{(\lambda+\mu)^2} \quad \operatorname{var}(X+Y) =$$

$$= \frac{\lambda\mu}{\lambda+\mu} + \frac{\lambda^2}{(\lambda+\mu)^2}(\lambda+\mu) = \frac{\lambda\mu+\lambda^2}{\lambda+\mu} = \frac{\lambda(\mu+\lambda)}{(\lambda+\mu)} = \lambda \ .$$

7.5 - Predição linear

Dado um vetor aleatório (X,Y) vamos estudar agora o problema de encontrar o "melhor" preditor linear de Y no sentido de que minimize a distância quadrada média esperada. Antes de dar resposta a êsse problema vamos introduzir várias definições prévias. Nesta seção vamos supor que $EX^2 < \infty$, $EY^2 < \infty$ e que $var(X) > 0$ e $var(Y)>0$ para evitar casos triviais.

DEFINIÇÃO 7.5.1 - O número $E(X-EX)(Y-EY)$ é chamado <u>covariância de X e Y</u> e é denotado por $cov(X,Y)$. Note que se L_2 denota o espaço das variáveis aleatórias tais que $EX^2 < \infty$ então a covariância é uma função de $L_2 \times L_2$ em $\mathbb{R}$.

O fato de que a covariância existe é uma conseqüência da desigualdade de Cauchy-Schwarz contida no Anexo 2.

A função $cov: L_2 \times L_2 \to \mathbb{R}$ tem as seguintes propriedades

PROPOSIÇÃO 7.5.1

a) $cov(X,Y) = E(XY) - (EX)(EY)$

b) $cov(X,Y) = cov(Y,X)$ (cov é simétrica)

c) Se $var(X) = 0$ (ou $var(Y) = 0$) então $cov(X,Y) = 0$

d) cov e bilineal ($cov(aX,Y) = a\, cov(X,Y)$ para $a \in \mathbb{R}$, e $cov(X+Z,Y) = cov(X,Y) + cov(Z,Y)$)

X,Y,Z variáveis aleatórias em L_2.

<u>Demonstração</u>:

a) $cov(X,Y) = E(X-EX)(Y-EY) = E[XY-(EX)Y-(EY)X+(EX)(EY)] =$

$= E\,XY - (EX)(EY).$

b) Decorre imediatamente de a).

c) Decorre de a) e do fato de que se $\operatorname{var}(X) = 0$ então $X = EX$, com probabilidade 1.

d) $\operatorname{cov}(aX,Y) = E(aX)Y - E(aX)EY = a[E(XY)-(EX)(EY)] =$

$= a \operatorname{cov}(X,Y)$

$\operatorname{cov}(X+Z,Y) = E(X+Z)Y - E(X+Z)EY = EXY + EZY -$

$- (EX)(EY) - (EZ)(EY) = \operatorname{cov}(X,Y) + \operatorname{cov}(Z,Y)$

DEFINIÇÃO 7.5.2 - Se $\operatorname{cov}(X,Y) = 0$, X e Y são ditas <u>não-correlacionadas</u>.

DEFINIÇÃO 7.5.3 - O número $\rho_{X,Y} = \dfrac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}$ é chamado o coeficiente de correlação entre X e Y.

<u>Nota</u>: Como $|\rho|^2 = \left|E\left(\frac{X-EX}{\sigma_X}\right)\left(\frac{Y-EY}{\sigma_Y}\right)\right|^2 \le E\left(\frac{X-EX}{\sigma_X}\right)^2 E\left(\frac{Y-EY}{\sigma_Y}\right)^2 = 1$

resulta que $|\rho| \le 1$.

(Ver desigualdade de Cauchy-Schwarz no Anexo 2).

PROPOSIÇÃO 7.5.2 - Se $E(X_i^2) < \infty$, $1 \le i \le n$.

$$\operatorname{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{var}(X_i) + \sum_{i<j} \operatorname{cov}(X_i, X_j).$$

<u>Demonstração</u>: $\operatorname{var}\left(\sum_{i=1}^{n} X_i\right) = E\left[\sum_{i=1}^{n} (X_i - EX_i)\right]^2 =$

$$= \sum_{i=1}^{n} E(X_i - EX_i)^2 + \sum_{i \ne j} E(X_i - EX_i)(X_j - EX_j) =$$

$$= \sum_{i=1}^{n} \operatorname{var}(X_i) + 2 \sum_{i<j} \operatorname{cov}(X_i, X_j) .$$

Vamos agora resolver o problema enunciado no comêço desta seção. Suponhamos que $EX = EY = 0$, $\operatorname{var}(X) = \operatorname{var}(Y) = 1$. Queremos encontrar α e β tais que $E(Y - \alpha X - \beta)^2$ seja mínimo. Agora $E(Y-\alpha X-\beta)^2 = E(Y-\alpha X)^2 + \beta^2 = EY^2 - 2\alpha E(XY) + \alpha^2 E(X^2) + \beta^2 = 1-2\alpha\rho_{X,Y} + \alpha^2 + \beta^2$.

Claramente é conveniente tomar $\beta = 0$. Esta última expressão (completando o quadrado) é igual a

$$(\alpha - \rho_{X,Y})^2 + 1 - \rho^2_{X,Y}$$

e portanto o mínimo se obtém tomando

$$\alpha = \rho_{X,Y} \quad e \quad \beta = 0 .$$

O valor de $E(Y-\alpha X-\beta)^2$ para êste particular par é $1-\rho^2$.

O caso geral é resolvido reduzindo o mesmo ao caso anterior. Sejam

$$Y = \sigma_Y Y' + EY \qquad X = \sigma_X X' + EX$$

onde $EX' = EY' = 0$, $\operatorname{var}(X') = \operatorname{var}(Y') = 1$.

Então

$$E(Y-aX-b)^2 = E(Y'\sigma_Y + EY - a\sigma_X X' - aEX - b)^2 =$$

$$= \sigma_Y^2 \, E\left[Y' - \frac{a\sigma_X}{\sigma_Y} X' + \frac{EY - aEX - b}{\sigma_Y}\right]^2 .$$

Quando a e b variam $\frac{a\sigma_X}{\sigma_Y}$ e $\frac{EY - aEX - b}{\sigma_Y}$ varrem todos os possíveis pares de números reais e, portanto, o mínimo se obtém tomando

$$\frac{a\sigma_X}{\sigma_Y} = \rho \qquad e \qquad EY - aEX - b = 0$$

ou seja

$$\hat{a} = \frac{\sigma_Y \rho}{\sigma_X} \qquad e \qquad \hat{b} = EY - \hat{a}EX .$$

PROPOSIÇÃO 7.5.3

a) $\rho_{aX+b,cY+d} = \operatorname{sgn}(ac)\rho_{X,Y}$ se $(ac) \neq 0$

b) Se $|\rho_{X,Y}| = 1$ então $Y = \hat{a}X+\hat{b}$ com probabilidade 1.

Demonstração:

a) $$\rho_{aX+b,cY+d} = \frac{\operatorname{cov}(aX+b,cY+d)}{\sigma_{aX+b}\,\sigma_{cY+d}} = \frac{\operatorname{cov}(aX,cY)}{|a|\sigma_X |c|\sigma_Y} =$$

$$= \frac{ac}{|a||c|} \frac{\operatorname{cov}(X,Y)}{\sigma_X\sigma_Y} = \operatorname{sgn}(ac)\rho_{X,Y}$$

a segunda e a terceira igualdades decorrem das propriedades ca covariância contidas na Proposição 7.5.1.

b) Dado X e Y definimos $X' = \frac{X-EX}{\sigma_X}$ e $Y' = \frac{Y-EY}{\sigma_Y}$.

$\rho_{X,Y} = 1$ implica $\rho_{X',Y'} = 1$, pela parte a).

Como $E(Y'-\rho X')^2 = 1-\rho^2$, $|\rho| = 1$ implica $1-\rho^2 = 0$ e, portanto, $Y'-\rho X' = 0$ com probabilidade 1 e desta identidade decorre a parte b) imediatamente.

7.6 - Independência de Vetores aleatórios

DEFINIÇÃO 7.6.1 - Seja $\{X_\alpha\}_{\alpha \in I}$ uma família de vetores aleatórios (com valores em $\mathbb{R}^n$). Dizemos que são independentes se

$\forall \alpha_1, \alpha_2, \ldots, \alpha_n$, $\forall A_1, \ldots, A_n$, $(A_i \subseteq \mathbb{R}^n, \quad 1 \leq i \leq n)$

$$P[X_{\alpha_i} \in A_i \; : 1 \leq i \leq n] = \prod_{i=1}^{n} P[X_{\alpha_i} \in A_i].$$

Note-se que a definição é similar a de independência de variáveis aleatórias (os conjuntos A_i agora são subconjuntos de $\mathbb{R}^n$). A grande maioria das proposições referentes a variáveis aleatórias podem ser generalizadas para cobrir o caso de vetores aleatórios. Nós não vamos utilizar essas generalizações. Para efeito de ser usada na seção seguinte só vamos observar que se μ é uma distribuição em $\mathbb{R}^n$ é possível construir vetores aleatórios $X_1, X_2, \ldots, X_k$ tais que são independentes e $\mathcal{L}(X_i) = \mu$ para todo $1 \leq i \leq k$. A demonstração é igual a correspondente para variáveis aleatórias. (Ver Proposição 4.4.1). Também funções de vetores independentes são independentes (Ver Proposição 5.4.3).

7.7 - Mais sôbre as distribuições de Poisson e Binomial

Consideremos a seguinte distribuição conjunta das variáveis V e W.

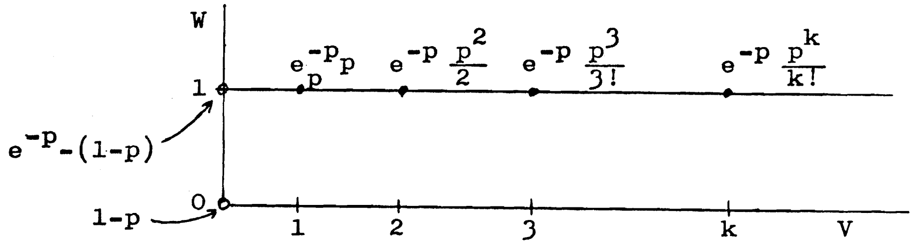


Figura 7.7.1

PROPOSIÇÃO 7.7.1 - $\mathcal{L}(V)$ e Poisson com parâmetro p,
$\mathcal{L}(W)$ e binomial $(1,p)$ e $P(V \neq W) \leq p^2$.

Demonstração: As duas primeiras afirmações são fàcilmente verificáveis.

Agora

$$P(V \neq W) = 1-(1-p) - e^{-p}p = p-e^{-p}p = p[1-e^{-p}] .$$

Como $e^{-p} \geq 1-p$ a última expressão é menor ou igual a

$$p[1-(1-p)] = p^2$$

que é o resultado desejado.

Sejam X e Y duas variáveis aleatórias que têm os mesmos valores $\{x_1, x_2, \ldots\}$.

PROPOSIÇÃO 7.7.2

$$\sum_{i=1}^{\infty} |p_X(x_i)-p_Y(x_i)| \le 2P[X\neq Y]$$

Demonstração: Seja $B = \{i: p_X(x_i) > p_Y(x_i)\}$ e $A = \{x_i: i \in B\}$. Como $1 = \sum_{i=1}^{\infty} p_X(x_i) =$ $= \sum_{i=1}^{\infty} p_Y(x_i)$ é fácil ver que $\sum_{i=1}^{\infty} |p_X(x_i)-p_Y(x_i)| =$

$= 2 \sum_{i\in B} p_X(x_i) - p_Y(x_i) = 2[P(X\in A) - P(Y\in A)] =$

$= 2[P(X\in A, X=Y) + P(X\in A, X\neq Y) - P(Y\in A, X\neq Y) - P(Y\in A, X=y)] =$

$= 2[P(X\in A, X\neq Y) - P(Y\in A, X\neq Y)] \le 2P(X\in A, X\neq Y) \le 2P[X\neq Y]$.

Consideramos agora uma distribuição de Poisson com parâmetro np $(p_{np}(k) = \frac{e^{-np}(np)^k}{k!}$ $k=0,1,\dots)$ e uma distribuição binomial com parâmetros n e p $(b_{n,p}(k) = \binom{n}{k} p^k q^{n-k}$ $k=0,1,\dots,n$, e 0 se $k > n)$. Temos o seguinte importante resultado.

PROPOSIÇÃO 7.7.3

$$\sum_{k=0}^{\infty} |b_{n,p}(k)-p_{np}(k)| \le 2np^2$$

(e portanto se np é limitado (por exemplo $np \to \lambda$) e $p \to 0$, $\sum_{k=0}^{\infty} |b_{n,p}(k)-p_{np}(k)| \to 0)$.

Demonstração: Sejam $(V_1,W_1),\dots,(V_n,W_n)$ vetores aleatórios independentes todos com a distribuição da Figura 7.7.1. Então como $V_1,V_2,\dots,V_n$ são independen-

tes $\mathcal{L}(\sum_{i=1}^{n} V_i)$ é Poisson np. (Ver Exemplo 7.4.1).
Igualmente $\mathcal{L}(\sum_{i=1}^{n} W_i)$ e Binomial n,p.

Pela Proposição 7.7.2 temos

$$\sum_{k=0}^{\infty} |b_{n,p}(k)-p_{np}(k)| \leq 2P[\sum_{i=1}^{n} V_i \neq \sum_{i=1}^{n} W_i]$$

$$\leq 2P(\bigcup_{i=1}^{n}[V_i \neq W_i]) \leq 2\sum_{i=1}^{n} P(V_i \neq W_i) = 2np^2 .$$

<u>Nota</u>: Não é difícil provar com métodos elementares que $\sum_{k=0}^{\infty} |b_{n,p}(k)-p_{np}(k)| \leq 4p$ (Ver [5]). Também é certo que $\sum_{k=0}^{\infty}|b_{n,p}(k)-p_{np}(k)| \leq 3p$ mas a prova é mais complicada. Para resultados mais elaborados ver [8].

EXERCÍCIOS

1. Provar que o melhor preditor constante em distância média quadrada de uma variável aleatória X com $EX^2 < \infty$ e a constante EX.

2. Provar que para todo $A \subseteq \Omega$, Y integrável
$$\int_A YdP = \int_A E(Y|X)dP$$

3. Suponhamos que a variável aleatória X tem uma distribuição de Pascal com parâmetro 2. Seja τ o tempo ao qual observa-se a primeira cara. Calcular a distribuição condicional de τ dado X.

4. Se a distribuição de X e Poisson com parâmetro $\lambda > 0$, e a distribuição condicional de Y dado X e Poisson com parâmetro $\lambda > 0$, qual é a distribuição de Y?

5. Se $\hat{Y} = \hat{a}X + \hat{b}$ provar que $\operatorname{cov}(Y-\hat{Y},X) = 0$.

6. Se $X_1,\dots,X_n$ não são correlacionadas então

$$\operatorname{var}\left(\sum_{i=1}^{n} a_i X_i + b_i\right) = \sum_{i=1}^{n} a_i^2 \operatorname{var}(X_i)$$

7. a) Provar que se X e Y são independentes então não são correlacionadas.

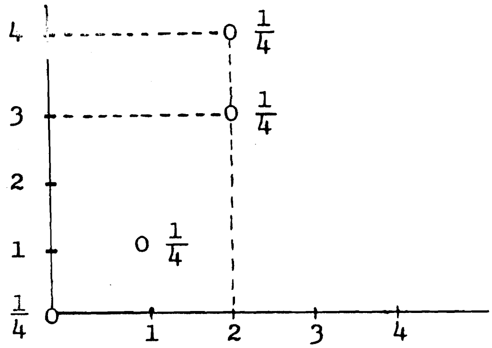
   b) Seja X uma variável aleatória tal que $\mathcal{L}(X)$ é simétrica. $(\mathcal{L}(X)(A) = \mathcal{L}(X)(-A))$. Seja $g:\mathbb{R} \to \mathbb{R}$ uma função par. Então

$$\operatorname{cov}(X,g(X)) = 0$$

   (E em geral X e $g(X)$ não são independentes).

8. $|\sigma_X - \sigma_Y| \leq \sigma_{X+Y} \leq \sigma_X + \sigma_Y$.

9. Calcular a reta de regressão de Y sôbre X e a função de regressão de Y sôbre X para a distribuição conjunta da figura.



10. Sejam X e Y duas variáveis aleatórias independentes tal que $\mathcal{L}(X)$ é binomial (n,p) e $\mathcal{L}(Y)$ é binomial (m,p).

Calcular:

a) A distribuição condicional de X dado $X+Y$.

b) A função de regressão de X sôbre $X+Y$.

c) $E(X|X+Y)$

d) Verificar que $E[E(X|X+Y)] = EX$.

CAPÍTULO 8

INTRODUÇÃO AO TEOREMA DO LIMITE CENTRAL

O objetivo da presente seção é introduzir a noção de convergência em distribuição e enunciar uma versão do Teorema do Limite Central. Não pretendemos dar nenhuma demonstração geral dêste importante resultado já que cremos que as técnicas usadas em sua demonstração excedem o nível dêste livro. Sua inclusão aqui tem o propósito de mostrar que se se pretende estudar o comportamento limite de sucessões de distribuições há necessidade de introduzir necessàriamente distribuições não concentradas em espaços discretos como os estudados até agora.

8.1 - A distribuição Normal

Consideremos a função

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad -\infty < x < +\infty$$

Esta função é chamada densidade normal e suas propriedades mais importantes estão contidas na seguinte proposição.

PROPOSIÇÃO 8.1.1

a) f é par, e ≥ 0.

b) f é decrescente em $[0,+\infty)$.

c) f é côncava em $[-1,1]$ e convexa em $\{x : x \geq 1\}$ e $\{x : x \leq -1\}$.

d) $\int_{-\infty}^{+\infty} f(x)dx < \infty$ mais ainda $\int_{-\infty}^{+\infty} f(x)dx = 1$.

<u>Demonstração</u>: a) e b) são imediatas.

c) derivando f duas vêzes temos $f'(x) = -xf(x)$ $f''(x) = (x^2-1)f(x)$. Portanto $f''(x) \leq 0$ se $|x| \leq 1$ e $f''(x) > 0$ se $|x| > 1$ o que prova c).

d) Como f é limitada e $f(x) \leq \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}}$ se $x \geq 1$, $\int_0^{+\infty} f(x)dx < \infty$ e portanto

$$\int_0^{+\infty} f(x)dx = 2\int_0^{+\infty} f(x)dx < \infty$$

Para provar que $\int_{-\infty}^{+\infty} f(x)dx = 1$ basta provar que $\left(\int_{-\infty}^{+\infty} f(x)dx\right)\left(\int_{-\infty}^{+\infty} f(y)dy\right) = 1$. Mas o membro esquerdo é igual a $\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{x^2+y^2}{2}} dxdy$. Passando a coordenadas polares $x = r\cos\varphi$, $y = r \operatorname{sen}\varphi$ $0 \leq r < \infty$, $0 \leq \varphi < 2\pi$, $\left|\frac{\partial(x,y)}{\partial(r,\varphi)}\right| = r$ temos que esta integral é igual a

$$\frac{1}{2\pi}\int_0^{+\infty}\int_0^{2\pi} e^{-\frac{r}{2}}\, r\, dr\, d\varphi = \int_0^{+\infty} e^{-\frac{r^2}{2}}\, r\, dr = 1$$ o que

prova d).

Às vêzes é importante poder dar ràpidamente uma cota superior para a cauda da distribuição normal e a rapidez da convergência a 0. A seguinte proposição é útil para isso.

PROPOSIÇÃO 8.1.2

$$\forall x > 0 \quad \frac{f(x)}{x}\left(1-\frac{1}{x^2}\right) < \int_x^{+\infty} f(x)dx < \frac{f(x)}{x}$$

Demonstração:

$$\frac{f(x)}{x} = \int_x^{+\infty} f(y)\left(1+\frac{1}{y^2}\right)dy$$

porque a derivada do membro direito é igual à do membro esquerdo (Teorema Fundamental do Cálculo) e então êles diferem uma constante. Essa constante deve ser 0 porque ambos os membros convergem a 0 quando $x \to \infty$.

De modo similar prova-se que

$$f(x)\left(\frac{1}{x}-\frac{1}{x^3}\right) = \int_x^{+\infty} f(y)\left(1-\frac{3}{y^4}\right)dy$$

Como

$$\int_x^{+\infty} f(y)\left(1-\frac{3}{y^4}\right)dy \le \int_x^{+\infty} f(y)dy \le \int_x^{+\infty} f(y)\left(1+\frac{1}{y^2}\right)dy$$

o resultado está provado.

Vamos indicar a função

$$\int_{-\infty}^{x} f(y)dy \quad \text{por} \quad \phi(x) .$$

Da Proposição 8.1.1 resulta imediatamente que ϕ é uma função de distribuição. Valores aproximados de ϕ estão contidos na Tabela nº 1. Note que dado que f é simétrica só os valores de ϕ para $x \geq 0$ são necessários.

8.2 - Convergência em distribuição

Sejam F e $\{F_n\}_{n=1,2,\ldots}$ uma função de distribuição e uma sucessão de funções de distribuição respectivamente.

DEFINIÇÃO 8.2.1 - Dizemos que F_n converge em distribuição a F (ou converge fracamente) se $F_n(t) \to F(t)$ sempre o que t é um ponto de continuidade de F. Se X_n são variáveis aleatórias com funções de distribuição F_n dizemos que X_n converge em istribuição a F (e escrevemos $\mathcal{L}(X_n) \to F$) se as funções de distribuição F_n convergem fracamente a F.

É fácil construir exemplos de funções de distribuição correspondentes a variáveis aleatórias discretas que convergem a funções de distribuição não discretas. Por exemplo se X_n toma os valores $\{0,\frac{1}{n},\frac{2}{n},\ldots,\frac{n-1}{n},1\}$ com probabilidades $\frac{1}{n+1}$ e F_n é a função de distribuição de X_n então $\mathcal{L}(X_n) \to F$ onde

$$F(t) = \begin{cases} 1 & t \geq 1 \\ t & 0 < t < 1 \\ 0 & t < 0 \end{cases}$$

O seguinte teorema pode ser utilizado para construir exemplos dêste fenômeno embora seja claro porque sua importância não reside nesse fato.

Seja $X_1,X_2,\ldots$ uma sucessão de variáveis aleatórias independentes e idênticamente distribuidas com $EX_1^2 < \infty$.

Seja $\mu = EX_1$ e $\sigma^2 = \text{var}(X_1)$.

TEOREMA 8.2.1 (Teorema do limite central)

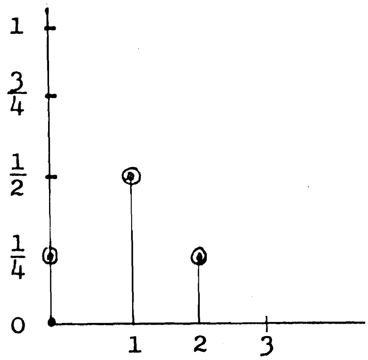
$$\forall a \leq b \quad P\left[a \leq \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma} \leq b\right] \to \int_a^b \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, d\bar{x} =$$

$$= \phi(b) - \phi(a) \; .$$

Vamos agora utilizar êste teorema em alguns exemplos particulares para calcular probabilidades aproximadamente.

Mais adiante vamos provar uma versão particular dêste teorema e ao mesmo tempo demonstrar que na fórmula de Stirling a constante $e^c = \sqrt{2\pi}$.

EXEMPLO 8.2.1 - Consideremos a distribuição da figura



Sejam $X_1, X_2, \ldots, X_{50}$, 50 variáveis aleatórias independentes e idênticamente distribuidas com a distribuição da figura. Calcular aproximadamente a probabilidade de que a soma seja ≤ 55.

Solução: A média e a variância da distribuição da figura são 1 e $\frac{1}{2}$. Então

$$P\left[\sum_{i=1}^{50} X_i \leq 55\right] = P\left[\frac{\sum_{i=1}^{50} X_i - 50}{\sqrt{\frac{50}{2}}} \leq \frac{55-50}{\sqrt{\frac{50}{2}}}\right] =$$

$$= P\left[\frac{\sum_{i=1}^{50} X_i - 50}{\sqrt{\frac{50}{2}}} \leq 1\right] \cong \int_{-\infty}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \phi(1) = 0.84$$

EXEMPLO 8.2.2 - Jogue uma moeda equilibrada 50 vêzes.

Calcule aproximadamente

a) A probabilidade de que o número de caras seja ≥ 20 e $\leq$ que 30.

b) A probabilidade de que o número de caras seja 23.

<u>Solução</u>: Seja X tal que $\mathcal{L}(X)$ é binomial com parâmetros 50 e $\frac{1}{2}$.

$$P[20 \leq X \leq 30] = P[19,5 \leq X \leq 30,5] -$$

$$= P\left[\frac{19,5-25}{\sqrt{\frac{50}{4}}} \leq \frac{X-25}{\sqrt{\frac{50}{4}}} \leq \frac{30,5-25}{\sqrt{\frac{50}{4}}}\right] =$$

$$= P\left[-1,56 \leq \frac{X-25}{\sqrt{\frac{50}{4}}} \leq 1,56\right] \cong \int_{-1,56}^{1,56} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \phi(1,56) -$$

$$- \phi(-1,56) = 2\phi(1,56)-1 = 0,88.$$

b) $P[X=23] = P[22,5 < X < 23,5] =$

$$= P\left[\frac{22,5-25}{\sqrt{\frac{50}{4}}} \leq \frac{X-25}{\sqrt{\frac{50}{4}}} \leq \frac{23,5-25}{\sqrt{\frac{50}{4}}}\right] =$$

$$= P\left[-0,71 \leq \frac{X-25}{\sqrt{\frac{50}{4}}} \leq -0,43\right] =$$

$$= \int_{-0,71}^{-0,43} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \phi(-0,43)-\phi(-0,71) =$$

$$= \phi(0,71) - \phi(0,43) = 0,76 - 0,67 = 0,09.$$

<u>Nota</u>: A adição e a subtração de 1/2 nos cálculos efetuados em a) e b) deve-se ao fato de que esta correção fornece uma melhor aproximação ao valor verdadeiro da probabilidade. Esta correção só é utilizada no caso da aproximação normal à distribuição binomial.

Voltamos agora ao Teorema do Limite Central em uma situação muito particular. Êste é um dos primeiros teoremas limite da teoria das probabilidades e cremos que sua demonstração é instrutiva.

TEOREMA 8.2.2 - Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes $X_i = \begin{cases} 1 & 1/2 \\ 0 & 1/2 \end{cases}$. Então

$$\forall x \quad P\left[\frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \le x\right] \xrightarrow[n \to \infty]{} \phi(x)$$

e a constante e^c da fórmula de Stirling é igual a $\sqrt{2\pi}$.

<u>Demonstração</u>: Vamos estudar primeiro o caso de n par.

$$P[|S_{2n}-n| < x\sqrt{\tfrac{n}{2}}] = P[n-x\sqrt{\tfrac{n}{2}} < S_{2n} < n+x\sqrt{\tfrac{n}{2}}] =$$

$$= \sum_{n-x\sqrt{\frac{n}{2}}<k<n+x\sqrt{\frac{n}{2}}} P[S_{2n} = k]$$

definindo $j = k-n$ temos que

$$P[|S_{2n}-n| < x\sqrt{\tfrac{n}{2}}] = \sum_{|j|<x\sqrt{\frac{n}{2}}} P[S_{2n} = n+j] \ .$$

Agora

$$P[S_{2n}=n+j] = \binom{2n}{n+j}\frac{1}{2^{2n}} = \frac{2n!}{(n+j)!(n-j)!}\frac{1}{2^{2n}} =$$

$$= \frac{2n!}{n!n!}\frac{n!}{(n+j)!}\frac{n!}{(n-j)!}\frac{1}{2^{2n}} =$$

$$= P[S_{2n}=n]\frac{n(n-1)\ldots(n-j+1)}{(n+j)(n+j-1)\ldots(n+1)} =$$

$$P[S_{2n}=n]\,\frac{1}{(1+\frac{j}{n})(1+\frac{j}{n-1})\ldots(1+\frac{j}{n-j+1})} =$$

$$= P[S_{2n}=n]\prod_{k=0}^{j-1}\frac{1}{(1+\frac{1}{n-k})} .$$

Definimos $\quad A_{n,j} = \prod_{k=0}^{j-1}\frac{1}{(1+\frac{j}{n-k})}$.

Então

$$\log A_{n,j} = -\sum_{k=0}^{j-1}\log(1+\frac{j}{n-k}) .$$

Como $\log(1+x) = x(1+\epsilon(x))$ para $|x| < 1$ onde $\epsilon(x) \to 0$ quando $x \to 0$ nós temos

$$\log A_{n,j} = -\sum_{k=0}^{j-1}\frac{j}{n-k}\,(1+\epsilon(\frac{j}{n-k})) .$$

Escolhamos $\epsilon_{n,j}$ tal que

$$\min_{0\le k\le j-1}\epsilon(\frac{j}{n-k}) \le \epsilon_{n,j} \le \max_{0\le k\le j-1}\epsilon(\frac{j}{n-k})$$

e

$$\log A_{n,j} = -(1+\epsilon_{n,j})\sum_{k=0}^{j-1}\frac{j}{n-k} .$$

Como j varia no conjunto $B_n = \{j : |j| < x\sqrt{\frac{n}{2}}\}$

$$\max_{j\in B_n}|\epsilon_{n,j}| \le \sup_{|y|\le\frac{x}{\sqrt{2\pi}}}|\epsilon(y)| \xrightarrow[n\to\infty]{} 0 .$$

Temos agora

$$\log A_{n,j} = -(1+\epsilon_{n,j})\sum_{k=0}^{j-1}\frac{j}{n}\,\frac{n}{n-k} = -(1+\epsilon_{n,j})\sum_{k=0}^{j-1}\frac{j}{n}(\frac{1}{1-\frac{k}{n}}) .$$

Fazendo um argumento similar ao anterior podemos escolher $e'_{n,j}$ tal que

$$\log A_{n,j} = -(1+e'_{n,j})\sum_{k=0}^{j-1} \frac{j}{n}$$

onde

$$\max_{j\in B_n} |e'_{n,j}| \to 0 .$$

Temos

$$\log A_{n,j} = -(1+e'_{n,j}) \frac{j^2}{n}$$

ou seja

$$A_{n,j} = e^{(-\frac{j^2}{n})(1+e'_{n,j})}$$

É fácil ver que existem números $\delta_{n,j}$ tais que

$$\sum_{j\in B_n} |\delta_{n,j}| \to 0$$

e

$$A_{n,j} = (1+\delta_{n,j})e^{-\frac{j^2}{n}} .$$

Temos finalmente

$$P[|S_{2n}-n|< x\sqrt{\tfrac{n}{2}}] = (\sum_{j\in B_n} (1+\delta_{n,j})e^{-\frac{j^2}{n}})P[S_{2n}=n] .$$

Pela fórmula de Stirling

$$P[S_{2n}=n) \cong \frac{1}{e^c}\sqrt{\frac{2}{n}}$$

e temos

$$P[|S_{2n}-n|< x\sqrt{\tfrac{n}{2}}] \cong \frac{(1+\delta_n)}{e^c} \sum_{j\in B_n} \sqrt{\frac{2}{n}}\, e^{-\frac{j^2}{n}}$$

fazendo a mudança de variável

$$y_j = j\sqrt{\frac{2}{n}} \quad \text{e} \quad \Delta_y = y_{j+1} - y_j = \sqrt{\frac{2}{n}}$$

o membro direito é igual a

$$\frac{(1+\delta_n)}{e^c} \sum_{-x<y_j<x} \Delta_y \, e^{-\frac{y_j^2}{2}}$$

quando $n \to \infty$ isto converge a

$$\frac{1}{e^c} \int_{-x}^{x} e^{-\frac{y^2}{2}} \, dy \quad .$$

Vamos agora estimar

$$P[\,|S_{2n}-n| > x\sqrt{\tfrac{n}{2}}] \quad .$$

$$P[\,|S_{2n}-n| > x\sqrt{\tfrac{n}{2}}] \le 2P[S_{2n} \ge n + [x\sqrt{\tfrac{n}{2}}\,]] \quad .$$

Pela desigualdade (5.2.2) esta última expressão é menor ou igual a

$$\frac{\{n+[x\sqrt{\frac{n}{2}}]\}\frac{1}{2}}{\{n+[x\sqrt{\frac{n}{2}}]-(2n+1)\frac{1}{2}\}^2} = \frac{\{n+[x\sqrt{\frac{n}{2}}]\}\frac{1}{2}}{\{[x\sqrt{\frac{n}{2}}]-\frac{1}{2}\}^2}$$

quando $n \to \infty$ a última expressão converge a $\frac{1}{x^2}$ e portanto

$$\lim_{x\to\infty} \limsup_{n} P[\,|\frac{S_{2n}-n}{\sqrt{\frac{n}{2}}}| > x] = 0 \; .$$

Pelo exercício 3 $\frac{1}{e^c}\int_{-\infty}^{+\infty} e^{\frac{y^2}{2}} \, dy = 1$ e portanto

$e^c = \sqrt{2\pi}$.

Do **exercício** 2 decorre agora imediatamente que $P[\frac{S_{2n}-n}{\sqrt{\frac{n}{2}}} < x] \to \phi(x)$.

O caso em que n é ímpar decorre do fato de que $\forall \epsilon > 0$

$$P\left[\left|\frac{S_{2n+1} - \frac{2n+1}{2}}{\sqrt{\frac{2n+1}{2}}} - \frac{S_{2n} - n}{\sqrt{\frac{n}{2}}}\right| \geq \epsilon\right] \xrightarrow[n\to\infty]{} 0$$

e do exercício 4.

EXERCÍCIOS

1. Se $X_n \xrightarrow{P} X$ então $\mathcal{L}(X_n) \to \mathcal{L}(X)$.

2. Se F_n e F são funções de distribuição e a e b, $a < b$, que são pontos de continuidade de F

$$F_n(b) - F_n(a) \to F(b) - F(a)$$

então $F_n \to F$.

3. Se $\{F_n\}_{n=1,2,\ldots}$ é uma sucessão de funções de distribuição e F é ≥ 0, crescente contínua à direita, $\lim\limits_{x\to-\infty} F(x) = 0$ $F_n(x) - F_n(-x) \to F(x) - F(-x)$ $\forall x$ ponto de continuidade da F e

$$\lim_{x\to+\infty} \limsup_n 1 - \{F_n(x) - F_n(-x)\} = 0$$

então F é uma função de distribuição.

4. Se $\forall \epsilon > 0$ $P[|X_n - Y_n| \geq \epsilon] \to 0$ e se $\mathcal{L}(X_n) \to F$ então $\mathcal{L}(Y_n) \to F$.

5. a) O número de descontinuidades de uma função de distribuição F e no máximo enumerável. (Portanto o conjunto de pontos de continuidade é denso).

   b) Se $F_n \to F$ então

   $$\forall a \in \mathbb{R} \quad \limsup_n F_n(a) \leq F(a)$$

   c) Se $F_n \to F$ e F é contínua em $a \in R$ $a_n \to a$ então $F_n(a_n) \to F(a)$.

   d) $F_n \to F$, a e b pontos de continuidade da F $a_n \to a$, $b_n \to b$ então

   $$[F_n(b_n) - F_n(a_n)] - [F(b_n) - F(a_n)] \xrightarrow[n\to\infty]{} 0$$

6. Provar que $\binom{2n}{n} \frac{1}{2^{2n}} \cong \frac{1}{\sqrt{\pi n}}$. (Sugestão: observe que o membro esquerdo é o têrmo central da distribuição binomial. Use o Teorema do Limite Central e o exercício 5 d)).

7. Provar que $\lim\limits_{n\to\infty} e^{-n} \sum\limits_{k=0}^{n} \frac{n^k}{n!} = \frac{1}{2}$.

   (Sugestão: Considerar variáveis aleatórias independendentes idênticamente distribuidas Poisson (1) e usar o Teorema do Limite Central).

8. Provar a Lei Fraca dos grandes números usando o Teorema do Limite Central.

9. Se $\{X_n\}_{n=1,2,\ldots}$ e X são variáveis aleatórias com valores no conjunto $\{x_1, x_2, \ldots\}$ e

$$p_{nk} = P[X_n = x_k] \to p_k = P[X = x_k] \quad \forall\, k = 1, 2, \ldots$$

então $\mathcal{L}(X_n) \to \mathcal{L}(X)$.

CAPÍTULO 9

INTRODUÇÃO AO PASSEIO AO ACASO

Êste capítulo tem como objetivo utilizar as técnicas e resultados dos capítulos anteriores em um modêlo particular de grande importância. Êste tema tem a peculiaridade de que com técnicas mais ou menos elementares é possível obter resultados bastante profundos e corrigir parcialmente a nossa intuição sôbre o comportamento das flutuações aleatórias. A referência natural para êste tema é [2].

9.1 - Notação e propriedades básicas

Nesta seção (Ω_n, P_n) denotará o seguinte espaço de probabilidades: $\Omega_n = \{-1,1\}^{I_n}$; se $\omega = \{x_1,\ldots,x_n\} \in \Omega_n$ então definimos $X_i(\omega) = x_i$ $i=1,\ldots,n$ $X_o = 0$ e $S_i = \sum_{j=0}^{i} X_j$ $i=0,1,\ldots n$; o número de 1 e -1 na sucessão ω (a e b respectivamente) estão dados pelas equações

$$a + b = n$$

$$a - b = S_n(\omega)$$

isto é

$$a = \frac{n + S_n(\omega)}{2} \qquad b = \frac{n - S_n(\omega)}{2}$$

Definimos para $0 < p < 1$, $q = 1-p$ $P_n(\omega) = p^a q^b$.

Em relação a P_n, $\{X_i\}_{i=1,\ldots,n}$ é um conjunto de variáveis aleatórias independentes e idênticamente distribuidas

$$P_n[X_i=1] = p \qquad P_n[X_i=-1] = 1-p = q \ .$$

Por trajetória associada ao ponto entendemos a função $i \mapsto S_i(\omega)$ $i=0,1,\ldots,n$; estas costumam ser representadas gràficamente por diagramas como o da Figura 9.1.1.

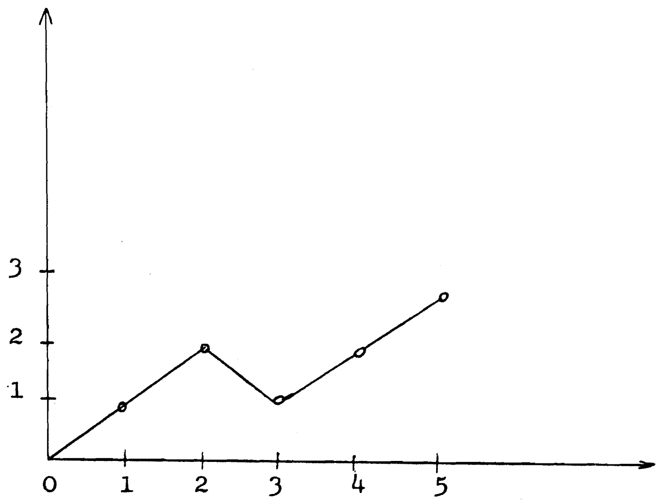


Figura 9.1.1

O experimento aleatório que sugere êste modêlo é o seguinte. Suponhamos uma partícula na origem da reta real. Por algum mecanismo aleatório selecionamos 1 ou -1 com probabilidade p e q respectivamente. Se o resultado é 1, a partícula se move uma unidade para a direita (de 0 a 1); se o resultado é -1, uma unidade para a esquerda (de 0 a -1). Se depois de i movimentos (tempo i) a partícula se encontra na posição k e o resultado é 1, a partícula se move para a posição $k+1$; se é -1, para a posição $k-1$. Note-se que S_i indica a posição no tempo i. Vamos calcular agora as probabilidades de alguns eventos que são de grande interêsse.

Começamos por

$$P_n[S_1 > 0,\ S_2 > 0, \ldots, S_n = x] \quad \text{onde} \quad x > 0.$$

Denotamos por $N_{(n,x),(n',x')}$ o número de trajetórias tais que $S_n = x$ $S_{n'} = x'$. Escreveremos $N_{(n,x)}$ em lugar de $N_{(0,0),(n,x)}$. Dados x e n o número de 1 e -1 (a e b respectivamente) está determinado. ($a = \frac{n+x}{2}$, $b = \frac{n-x}{2}$). Claramente $N_{(n,x)} = \binom{n}{\frac{n+x}{2}}$. Resulta que $P_n[S_1>0,\ldots,S_n=x]$ = (número de trajetórias desde (0,0) até (n,x) que são sempre >0) $p^a q^b = \{N_{(1,1),(n.x)}$ - [número de trajetórias desde (1,1) até (n,x) que tocam ou cruzam o eixo]} $p^a q^b = \{N_{(0,0),(n-1,x-1)}$ - [número de trajetórias desde (1,1) até (n,x) que tocam ou cruzam o eixo]}.

Para calcular o segundo membro da expressão entre colchetes usaremos o chamado <u>princípio da reflexão</u>.

Sejam A e B dois pontos que estão sôbre o eixo das abcissas (Ver figura 9.1.2).

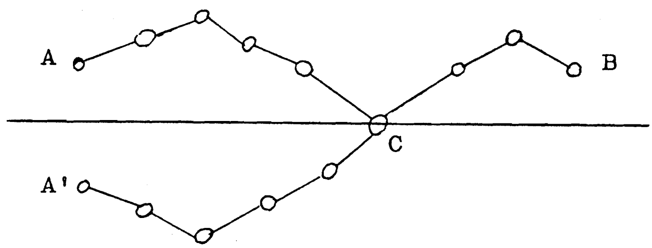


Figura 9.1.2

Dada uma trajetória que toca ou cruza o eixo e que vai desde A a B reflete-se o segmento que vai desde A até o primeiro zero da trajetória (ponto C na figura 9.1.2) Obtemos assim uma trajetória desde A' até B. É fácil ver que esta construção estabelece uma correspondência 1-1 entre as trajetórias que vão desde A a B e tocam ou cruzam o eixo e tôdas as trajetórias desde A' a B. Êste resultado é conhecido como princípio da reflexão. Temos então

$$P_n[S_1>0,\dots,S_n=x] = p^a q^b\{N_{(n-1,x-1)} - N_{(1,-1),(n,x)}\} =$$

$$= p^a q^b\{N_{(n-1,x-1)} - N_{(n-1,x+1)}\} =$$

$$= p^a q^b\{\binom{n-1}{\frac{n+x-1}{2}} - \binom{n-1}{\frac{n+x}{2}}\} = p^a q^b\{\binom{n-1}{a-1} - \binom{n-1}{a}\} .$$

Agora

$$\binom{n-1}{a-1} - \binom{n-1}{a} = \frac{(n-1)!}{(a-1)!(n-a)!} - \frac{(n-1)!}{a!(n-1-a)!} = \binom{n}{a}\frac{a-b}{n} \quad (9.1.1)$$

Êste é o número de trajetórias de $(0,0)$ até (n,x) que são sempre > 0.

Resulta então

$$P_n[S_1>0, S_2>0, \ldots, S_n=x] = \binom{n}{a} \frac{a-b}{n} p^a q^b .$$

Como $P_n[S_n=x] = \binom{n}{a} p^a q^b$ resulta que

$$P_n[S_1>0, S_2>0, \ldots, S_{n-1}>0 | S_n=x] = \frac{a-b}{n} = \frac{a-b}{a+b}$$

que é uma probabilidade condicional de interêsse.

Um problema relacionado com o cálculo anterior é o seguinte. Em uma eleição com dois candidatos, o candidato número 1 recebe a votos e o candidadto número 2 b votos com $a > b$ e $a+b = n$. Suponhamos que os votos são contados um a um. A probabilidade de que durante tôda a apuração o candidato número 1 leve vantagem é $\frac{a-b}{a+b}$. Isto segue do fato de que o número total de trajetórias possíveis é $\binom{n}{a}$ e o número de trajetórias desde $(0,0)$ até $(n,a-b)$ que estejam sempre acima do eixo é $\binom{n}{a} \frac{a-b}{a+b}$.

Vamos agora calcular

$$\begin{aligned} f_{2n} &= P_{2n} \text{ [o primeiro retôrno à origem ocorre no tempo} 2n] \\ &= P_{2n}[S_1 \neq 0, S_2 \neq 0, \ldots, S_{2n-1} \neq 0, S_{2n} = 0] \\ &= 2 \text{ [número de trajetórias de } (0,0) \text{ até } (2n,0) \text{ que são} \\ &\quad >0 \text{ exceto nos tempos } 0 \text{ e } 2n]\ p^n q^n = \end{aligned}$$

$= 2$ [número de trajetórias desde $(0,0)$ até $(2n-1,1)$ que são >0] $p^n q^n$

$$= 2\binom{n-1}{n} \frac{1}{2n-1} \, p^n q^n \quad \text{por} \quad (9.1.1)$$

$$= \frac{2(2n-1)!}{n!(n-1)!} \frac{1}{(2n-1)} \, p^n q^n = \frac{2}{n} \binom{2n-2}{n-1} (pq)^n \ .$$

Se o passeio ao acaso é simétrico podemos obter uma expressão aproximada para f_{2n} usando a Fórmula de Stirling. Temos

$$f_{2n} = \frac{2 \ (2n-2)^{(2n-2)} \ e^{-(2n-2)} \sqrt{2\pi(2n-2)}}{n 2^n \ (n-1)^{n-1} \ e^{-(n-1)} \sqrt{2\pi(n-1)} \ (n-1)^{n-1} e^{-(n-1)} \sqrt{2\pi(n-1)}}$$

$$= \frac{1}{2\sqrt{\pi} \ n\sqrt{n-1}} \cong \frac{1}{2\sqrt{\pi}} \frac{1}{n^{3/2}} \ .$$

Vamos agora calcular dado $\ell > 0$.

$P_{\ell+2k}$ [a primeira passagem por $\ell > 0$ ocorra no tempo $\ell+2k$] (Observe-se que $\ell+2k$, $k=0,1,\ldots$ são os únicos tempos nos quais uma primeira passagem pode ocorrer). Êste cálculo pode ser efetuado utilizando o chamado princípio de dualidade, que vamos utilizar sem formular precisamente. Se observamos a trajetória da Figura 9.1.3.

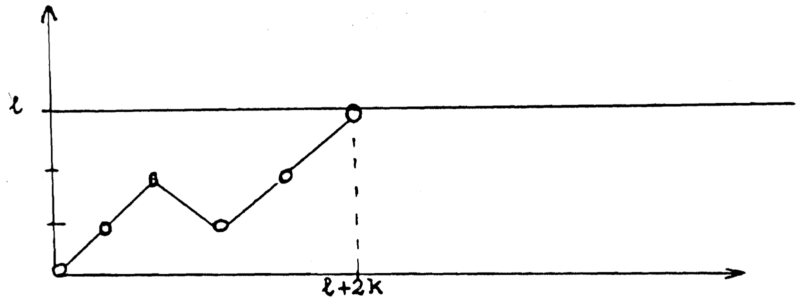


Figura 9.1.3

vemos que está sempre abaixo de ℓ e que percorre uma distância vertical de ℓ unidades em $\ell+2k$ unidades de tempo. Se a observamos em sentido contrário, temos uma trajetória >0 que chega no tempo $\ell+2k$ a $\ell > 0$. Portanto o número de trajetórias contidas no evento é igual ao número de trajetórias desde $(0,0)$ até $(\ell+2k,\ell)$ que são >0 e então a probabilidade procurada é

$$\binom{\ell+2k}{a} \frac{a-b}{\ell+2k} p^a q^b \quad \text{onde} \quad a+b = \ell+2k \quad \text{e} \quad a-b = \ell.$$

Resulta então que esta probabilidade é igual a

$$\binom{\ell+2k}{\ell+k} \frac{\ell}{\ell+2k} p^{\ell+k} q^k = \binom{\ell+2k}{k} \frac{\ell}{\ell+2k} p^{\ell+k} q^k .$$

9.2 - Passeio ao acaso simétrico

Vamos a considerar desde agora em diante só passeios ao ocaso simétricos. $(p = q = \frac{1}{2})$.

PROPOSIÇÃO 9.2.1

$$P_{2n}[S_1\geq 0 \dots S_{2n}\geq 0] = P_{2n}[S_1\neq 0,\dots,S_{2n}\neq 0]$$

<u>Demonstração</u>:

$$P_{2n}[S_1\neq 0,\dots,S_{2n}\neq 0] = 2P_{2n}[S_1>0,\dots,S_{2n}>0] =$$

$= \frac{2}{2^{2n}}$ [número de trajetórias desde (0,0), > 0] =

$= \frac{2}{2^{2n}}$ [número de trajetórias desde (1,1), ≥ 1] =

$= \frac{1}{2^{2n-1}}$[número de trajetórias desde (0,0), ≥ 0 desde 0 até 2n-1] =

$$= P_{2n-1}[S_1\geq 0,\dots,S_{2n-1}\geq 0] = P_{2n}[S_1\geq 0,\dots,S_{2n}\geq 0]$$

(porque S_{2n-1} é sempre $\neq 0$).

Denotemos com $u_{2n} = P_{2n}[S_{2n}=0]$.

PROPOSIÇÃO 9.2.2

$$P_{2n}[S_1\neq 0,\dots,S_{2n}\neq 0]= P[S_{2n}=0] = u_{2n}$$

<u>Demonstração</u>:

$$P_{2n}[S_1\neq 0,\dots,S_{2n}\neq 0] = 2P[S_1>0,\dots,S_{2n}>0] =$$

$$= 2\sum_{j=1}^{n} P_{2n}[S_1>0,\dots,S_{2n}=2j] =$$

$$= 2\sum_{j=1}^{n} (N_{2n-1,2j-1} - N_{2n-1,2j+1}) \frac{1}{2^{2n}} =$$

(ver demonstração da fórmula (9.1.1))

$$= \frac{2}{2^{2n}}[(N_{2n-1,1}-N_{2n-1,3})+(N_{2n-1,3}-N_{2n-1,5})+\dots+(N_{2n-1,2n-1})]$$

$$= \frac{2}{2^{2n}} N_{2n-1,1} = \frac{N_{2n,0}}{2^{2n}} = P_{2n}[S_{2n}=\emptyset] = u_{2n} \quad .$$

Vamos agora a estudar a distribuição e comportamento limite da variável introduzida na seguinte

DEFINIÇÃO 9.2.1 - $M_n = \max\limits_{1\leq i\leq n} S_i$.

PROPOSIÇÃO 9.2.3

$$p_{n,k} = P_n[M_n=k] = \frac{1}{2^n}\binom{n}{[\frac{n-k}{2}]} \quad k=0,1,\ldots,n$$

Demonstração: Definimos $\bar{M}_n = \max\limits_{2\leq i\leq n+1} S_i$; para $k\geq 1$ temos

$$p_{n+1,k} = P_{n+1}[M_{n+1}=k] = P_{n+1}[X_1=1, M_{n+1}=k]$$

$$+ \; P_{n+1}[X_1=-1, M_{n+1}=k] = P_{n+1}[X_1=1 \; \bar{M}_n = k-1]$$

$$+ \; P_{n+1}[X_1=-1, \bar{M}_n=k+1] = \frac{1}{2}(p_{n,k-1} + p_{n,k+1}) \; .$$

A última igualdade decorre do fato que X_1 é independente de $\bar{M}_n$ e de que $\mathcal{L}(M_n) = \mathcal{L}(\bar{M}_n)$.

Para $k = 0$ temos também

$$p_{n+1,0} = P_{n+1}[X_1=-1, \bar{M}_n\leq 1] = \frac{1}{2} P_n[M_n\leq 1] =$$

$$= \frac{1}{2}\{P_n[M_n=1] + P_n[M_n=0]\} = \frac{1}{2}(p_{n,1} + p_{n,0})$$

Ou seja provamos as duas seguintes fórmulas

$$p_{n+1,0} = \frac{1}{2}(p_{n,1} + p_{n,0}) \tag{9.2.1}$$

$$p_{n+1,k} = \frac{1}{2}(p_{n,k-1} + p_{n,k+1}) \tag{9.2.2}$$

Agora

$$p_{n,0} = P_n[M_n=0] = P_n[S_1\le 0,\dots,S_n\le 0] =$$

$$= P_{2[\frac{n+1}{2}]}[S_1\le 0,\dots,S_{2[\frac{n+1}{2}]}\le 0] = \binom{2[\frac{n+1}{2}]}{[\frac{n+1}{2}]}\frac{1}{2^{2[\frac{n+1}{2}]}} =$$

$$= \frac{1}{2^n}\binom{n}{[\frac{n}{2}]}.$$

A penúltima igualdade é uma conseqüência das Proposições 9.2.1 e 9.2.2.

A fórmula (9.2.1) permite calcular $p_{n,1}$ conhecendo $p_{n,0}$ e a fórmula (9.2.2) calcular $p_{n,2}$ conhecendo $p_{n,0}$ e $p_{n,1}$ etc. Ou seja, $p_{n,0}$ determina $p_{n,k}$ para todo $k\ge 1$. Agora é fácil calcular $p_{n,k}$ ou então verificar diretamente que a função $u_{n,k} = \frac{1}{2^n}\binom{n}{[\frac{n-k}{2}]}$ satisfaz as equações (9.1.1) e (9.1.2) e $u_{n,0} = p_{n,0}$. Portanto $u_{n,k} = p_{n,k}$ $n\ge 1$, $0 \le k \le n$.

PROPOSIÇÃO 9.2.4

$$\forall\ x \ge 0 \quad P_n[\frac{M_n}{n} \le x] \to \sqrt{\frac{2}{\pi}}\int_0^x e^{-\frac{u^2}{2}}\,du$$

Demonstração:

$$P_n[\frac{M_n}{n} \le x] = \frac{1}{2^n}\sum_{k<x\sqrt{n}}\binom{n}{[\frac{n-k}{2}]} \quad .$$

Mudando de índice temos

$$\frac{2}{2^n} \sum_{\frac{n-x\sqrt{n}}{2} < j < \frac{n}{2}} \binom{n}{j} - \frac{\delta_n}{2^n} \binom{n}{[\frac{n}{2}]}$$

onde $\delta_n = 1$ se n é par e 0 se n é ímpar.

Como

$$\binom{n}{[\frac{n}{2}]} \frac{1}{2^n} \cong \sqrt{\frac{2}{\pi n}}$$

temos

$$\lim_{n\to\infty} P_n[\frac{M_n}{n} \leq x] = 2 \lim_{n\to\infty} \frac{1}{2^n} \sum_{\frac{n-x\sqrt{n}}{2} < j < \frac{n}{2}} \binom{n}{j} =$$

$$= 2 \frac{1}{\sqrt{2\pi}} \int_{-x}^{0} e^{-\frac{u^2}{2}} du = \sqrt{\frac{2}{\pi}} \int_{0}^{x} e^{-\frac{u^2}{2}} du \quad .$$

9.3 - Uma introdução informal à lei do arco seno

Vamos definir agora uma sucessão de variáveis aleatórias da seguinte maneira:

τ_n = número de têrmos na sucessão $S_1, \ldots, S_n$ que são positivos ou que são iguais a 0 mas o têrmo anterior é positivo.

A prova da seguinte proposição pode ser encontrada em [2], Capítulo 2.

PROPOSIÇÃO 9.3.1

$$P_{2n}[\tau_{2n}=2k] = P_{2k}[S_{2k}=0] \quad P_{2(n-k)}[S_{2(n-k)}=0]$$

$$= \frac{\binom{2k}{k}\binom{2n-k}{n-k}}{2^{2n}} .$$

Note que esta função tem a forma de U e portanto os valores mais prováveis são 0 e 2n; os menos prováveis são os valores próximos a n. Pelo exercício 6 do Capítulo 8 (ou pela Fórmula de Stirling diretamente) esta última expressão é aproximadamente igual a

$$\frac{1}{\sqrt{\pi k}\,\sqrt{\pi(n-k)}} = \frac{1}{\pi n \sqrt{\frac{k}{n}(1-\frac{k}{n})}} .$$

Se $0 < a < b < 1$ e

$$g(y) = \begin{cases} \dfrac{1}{\pi\sqrt{y(1-y)}} & 0 < y < 1 \\ 0 & \begin{matrix} y \leq 0 \\ y \geq 1 \end{matrix} \end{cases}$$

é natural pensar que

$$P[a < \frac{\tau_{2n}}{2n} < b] \xrightarrow[n\to\infty]{} \int_a^b g(y)dy .$$

Como $\int_0^1 g(y)dy = 1$ $\left(\frac{d}{dy} (\text{arc sen} \sqrt{y}) = \frac{1}{2\sqrt{y(1-y)}} \right)$ resulta então que a função

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \displaystyle\int_0^x g(y)dy = \frac{2}{\pi} \text{ arc sen } \sqrt{x} & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}$$

é uma função de distribuição e

$$P[a < \frac{\tau_{2n}}{2n} < b] \to F(b)-F(a) \quad \forall\, 0 < a < b < 1$$

É fácil ver então que

$$\lim_{n\to\infty} P[\frac{\tau_{2n}}{2n} \leq x] = F(x) .$$

Como $0 \leq \tau_{2n+1} - \tau_{2n} \leq 1$ temos em geral que

$$P[\frac{\tau_n}{n} \leq x] \xrightarrow[n\to\infty]{} 0 \quad \frac{2}{\pi} \text{ arc sen } \sqrt{x}$$

que é a chamada lei do arco seno.

Note que a função g tem forma de U e é simétrica com respeito ao ponto $\frac{1}{2}$ no qual assume o seu valor mínimo. A Proposição 9.3.1 e a lei do arco seno dão idéia de que a intuição a respeito das flutuações aleatórias pode estar equivocada.

EXERCÍCIOS

1. Provar que

$$P[S_1 \geq 0, \ldots, S_{2n-1} \geq 0, S_{2n}=0] = \frac{P[S_{2n}=0]}{n+1}$$

$$(P[S_{2n}=0] = \binom{2n}{n}(pq)^n) \ .$$

2. Seja $p = q = \frac{1}{2}$. Provar que

a) $h_{2n} = \dfrac{u_{2n-2}}{2n}$ $\qquad (u_{2n} = \binom{2n}{n}\dfrac{1}{2^{2n}})$

b) $\dfrac{u_{2n}}{u_{2n-2}} = 1 - \dfrac{1}{2n}$ $\qquad$ (portanto $h_{2n} = u_{2n-2} - u_{2n}$)

3. Provar que $\sum\limits_{n=1}^{\infty} 2n\, f_{2n} = \infty$ (o tempo médio para retornar à origem é infinito).

(<u>Sugestão</u>: usar a aproximação de f_{2n} dada pela fórmula de Stirling).

4. Dois jogadores jogam uma moeda equilibrada n vêzes. Provar que a probabilidade de que ambos os jogadores obtenham o mesmo número de caras é $\dfrac{1}{2^{2n}} \sum\limits_{k=0}^{n} \binom{n}{k}^2$.

5. Provar que $\sum\limits_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}$.

ANEXO 1

ALGUNS RESULTADOS SÔBRE SÉRIES NUMÉRICAS

DEFINIÇÃO 1 - Dada uma sucessão $\{a_i\}_{i=0,1,\ldots}$, seja $S_n = \sum_{i=0}^{n} a_i$. A sucessão de <u>somas parciais</u> $\{S_n\}_{n=0,1,\ldots}$, denominamos série $\sum_{k=0}^{\infty} a_k$, onde a_n é o <u>enésimo</u> (n-gésimo) têrmo da série.

DEFINIÇÃO 2 - Se $\lim_{n\to\infty} S_n = \ell$, dizemos que ℓ é a soma da série, e escrevemos $\sum_{k=0}^{\infty} a_k = \ell$. Caso a série seja de têrmos positivos, escrevemos $\sum_{k=0}^{\infty} a_k = +\infty$, se $\lim_{n\to\infty} S_n = +\infty$.

DEFINIÇÃO 3 - Dizemos que a série $\sum_{k=0}^{\infty} a_k$ é <u>absolutamente, convergente</u> se $\lim_{n\to\infty} \sum_{i=0}^{\infty} |a_i| < +\infty$. Escrevemos então, $\sum_{k=0}^{\infty} |a_k| < +\infty$.

É simples verificar que a <u>convergência absoluta</u> implica <u>convergência</u>.

Seja $\mathbb{N} = \{0,1,\ldots\}$, e $h:\mathbb{N} \to \mathbb{N}$ uma bijeção. A <u>sucessão</u> $\{a_{h(n)}\}_{n=0,1,\ldots}$ é dita um <u>reordenamento</u> da <u>sucessão</u> $\{a_n\}_{n=0,1,\ldots}$, e a série $\sum_{u=0}^{\infty} a_{h(n)}$ um <u>reordenamento</u> da série $\sum_{k=0}^{\infty} a_k$.

TEOREMA 1 - Se uma série é absolutamente convergente (ou de têrmos positivos), qualquer de seus <u>reordenamentos</u> é absolutamente convergente (convergente), e tôdas as séries obtidas por <u>reordenação</u> convergem para o mesmo limite. Recìprocamente, se todo <u>reordenamento</u> converge para um mesmo limite finito, a série é absolutamente convergente.

DEFINIÇÃO 4 - Dizemos que uma série dupla $\sum\limits_{n=0}^{\infty} \sum\limits_{m=0}^{\infty} a_{nm}$ é <u>absolutamente convergente</u>, se $\exists M$ tal que $\forall k, \forall s, \sum\limits_{n=0}^{k} \sum\limits_{m=0}^{s} |a_{nm}| \leq M$.

Se $h: \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ é uma bijeção, a sucessão (a série) $\{a_{h(n)}\}_{n=0,1,\ldots}$ $(\sum\limits_{n=0}^{\infty} a_{h(n)})$ é dita um <u>reordenamento</u> da sucessão $\{a_{nm}\}\ \begin{matrix} n=0,1,\ldots \\ m=0,1,\ldots \end{matrix}$ (da série $\sum\limits_{n} \sum\limits_{m} a_{nm}$).

TEOREMA 2 - Se a série $\sum\limits_{n} \sum\limits_{m} a_{nm}$ converge absolutamente (ou é de têrmos positivos), então qualquer reordenamento tem a mesma soma. Êsse valor comum é denominado <u>soma da série</u>.

TEOREMA 3 - Se a série $\sum\limits_{n} \sum\limits_{m} a_{nm}$ é absolutamente convergente (ou de têrmos positivos), então $\forall n, \forall m$, $\sum\limits_{m=0}^{\infty} a_{nm}$ e $\sum\limits_{n=0}^{\infty} a_{nm}$ são absolutamente convergentes (convergentes), e $\sum\limits_{n=0}^{\infty} \sum\limits_{m=0}^{\infty} a_{nm} = \sum\limits_{n=0}^{\infty} \left[\sum\limits_{m=0}^{\infty} a_{nm}\right] = \sum\limits_{m=0}^{\infty} \left[\sum\limits_{n=0}^{\infty} a_{nm}\right]$.

TEOREMA 4 - Seja $\{a_i\}_{i=0,1,\ldots}$ uma sucessão alternada $(a_0 \geq 0,\ a_1 \leq 0,\ a_2 \geq 0, \ldots)$ tal que $|a_{i+1}| \geq |a_i|$ e

$|a_i| \to 0$ quando $i \to \infty$. Então $S_n = \sum_{i=0}^{n} a_i$ converge.

EXERCÍCIOS

1. Calcular $\sum_{n=1}^{\infty} n\,\rho^n$, onde $0 \le \rho < 1$, usando o Teorema 3.

2. Provar que $\sum_{k=0}^{\infty} (k+r-1)\ldots\ldots(k+1)q^k = \frac{(r-1)!}{(1-q)^r}$ para $|q| < 1$, $r = 1,2,\ldots$.

ANEXO 2

A apresentação dêste material clássico está baseada essêncialmente em [4]. Seja X uma variável aleatória sôbre (Ω,P) um espaço de probabilidades. Definimos para $p > 1$ a norma-p de X como o número

$$\|X\|_p = (E|X|^p)^{\frac{1}{p}}$$

Antes de provar o Teorema 2, necessitamos alguns resultados prévios.

TEOREMA 1 (Desigualdade de Young) - Seja $\varphi:[0,+\infty) \to [0,+\infty)$ contínua, estritamente crescente, $\varphi(u) \xrightarrow[n\to+\infty]{} +\infty$, $\varphi(0) = 0$. Seja $\psi = \varphi^{-1}$. Definamos para todo $x \geq R$

$$\phi(x) = \int_0^x \varphi(u)du$$

$$\Psi(x) = \int_0^x \psi(v)dv$$

Então para todo $a,b \in [0,+\infty)$

$$ab \leq \phi(a) + \Psi(b)$$

a igualdade vale se e sòmente se $b = \varphi(a)$.

<u>Demonstração</u>: Claramente

$$\int_0^a \varphi(u)du + \int_0^{\varphi(a)} \psi(v)dv = a\,\varphi(a)$$

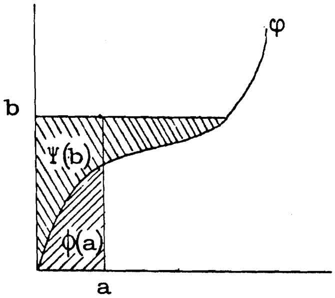
Isto é

$$\phi(a) + \Psi(\varphi(a)) = a\,\varphi(a)$$

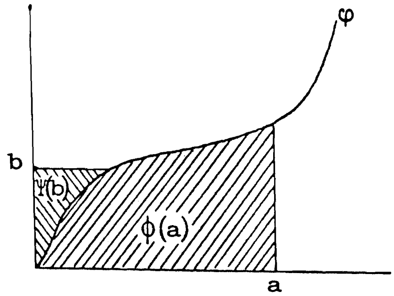
De aqui

$$\Phi(a) + \Psi(b) = a\,\varphi(a) + \Psi(b) - \Psi(\varphi(a))$$

A desigualdade decorre agora considerando os casos $\varphi(a) \leq b$ e $\varphi(a) > b$ nas figuras



$\varphi(a) \leq b$



$\varphi(a) > b$

Para $p > 0$, $p \neq 1$, definimos $q = \frac{p}{p-1}$. Então $\frac{1}{p} + \frac{1}{q} = 1$.

COROLÁRIO 1 - Para $p > 1$, $a \geq 0$, $b \geq 0$

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

com igualdade se e sòmente se $a^p = b^q$.

<u>Demonstração</u>: Seja $\varphi(u) = u^{p-1}$. Então

$$\phi(a) = \int_0^a u^{p-1}\,du = \left.\frac{u^p}{p}\right|_0^a = \frac{a^p}{p} \quad ,$$

$$\psi(v) = v^{\frac{1}{p-1}}, \quad e \quad \Psi(b) = \int_0^b v^{\frac{1}{p-1}}\, dv = \frac{v^q}{q}\Big|_0^b = \frac{b^q}{q}.$$

Por o teorema anterior temos $ab \le \frac{a^p}{p} + \frac{b^q}{q}$ com igualdade se e sòmente se $b = \varphi(a) = a^{p-1}$. Isto é

$$b^{\frac{1}{p-1}} = a \quad \text{que é equivalente a}$$

$$b^{\frac{p}{p-1}} = b^q = a^p .$$

Sejam agora X e Y duas variáveis aleatórias sôbre (Ω, P).

TEOREMA 2 (Desigualdade de Hölder para $p>1$) - Seja

$$X \in L_p \quad e \quad Y \in L_q \quad \text{onde} \quad p > 1 \quad e \quad q = \frac{p}{p-1}.$$

Então

$$E|XY| \le \|X\|_p \, \|Y\|_q$$

a igualdade vale se e sòmente se $\exists\ A \ge 0$ e $B > 0$ tal que $A|X|^p = B|Y|^q$.

<u>Demonstração</u>: Se $P[X=0] = 1$ ou $P[Y=0] = 1$ a desigualdade é trivial. Se não é êste o caso pelo corolário anterior

$$\frac{|X|}{\|X\|_p} \frac{|Y|}{\|Y\|_q} \le \frac{1}{p} \frac{|X|^p}{\|X\|_p^p} + \frac{1}{q} \frac{|Y|^q}{\|Y\|_q^q}$$

e tomando esperanças resulta

$$\frac{E|X|\,|Y|}{\|X\|_p \, \|Y\|_q} \le \frac{1}{p} + \frac{1}{q} = 1$$

ou seja $E|XY| \le \|X\|_p \, \|Y\|_q$.

A última afirmação decorre imediatamente do corolá rio.

No caso em que $p = q = 2$ obtemos a chamada desigualdade de Cauchy-Schwarz

$$E|XY| \le E^{1/2}|X|^2 \, E^{1/2}|Y|^2$$

REFERÊNCIAS

[1] - LUBELL, A short proof of Sperner's lemma. Introduction to Combinatorial Theory, vol.1 1966, pg. 199.

[2] - FELLER WILLIAM, An Introduction to probability theory and its applications, vol. 1, 1950 John Wiley & Sons, Inc.

[3] - BOURBAKI N. (1968) Theory of Sets, Hermann, Paris, France.

[4] - EDWIN HEWITT, KARL STROMBERG (1965) Real and Abstract Analysis, Springer-Verlag.

[5] - LUCIEN LE CAM, (1963). On the distributions of sums of Independent Random Variables, Proceedings of an International Research Seminar. Statistical Laboratory. University of California, Berkeley, Springer-Verlag.

[6] - ARAM J. THOMASIAN (1969) The Structure of Probability Theory with Applications. Mc Graw-Hill.

[7] - PAUL M. MEYER (1965) Introductory Probability with Statistical Applications. Addison-Wesley Publishing Company Inc.

[8] - PROHOROV, UY V. (1953) Asymptotic behavior of the binomial distribution. Uspehi Mat. Nauk. 8, 135.

TABELA Nº 1

Distribuição normal

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$$

x.	0,00.	0,02.	0,04.	0,06.	0,08.
0,0	0,5000	0,5080	0,5160	0,5239	0,5319
0,1	0,5398	0,5478	0,5557	0,5636	0,5714
0,2	0,5793	0,5871	0,5948	0,6026	0,6103
0,3	0,6179	0,6255	0,6331	0,6406	0,6480
0,4	0,6554	0,6628	0,6700	0,6772	0,6844
0,5	0,6915	0,6985	0,7054	0,7123	0,7190
0,6	0,7257	0,7324	0,7389	0,7454	0,7517
0,7	0,7580	0,7642	0,7704	0,7764	0,7823
0,8	0,7881	0,7939	0,7995	0,8051	0,8106
0,9	0,8159	0,8212	0,8264	0,8315	0,8365
1,0	0,8413	0,8461	0,8508	0,8554	0,8599
1,1	0,8643	0,8686	0,8729	0,8770	0,8810
1,2	0,8849	0,8888	0,8925	0,8962	0,8997
1,3	0,9032	0,9066	0,9099	0,9131	0,9162
1,4	0,9192	0,9222	0,9251	0,9279	0,9306
1,5	0,9332	0,9357	0,9382	0,9406	0,9429
1,6	0,9452	0,9474	0,9495	0,9515	0,9535
1,7	0,9554	0,9573	0,9591	0,9608	0,9625
1,8	0,9641	0,9656	0,9671	0,9[illegible]86	0,9699
1,9	0,9713	0,9726	0,9738	0,9[illegible]50	0,9761
2,0	0,9772	0,9783	0,9793	0,9803	0,9812
2,1	0,9821	0,9830	0,9838	0,9846	0,9854
2,2	0,9861	0,9868	0,9875	0,9881	0,9887
2,3	0,9893	0,9898	0,9904	0,9909	0,9913
2,4	0,9918	0,9922	0,9927	0,9931	0,9934
2,5	0,9938	0,9941	0,9945	0,9948	0,9951
2,6	0,9953	0,9956	0,9959	0,9961	0,9963
2,7	0,9965	0,9967	0,9969	0,9971	0,9973
2,8	0,9974	0,9976	0,9977	0,9979	0,9980
2,9	0,9981	0,9982	0,9984	0,9985	0,9986
3,0	0,9987	0,9987	0,9988	0,9989	0,9990

TABELA Nº 2

Distribuição de Poisson

$$P_\lambda(k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

k \ λ	0,1.	0,2,	0,3,	0,4.	0,5.	0,6.
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488
1	0,0905	0,1637	0,2222	0,2681	0,3033	0,3293
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198
4	—	0,0001	0,0002	0,0007	0,0016	0,0030
5	—	—	—	0,0001	0,0002	0,0004
6	—	—	—	—	—	—

k \ λ	0,7.	0,8.	0,9.	1,0.	1,5.	2,0.
0	0,4966	0,4493	0,4066	0,3679	0,2231	0,1353
1	0,3476	0,3595	0,3659	0,3679	0,3347	0,2707
2	0,1217	0,1438	0,1647	0,1839	0,2510	0,2707
3	0,0284	0,0383	0,0494	0,0613	0,1255	0,1804
4	0,0050	0,0077	0,0111	0,0153	0,0471	0,0902
5	0,0007	0,0012	0,0020	0,0031	0,0141	0,0361
6	0,0001	0,0002	0,0003	0,0005	0,0035	0,0120
7	—	—	-	0,0001	0,0008	0,0034
8	—	—	—	—	0,0001	0,0009
9	—	—	—	—	—	0,0002

k \ λ	2,5.	3,0.	3,5.	4,0.	4,5.	5,0.
0	0,0821	0,0498	0,0302	0,0183	0,0111	0,0067
1	0,2052	0,1494	0,1057	0,0733	0,0500	0,0337
2	0,2565	0,2240	0,1850	0,1465	0,1125	0,0842
3	0,2138	0,2240	0,2158	0,1954	0,1687	0,1404
4	0,1336	0,1680	0,1888	0,1954	0,1898	0,1755
5	0,0668	0,1008	0,1322	0,1563	0,1708	0,1755
6	0,0278	0,0504	0,0771	0,1042	0,1281	0,1462
7	0,0099	0,0216	0,0385	0,0595	0,0824	0,1044
8	0,0031	0,0081	0,0169	0,0298	0,0463	0,0653
9	0,0009	0,0027	0,0066	0,0132	0,0232	0,0363
10	0,0002	0,0008	0,0023	0,0053	0,0104	0,0181
11	—	0,0002	0,0007	0,0019	0,0043	0,0082
12	—	0,0001	0,0002	0,0006	0,0016	0,0034
13	—	—	0,0001	0,0002	0,0006	0,0013
14	—	—	—	0,0001	0,0002	0,0005
15	—	—	—	—	0,0001	0,0002

TABELA Nº 3

Distribuição binomial

$$b_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

n.	k. \ p.	0,05.	0,10.	0,20.	0,30.	0,40.	0,50.
4	0	0,8145	0,6561	0,4096	0,2401	0,1296	0,0625
	1	0,1715	0,2916	0,4096	0,4116	0,3456	0,2500
	2	0,0135	0,0486	0,1536	0,2646	0,3456	0,3750
	3	0,0005	0,0036	0,0256	0,0756	0,1536	0,2500
	4	0,0000	0,0001	0,0016	0,0081	0,0256	0,0625
6	0	0,7351	0,5314	0,2621	0,1176	0,0467	0,0156
	1	0,2321	0,3543	0,3932	0,3025	0,1866	0,0938
	2	0,0305	0,0984	0,2458	0,3241	0,3110	0,2344
	3	0,0021	0,0146	0,0819	0,1852	0,2765	0,3125
	4	0,0001	0,0012	0,0154	0,0595	0,1382	0,2344
	5	0,0000	0,0001	0,0015	0,0102	0,0369	0,0938
	6	0,0000	0,0000	0,0001	0,0007	0,0041	0,0156
8	0	0,6634	0,4305	0,1678	0,0576	0,0168	0,0039
	1	0,2793	0,3826	0,3355	0,1977	0,0896	0,0312
	2	0 0515	0,1488	0,2936	0,2965	0,2090	0,1094
	3	0,0054	0,0331	0,1468	0,2541	0,2787	0,2188
	4	0,0004	0,0046	0,0459	0,1361	0,2322	0,2734
	5	0,0000	0,0004	0,0092	0,0467	0,1239	0,2188
	6	0,0000	0,0000	0,0011	0,0100	0,0413	0,1094
	7	0,0000	0,0000	0,0001	0,0012	0,0079	0,0312
	8	0,0000	0,0000	0,0000	0,0001	0,0007	0,0039
10	0	0,5987	0,3487	0,1074	0,0282	0,0060	0,0010
	1	0,3151	0,3874	0,2684	0,1211	0,0403	0,0098
	2	0,0746	0,1937	0,3020	0,2335	0,1209	0,0439
	3	0,0105	0,0574	0,2013	0,2668	0,2150	0,1172
	4	0,0010	0,0112	0,0881	0,2001	0,2508	0,2051
	5	0,0001	0,0015	0,0264	0,1029	0,2007	0,2461
	6	0,0000	0,0001	0,0055	0,0368	0,1115	0,2051
	7	0,0000	0,0000	0,0008	0,0090	0,0425	0,1172
	8	0,0000	0,0000	0,0001	0,0014	0,0106	0,0439
	9	0,0000	0,0000	0,0000	0,0001	0,0016	0,0098
	10	0,0000	0,0000	0,0000	0,0000	0,0001	0,0010

Distribuição binomial

(continuação)

n.	k. \ p.	0,05.	0,10.	0,20.	0,30.	0,40.	0,50.
15	0	0,4633	0,2059	0,0352	0,0047	0,0005	0,0000
	1	0,3658	0,3432	0,1319	0,0305	0,0047	0,0005
	2	0,1348	0,2669	0,2309	0,0916	0,0219	0,0032
	3	0,0307	0,1285	0,2501	0,1700	0,0634	0,0139
	4	0,0049	0,0428	0,1876	0,2186	0,1268	0,0417
	5	0,0006	0,0105	0,1032	0,2061	0,1859	0,0916
	6	0,0000	0,0019	0,0430	0,1472	0,2066	0,1527
	7	0,0000	0,0003	0,0138	0,0811	0,1771	0,1964
	8	0,0000	0,0000	0,0035	0,0348	0,1181	0,1964
	9	0,0000	0,0000	0,0007	0,0116	0,0612	0,1527
	10	0,0000	0,0000	0,0001	0,0030	0,0245	0,0916
	11	0,0000	0,0000	0,0000	0,0006	0,0074	0,0417
	12	0,0000	0,0000	0,0000	0,0001	0,0016	0,0139
	13	0,0000	0,0000	0,0000	0,0000	0,0003	0,0032
	14	0,0000	0,0000	0,0000	0,0000	0,0000	0,0005
	15	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
20	0	0,3585	0,1216	0,0115	0,0008	0,0000	0,0000
	1	0,3774	0,2702	0,0576	0,0068	0,0005	0,0000
	2	0,1887	0,2852	0,1369	0,0278	0,0031	0,0002
	3	0,0596	0,1901	0,2054	0,0716	0,0123	0,0011
	4	0,0133	0,0898	0,2182	0,1304	0,0350	0,0046
	5	0,0022	0,0319	0,1746	0,1789	0,0746	0,0148
	6	0,0003	0,0089	0,1091	0,1916	0,1244	0,0370
	7	0,0000	0,0020	0,0545	0,1643	0,1659	0,0739
	8	0,0000	0,0004	0,0222	0,1144	0,1797	0,1201
	9	0,0000	0,0001	0,0074	0,0654	0,1597	0,1602
	10	0,0000	0,0000	0,0020	0,0308	0 1171	0,1762
	11	0,0000	0,0000	0,0005	0,0120	0,0710	0,1602
	12	0,0000	0,0000	0,0001	0 0039	0,0355	0,1201
	13	0 0000	0,0000	0,0000	0,00 0	0,0146	0,0739
	14	0,0000	0,0000	0 0000	0,0002	0,0049	0,0370
	15	0 0000	0,0000	0,0000	0,0000	0,0013	0 0148
	16	0 0000	0,0000	0,0000	0 0000	0,0003	0,0046
	17	0 0000	0,0000	0,0000	0,0000	0,0000	0,0011
	18	0 0000	0,0000	0,0000	0000	0,0000	0,0002
	19	0 0000	0,0000	0,0000	0,0000	0,0000	0,0000
	20	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

NOTAS AOS CAPÍTULOS

CAPÍTULO 1

Exercício 4 - Seja $\{x_1, x_2, \ldots\}$ uma enumeração dos racionais de $\mathbb{R}$. Seja $A_n = (x_n - 1, x_n + 1)$. Calcular $\lim_n \sup A_n$ e $\lim_n \inf A_n$.

CAPÍTULO 2

Talvez seja conveniente remarcar que uma das vantagens do tratamento formal da análise conbinatória é que permite, por exemplo, dizer explìcitamente qual é o espaço amostral de muitos experimentos e cremos também que evita erros (ou pelo menos os torna mais difíceis) nos cálculos das probabilidades de certos eventos.

A demonstração da Proposição 2.8.1 pode ser descrita em linguagem mais intuitiva. Seja ω um ponto fixo de um conjunto Ω contendo n+1 elementos. Os subconjuntos de Ω com k elementos (que são $\binom{n+1}{k}$) podem ser classificados em dois grupos; os que contém ω e os que não. O primeiro grupo tem $\binom{n}{k-1}$ conjuntos; o segundo $\binom{n}{k}$. Portanto $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$.

EXERCÍCIOS

13. Uma urna contém n-1 bolinhas brancas e uma preta. Se fazem extrações sem reposição até esgotar o conteúdo

da urna. Calcular a probabilidade de escolher a bolinha preta na extração i, $1 \leq i \leq n$.

CAPÍTULO 3

EXERCÍCIOS

14. No Exemplo 3.4.2, tomemos como espaço de probabilidades $C^R_{r,n}$ com a probabilidade usual. Neste espaço os casos favoráveis se identificam com $C'_{r,n}$. Mas a probabilidade de $C'_{r,n}$ neste espaço é diferente da calculada no Exemplo 3.4.2. Onde esta a falacia?

CAPÍTULO 4

EXERCÍCIOS

13. Seja (Ω, P) um espaço de probabilidades e $\{A_n\}_{n=1,2,\ldots}$ uma sucessão de conjuntos independentes. Para $0 \leq x \leq 1$ seja $\varphi(x) = x \wedge (1-x)$. Então $\sum_{n=1}^{\infty} \varphi(P(A_n)) < \infty$.

CAPÍTULO 6

PROPOSIÇÃO 6.1.7 - $0 \leq X_n \uparrow X$ então $E\, X_n \uparrow EX$.

<u>Demonstração</u>: Seja $U_{n,m}$ uma sucessão de variáveis aleatórias simples tais que $0 \leq U_{n,m} \uparrow X_n$.

Seja $V_m = \sup_{1 \leq n \leq m} U_{n,m}$; temos que $V_m \leq V_{m+1}$ e $U_{n,m} \leq$ $\leq V_m \leq X_m$ se $m \geq n$, portanto $E\, U_{n,m} \leq E\, V_m \leq E\, X_m$, $m \geq n$. Se $m \to \infty$ obtemos $X_n \leq \lim_m V_m \leq X$ e $E\, X_n \leq \lim_m E\, V_m \leq$

$\leq \lim_m E X_m$; se agora $n \to \infty$ temos $X \leq \lim_m V_m \leq X$ e $\lim_n EX_n \leq \lim_m E V_m \leq \lim_m E X_m$ ou seja $V_m \uparrow X$ e $\lim_m E V_m = \lim_n E X_n$. Mas as V_m são variáveis aleatórias simples, portanto $EX = \lim_n E V_m = \lim_n E X_n$ o que prova a proposição. (O exercício 2 decorre fàcilmente desta proposição).

Na Definição 6.3.2 na realidade a hipótese $EX^2 < \infty$ não é usada. Só é necessário que $E|X| < \infty$. Claro que então var(X) pode ser $+\infty$. A raiz quadrada de σ_X^2 e denotada por σ_X ; e chamada desvio-padrão.

Uma aplicação muito importante da desigualdade de Chebycher está contida na seguinte proposição.

Lei fraca dos grandes números

Seja $\{X_n\}_{n=1,2,\ldots}$ uma sucessão de variáveis aleatórias independentes identicamente distribuidas tais que $E(X_1^2) < \infty$. Sejam $\mu = EX_1$, $\sigma^2 = \mathrm{var}(X_1)$ e $S_n = \sum_{i=1}^{n} X_i$. Temos o seguinte resultado

PROPOSIÇÃO (Lei fraca dos grandes números):

$$\frac{S_n}{n} \xrightarrow{P} \mu .$$

Demonstração: Seja $\epsilon > 0$, $P[|\frac{S_n}{n} - \mu| \geq \epsilon] =$

$$= P[|\sum_{i=1}^{n}(X_i - \mu)| \geq n\epsilon] \overset{\text{(Chebychev)}}{\leq}$$

$$\leq \frac{\operatorname{var}(\sum_{i=1}^{n}(X_i-\mu))}{n^2\epsilon^2} = \frac{n\sigma^2}{n^2\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \xrightarrow[n\to\infty]{} 0 .$$

EXERCÍCIOS

9. Provar que $E|X| < \infty$ se e sòmente se $\sum_{n=1}^{\infty} n\, P[|X| \geq n] < \infty$.

10. Sejam X e Y variáveis aleatórias simples com valores $x_1, x_2, \ldots, x_N$ e $y_1, y_2, \ldots, y_M$ respectivamente. Se $E\, X^n\, Y^m = E\, X^n\, E\, Y^m$ para todo $n = 1,2,\ldots.,N-1$; $m = 1,2,\ldots.,M-1$, então X e Y são independentes.

11. (Produtos de variáveis aleatórias independentes).

Sejam $X_1, X_2, \ldots, X_n$ n variáveis aleatórias independentes idênticamente distribuidas tais que $P[X_1 > 0] = 1$ e $E\,(\log X_1)^2 < \infty$. Então $\forall\, \epsilon > 0$

$$P[e^{n[E(\log X_1)-\epsilon]} < \prod_{i=1}^{n} X_i < e^{n[E(\log X_1)+\epsilon]}] \geq$$

$$1 - \frac{\operatorname{var}(\log X_1)}{n\epsilon^2} .$$

(<u>Sugestão</u>: tomar logarítmos e usar a desigualdade de Chebychev). (Êste resultado indica que o produto cresce como $[e^{E(\log X_1)}]^n$ e não como $(E\, X_1)^n$).

12. Generalizar a Lei fraca dos grandes números ao caso no qual as variáveis aleatórias são independentes mas não idênticamente distribuidas supondo que $\forall\, n$ $\operatorname{var}(X_n) \leq$ $\leq M$ (variâncias uniformemente limitadas).

13. Consideremos uma variável aleatória X com a seguinte distribuição:

$$P[X = n] = \frac{1}{e^{\lambda}-1}\,\frac{\lambda^{n}}{n!}\ , \quad \lambda > 0\ ,$$

$n=1,2,\ldots$ (Ver exercício 13 do Capítulo 3). Calcular EX e $\mathrm{var}(X)$.

CAPÍTULO 7

As hipóteses feitas ao início da seção 7.5 ($\mathrm{var}(X) > 0$, $\mathrm{var}(Y) > 0$) só começam a ser utilizadas na Definição 7.5.3.

A distribuição da soma de variáveis aleatórias independentes não foi especialmente tratada em nenhuma seção dêste livro. Mas note-se que alguns casos particulares foram estudados. No Exemplo 7.4.1 provou-se que a soma de duas variáveis aleatórias independentes com distribuições de Poisson com parâmetros λ e μ respectivamente e novamente Poisson $(\lambda+\mu)$. No Exercício 10 do Capítulo 7 um resultado sôbre a distribuição da soma de duas variáveis aleatórias independentes com distribuições binomiais ((n,p) e (m,p) respectivamente) é necessário. (A distribuição da soma e binomial $(n+m,p)$).

CAPÍTULO 8

Talvez seja conveniente dar ênfase à propriedade crucial contida no Teorema Central do Limite, que o fato que a distribuição limite (normal) e a mesma qualquer que seja a distribuição comum às X_i .

EXERCÍCIOS

10. Jogue uma moeda equilibrada 36 vêzes. Calcule aproximadamente a probabilidade de que o número de caras seja ≥ 15 e ≤ 20.

11. Se $F_n \to F$ e F é contínua, então F_n converge uniformemente a F.

12. Se $\{X_n\}_{n=1,2,\ldots}$ é uma sucessão de variáveis aleatórias tais que

função de distribuição de $X_n = F_n \to F$

onde $F = I_{[a,+\infty)}$ então $X_n \xrightarrow{P} a$.

CAPÍTULO 9

Os subíndices nos espaços (Ω_n, P_n) poderiam ser eliminados se um construisse um espaço de probabilidades e uma sucessão infinita de variáveis aleatórias independentes $\{X_n\}_{n=1,2,\ldots}$ tais que $X_n = \begin{cases} 1 & p \\ -1 & q \end{cases}$. Tal construção é possível mas as técnicas usadas nela estão por cima do nível desta monografia. Tal espaço "grande" conteria os (Ω_n, P_n) como "subespaços" e as P_n seriam restrições de uma probabilidade P definida sôbre uma família de subconjuntos dêsse espaço.

SOLUÇÕES E RESPOSTAS PARA ALGUNS EXERCÍCIOS

CAPÍTULO 1

3. $A \cup B = (A \Delta B) \Delta (A \cap B)$

   $A - B = A \Delta (A \cap B)$

4. $\lim_n \sup A_n = \mathbb{R}$ $\qquad$ $\lim_n \inf A_n = \emptyset$

CAPÍTULO 2

4. Consideremos a identidade

$$(1+t)^n = \sum_{k=0}^{n} \binom{n}{k} t^k \tag{I}$$

   Derivando sucessivamente temos

$$n(1+t)^{n-1} = \sum_{k=1}^{n} \binom{n}{k} k\, t^{k-1} \tag{II}$$

$$n(n-1)(1+t)^{n-2} = \sum_{k=2}^{n} \binom{n}{k} k(k-1)\, t^{k-2} \tag{III}$$

Substituindo t por -1 em (I) obtemos a)

Substituindo t por 1 em (II) obtemos b)

Substituindo t por -1 em (II) obtemos c)

Substituindo t por 1 em (III) obtemos d).

5. a) Para $1 \leq j \leq k \leq i$ temos $\binom{i}{k}\binom{k}{j} = \binom{i}{j}\binom{i-j}{k-j}$.

Isto pode ser provado idretamente ou senão observando que ambos os membros correspondem a duas formas diferentes de calcular o número de pares de subconjuntos $A \subseteq B$ de um conjunto com i elementos tais que $\#(A) = j$ e $\#(B) = k$. No membro esquerdo escolhe-se primeiro um subconjunto de k elementos e depois um subconjunto dêste com j elementos. No membro direito escolhe-se primeiro um subconjunto com j elementos e depois $k-j$ dos restantes $i-j$ pontos para formar o subconjunto B de cardinal k. Temos que

$$\binom{n}{k}(1+t)^k = \binom{n}{k}\sum_{v=o}^{k}\binom{k}{v}t^v = \sum_{v=o}^{k}\binom{n}{k}\binom{k}{v}\,t^v$$

agora a) decorre da fórmula

$$\binom{n}{k}\binom{k}{v} = \binom{n}{v}\binom{n-v}{k-v}$$

b) Substituir t por -1.

6. $$\sum_{v=0}^{n}(-1)^v\binom{a}{v} = \binom{a}{0} + \sum_{v=1}^{n}(-1)^v\left[\binom{a-1}{v} + \binom{a-1}{v-1}\right]$$

$$= \binom{a}{0} + \sum_{v=1}^{n}(-1)^v\binom{a-1}{v} + \sum_{v=1}^{n}(-1)^v\binom{a-1}{v-1} =$$

$$= \binom{a}{0} + \sum_{v=1}^{n}(-1)^v\binom{a-1}{v} + \sum_{v=0}^{n-1}(-1)^{1+v}\binom{a-1}{v} =$$

$$= \binom{a}{0} + (-1)^n\binom{a-1}{n} + (-1)^1\binom{a-1}{0} = (-1)^n\binom{a-1}{n}\ .$$

7. $\binom{n+1}{m+1} = \binom{n}{m+1} + \binom{n}{m} = \binom{n-1}{m+1} + \binom{n-1}{m} + \binom{n}{m} =$

$= \binom{m+1}{m+1} + \binom{m+1}{m} + \binom{m+2}{m} + \dots + \binom{n-1}{m} + \binom{n}{m} =$

$= \binom{m}{m} + \binom{m+1}{m} + \binom{m+2}{m} + \dots + \binom{n-1}{m} + \binom{n}{m}$.

9. Substituir em 8 a, b e ℓ por n.

10. $\binom{n+m}{n}$ $\binom{n+m-1}{m-1}$

13. $\frac{1}{n}$

CAPÍTULO 3

1. a) $\bigcup_{k=n}^{\infty} A_k \downarrow \lim_n \sup A_n$

Portanto $P(\lim_n \sup A_n) = \lim_n P(\bigcup_{k=n}^{\infty} A_k) \geq \lim_n \sup P(A_n)$

Em forma semelhante prova-se a outra desigualdade.

b) Decorre imediatamente de a).

2. Seja $a \in \Omega$ então $a = \bigcap_{n=1}^{\infty} [a, a+\frac{1}{n})$.

Portanto se P fôsse uma probabilidade $P(\{a\}) =$

$= \lim_n P([a, a+\frac{1}{n})) = \lim_n \frac{1}{n} = 0$. Como Ω é enumerável isto implicaria $P(\Omega) = 0$ o que é impossível.

3. $\bigcup_{i=1}^{\infty} A_i = \sum_{i=1}^{\infty} B_i$ onde $B_1 = A_1$ e $B_i = A_i - \bigcup_{j=1}^{i-1} A_j$

$i \geq 2$. Como $B_i \subseteq A_i$ $\forall i$ temos que

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

4. $$1 - \left[1 - \sum_{k=0}^{n} \frac{(-1)^k}{k!}\right] = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \xrightarrow[n\to\infty]{} e^{-1} .$$

5. $\frac{1}{5}$

7. a) $\Omega = \{f;\ f: I_3 \to I_q + \{0\}\}$

$$A = \{f:\ f \in \Omega, \sum_{i=1}^{3} f(i) > 9\}$$

$$P(A) = 1 - P\{f:\ f \in \Omega, \sum_{i=1}^{3} f(i) \leq 9\}$$

Pela Proposição 2.9.4 temos

$$P(A) = 1 - \frac{\binom{3+9}{3}}{1000} = 1 - \frac{22}{100} = 0{,}78 .$$

b) 0,384.

8. $$1 - \frac{\binom{20}{10}}{\binom{35}{10}}$$

10. Se $m \geq 1$ temos

$$\left(1 - \frac{1}{N}\right)^m = \left(1 - \frac{1}{N}\right)^{m-1} \left(1 - \frac{1}{N}\right) = \left(1 - \frac{1}{N}\right)^{m-1} -$$

$$- \frac{1}{N}\left(1 - \frac{1}{N}\right)^{m-1} \geq \left(1 - \frac{1}{N}\right)^{m-1} - \frac{1}{N} .$$

Reiterando a desigualdade temos que

$$\left(1 - \frac{1}{N}\right)^n \geq 1 - \frac{n}{N}$$

e portanto $1 - \left(1 - \frac{1}{N}\right)^n \leq 1 - \left(1 - \frac{n}{N}\right) = \frac{n}{N}$ ou seja que o jogo "frio" é melhor (apesar de que em geral parece provo car menos "satisfação" porque um joga menos tempo).

11. $P(\lim_n \sup A_n) = \lim_n P(\bigcup_{k=n}^{\infty} A_k) \leq \lim_n \sup(\sum_{k=n}^{\infty} P(A_k)) = 0.$

14. Seja $\varphi : I_n^{I_r} \to C^R_{r,n}$ a aplicação que transforma a função $f \in I_n^{I_r}$ na função de $C^R_{r,n}$ que se obtêm ordenando em forma crescente os valores da f. Isto é $f(I_r) = \{i_1, i_2, \ldots, i_p\}$ com $1 \leq i_1 < i_2 < \ldots < i_p \leq n$ e $\#(f^{-1}(\{i_j\})) = k_j$ $\varphi(f) = \sum_{j=1}^{p} i_j I_{[\sum_{i=1}^{j-1} k_i+1, \sum_{i=1}^{j} k_i]}$ $(\sum_{i=1}^{0} k_i = 0)$. Agora, $\#(\varphi^{-1}(f)) = \frac{r!}{k_1! k_2! \ldots k_p!}$ que claramente depende da f. Ou seja, se P é a probabilidade uniforme sôbre $I_n^{I_r}$, $P(\varphi^{-1}(\{f\})) \neq \frac{1}{\#(C^R_{r,n})}$ e portanto se um usa como espaço amostral $C^R_{r,n}$ os resultados não vão ser os mesmos. A escolha de $C^R_{r,n}$ como espaço amostral não é a adequada.

CAPÍTULO 4

3. $\frac{1}{6}$

5. a) $\frac{10}{12}$ b) não.

6. $\frac{.3.4}{.3.4+.7.6} = \frac{12}{54}$.

7. $\frac{9}{16}$ $\frac{285}{985}$

11. Seja $A = \{x: x = js, j=1,2,...\}$

$$P(A) = \sum_{j=1}^{\infty} \frac{1}{c_\gamma} \frac{1}{(js)^{1+\gamma}} = \frac{1}{s^{1+\gamma}} \left(\sum_{j=1}^{\infty} \frac{1}{c_\gamma} \frac{1}{j^{1+\gamma}} \right) = \frac{1}{s^{1+\gamma}} .$$

Seja agora $1 \leq j_1 < j_2 < \ldots < j_n$. Como os p_j são primos temos $\bigcap_{i=1}^{n} A_{j_i} = \{x: x$ é divisível por $\prod_{i=1}^{n} p_{j_i}\}$ ..

Portanto

$$P\left(\bigcap_{i=1}^{n} A_{j_i}\right) = \frac{1}{\left(\prod_{i=1}^{n} p_{j_i}\right)^{1+\gamma}} = \prod_{i=1}^{n} \frac{1}{(p_{j_i})^{1+\gamma}} = \prod_{i=1}^{n} P(A_{j_i})$$

o que prova a independência dos A_j $j=1,2,...$

$$\frac{1}{c_\gamma} = P(\{1\}) = \lim_n P\left(\bigcap_{j=1}^{n} A_j^c\right) = \lim_n \prod_{i=1}^{n} P(A_j^c) =$$

$$= \lim_n \prod_{i=1}^{n} [1 - P(A_j)] = \lim_n \prod_{i=1}^{n} \left(1 - \frac{1}{p_j^{1+\gamma}}\right) .$$

Esta identidade claramente implica o resultado.

12. a) $\frac{7}{12}$. b) $\frac{2}{5}$.

13. Seja $\omega \in \Omega$; $\forall n$ $\omega \in \bigcap_{i=1}^{n} B_i$ onde B_i é igual a A_i ou A_i^c. $P(B_i) \leq p_i \vee (1-p_i) = 1 - \varphi(p_i)$ onde $p_i = P(A_i)$. Portanto $P(\{\omega\}) \leq \prod_{i=1}^{n} [1 - \varphi(p_i)] \leq e^{-\sum_{i=1}^{n} \varphi(p_i)}$.

Se $\sum_{n=1}^{\infty} \varphi(p_n) = \infty$ isto implicaria que para todo ω, $P(\{\omega\}) = 0$ o qual é absurdo porque Ω é enumerável.

CAPÍTULO 5

4. $$P(B_{r+s}|B_r) = \frac{P(B_{r+s})}{P(B_r)} = \frac{\sum_{k=r+s+1}^{\infty} p\,q^{k-1}}{\sum_{k=r+1}^{\infty} p\,q^{k-1}} = \frac{p\,q^{r+s}}{p\,q^{r}} =$$

$$= q^s = P(B_s).$$

A equação $Q(B_{r+1}) = Q(B_r)\,Q(\{1\})$ implica ràpidamente que

$$Q(B_k) = Q^k(B_1) = [1 - Q(\{1\})]^k \quad k=1,2,\ldots.$$

para $k \geq 2$ temos

$$Q(\{k\}) = Q(B_{k-1} - B_k) = Q(B_{k-1}) - Q(B_k) =$$

$$= [Q(B_1)]^{k-1} - [Q(B_1)]^k = [Q(B_1)]^{k-1}\,[1-Q(B_1)] =$$

$$= Q(\{1\})\,[1 - Q(\{1\})]^{k-1}.$$

5. Temos $\frac{P(\{n+1\})}{P(\{n\})} = \frac{\lambda}{n+1}$. Tomemos $k \geq \lambda$. Seja

$$P(\{k+j\}) = P(\{k\})\,\frac{P(\{k+1\})}{P(\{k\})}\cdots\frac{P(\{k+j\})}{P(\{k+j-1\})} =$$

$$= P(\{k\})\left(\frac{\lambda}{k+1}\right)\left(\frac{\lambda}{k+2}\right)\cdots\left(\frac{\lambda}{k+j}\right) \leq$$

$$\leq P(\{k\})\left(\frac{\lambda}{k+1}\right)^j.$$

Para $j \geq 0$ temos em geral

$$P(\{k+j\}) \leq P(\{k\})\left(\frac{\lambda}{k+1}\right)^j.$$

Agora

$$P(\{k\} \mid \{x:\ x\geq k\}) \geq \frac{P(\{k\})}{P(\{k\})\sum_{j=0}^{\infty}\left(\frac{\lambda}{k+1}\right)^j} =$$

$$= \frac{1}{\dfrac{1}{1 - \dfrac{\lambda}{k+1}}} \xrightarrow[k\to\infty]{} 1 .$$

8. $\Omega = \{x_1, x_2, \ldots\}$. Se $A \subseteq \Omega$ $P(A) = \sum_{i=1}^{\infty} I_A(x_i) p_i$ e

$X: \Omega \to \mathbb{R}$ definida por $X(x_i) = x_i$.

9. Dado $\epsilon > 0$ seja n tal que $\sum_{i=n+1}^{\infty} P[X=x_i] < \epsilon$.

Seja $\delta > 0$ tal que $x_i \notin (x-\delta, x+\delta)$ para $i=1,2,\ldots,n$

Então se y pertence a êsse intervalo $|F(x)-F(y)| < \epsilon$.

CAPÍTULO 6

2. Vamos provar que se $A_n \uparrow \Omega$ então $E\, I_{A_m} |Y| \uparrow E|y|$.

(Isto implica imediatamente o resultado).

Sejam $\{y_1, y_2, \ldots.\}$ os valores de Y e $B_m = \sum_{i=1}^{m} [Y=y_i]$. Como $I_{B_m}|Y|$ é uma variável aleatória simples temos que $E|Y| = \lim_m E\, I_{B_m}|Y|$. Agora

$$\lim_n E(I_{A_n}|Y|) = \lim_n \lim_m E(I_{A_n} I_{B_m}|Y|) =$$

$$= \lim_m \lim_n E(I_{A_n} I_{B_m}|Y|) = \lim_m E\, I_{B_m}|Y| =$$

$$= E|Y| .$$

A primeira igualdade decorre do fato que $I_{A_m} I_{B_m}|Y|$ é uma r.a. simples, a segunda porque $E(I_{A_n} I_{B_m}|Y|)$ é crescente em n e m.

3. $\frac{n}{2}$ $\quad \frac{n^2}{12} + \frac{n}{16}$

5. Podemos supor $\epsilon \leq 1$. Seja $h(x) = x^2 - \epsilon^2$. Temos $I_{\{x:|x|\geq\epsilon\}} \geq h$ sôbre $[-1,1]$. Agora

$$P[|x|\geq\epsilon] = E(I_{\{x:|x|\geq\epsilon\}} \circ X) \geq E(h \circ X) = EX^2 - \epsilon^2 \ .$$

CAPÍTULO 7

1. $E(X-a)^2 = E(X-EX)^2 + [E(X)-a]^2 + 2\,E(X-EX)(EX-a) =$

$$= E(X-EX)^2 + (EX-a)^2 \geq E(X-EX)^2 \ .$$

3. $P[\tau = j \mid X=i] = \dfrac{P[\tau = j, X=i]}{P[X=i]}$ $\quad i=2,3,\ldots$ $\quad j=1,2,\ldots,i-1$

$$= \frac{p^2\, q^{i-2}}{(i-1)\, p^2\, q^{i-2}} = \frac{1}{i-1}$$

ou seja a distribuição condicional e uniforme sôbre $\{1,2,\ldots,i-1\}$.

5. Temos que provar que $\operatorname{cov}(Y,X) = \operatorname{cov}(\hat{Y},X)$. Pelas propriedades da covariância, $\operatorname{cov}(\hat{Y},X) = \hat{a}\,\sigma_X^2$. A identidade a verificar e portanto equivalente a $\operatorname{cov}(Y,X) = \hat{a}\,\sigma_X^2$, a qual é imediata da fórmula $\hat{a} = \dfrac{\sigma_Y \rho}{\sigma_X}$.

8. $\sigma_{X+Y}^2 = \operatorname{var}(X+Y) = \sigma_X^2 + \sigma_Y^2 + 2\operatorname{cov}(X,Y)$ (Proposição 7.5.2).

A desigualdade de Cauchy-Schwarz (Anexo 2) implica

$$|\operatorname{cov}(X,Y)| \leq \sigma_X\,\sigma_Y$$

ou seja $-\sigma_X\sigma_Y \le \mathrm{cov}(X,Y) \le \sigma_X\sigma_Y$ e portanto

$$\sigma_X^2 + \sigma_Y^2 - 2\sigma_X\,\sigma_Y \le \sigma_{X+Y}^2 \le \sigma_X^2 + \sigma_Y^2 + 2\,\sigma_X\,\sigma_Y \quad \text{ou}$$

$$(\sigma_X - \sigma_Y)^2 \le \sigma_{X+Y}^2 \le (\sigma_X + \sigma_Y)^2$$

o que implica $|\sigma_X - \sigma_Y| \le \sigma_{X+Y} \le \sigma_X + \sigma_Y$.

CAPÍTULO 8

1. $P[X_n \le x] = P[X_n \le x,\ X \le x+\epsilon] + P[X_n \le x,\ X > x+\epsilon] \le$
$$\le P[X \le x+\epsilon] + P[|X_n - X| \ge \epsilon]$$

isto é

$$F_n(x) - P[|X_n - X| \ge \epsilon] \le F(x+\epsilon)\ .$$

Tomando limites um prova que

$$\forall\, \epsilon > 0 \quad \lim_n \sup F_n(x) \le F(x+\epsilon)\ .$$

Em forma semelhante é fácil provar que

$$\forall\, \epsilon > 0, \quad F(x-\epsilon) \le \lim_n \inf F_n(x).$$

O resultado decorre do fato que F é contínua em x.

2. Seja x um ponto de continuidade da F. Escolhendo $a = x$ e $b > a$ temos

$$\lim_n \inf\ (1 - F_n(x)) \ge \lim_n \inf\ [F_n(b) - F_n(x)] = F(b) - F(x)$$

ou seja $1 - \lim_n \sup F_n(x) \ge F(b) - F(x),\ \forall\, b$.

Portanto $1 - \lim_n \sup F_n(x) \ge 1 - F(x)$ o que implica que $F(x) \ge \lim_n \sup F_n(x)$. Tomando $x = \smile$ e $a < x$ em forma semelhante obtemos

$$\lim_n \inf F_n(x) \geq F(x)$$

o que prova o resultado.

7. Vamos provar primeiro que se X e Y são duas variáveis independentes tais que $\mathcal{L}(X) = \text{Poisson}(\lambda)$ e $\mathcal{L}(Y) = \text{Poisson}(\mu)$, então $\mathcal{L}(X+Y) = \text{Poisson}(\lambda+\mu)$

$$P[X+Y = k] = \sum_{i=0}^{k} P[X=i,\ Y=k-i] = \sum_{i=0}^{k} P[X=i]\ P[Y=k-i] =$$

$$= \sum_{i=0}^{k} \frac{e^{-\lambda}\lambda^{i}\, e^{-\mu}\mu^{k-i}}{i!\,(k-i)!} = \frac{e^{-(\lambda+\mu)}}{k!} \sum_{i=0}^{k} \binom{k}{i} \lambda^{i} \mu^{k-i} =$$

$$= \frac{e^{-(\lambda+\mu)}}{k!} (\lambda+\mu)^{k} .$$

Seja agora $\{X_n\}_{n=1,2,\ldots}$ uma sucessão de variáveis aleatórias independentes idênticamente distribuidas Poisson (1). $E\,X_n = 1$, $\text{var}(X_n) = 1$. Pelo Teorema Central do Limite

$$P\left[\frac{\sum_{i=1}^{n} X_i - n}{\sqrt{n}}\right] \leq 0 \longrightarrow \int_{-\infty}^{0} f(x)dx = \frac{1}{2}$$

ou seja

$$P[\sum_{i=1}^{n} X_i \leq n] \to \frac{1}{2} .$$

Como $\mathcal{L}(\sum_{i=1}^{n} X_i) = \text{Poisson}(n)$ temos que

$$P[\sum_{i=1}^{n} X_i \leq n] = \sum_{k=0}^{n} \frac{e^{-n}\, n^{k}}{k!}$$

o que prova o resultado.

9. Vamos provar uma proposição que é mais geral e que tem um interêsse independente. Seja Ω um conjunto enumerável e $\{P_n\}_{n=1,2,\ldots}$ uma sucessão de probabilidades determinadas por sucessões $\{p_{nk}\}_{\substack{n=1,2,\ldots\\ k=1,2,\ldots}}$ $p_{nk} \geq 0$, $\sum_{k=1}^{\infty} p_{nk} = 1$.

Seja P outra probabilidade sôbre Ω determinada por sucessão $\{p_k\}_{k=1,2,\ldots}$. Se $\forall k$, $\lim_{n\to\infty} p_{nk} = p_k$ então $\sum_{k=1}^{\infty} |p_{nk} - p_k| \to 0$.

Demonstração:

$$\sum_{k=1}^{\infty} |p_{nk}-p_k| \leq \sum_{k=1}^{N} |p_{nk}-p_k| + \sum_{k=N+1}^{\infty} p_{nk} + \sum_{k=N+1}^{\infty} p_k$$

agora

$$\sum_{k=N+1}^{\infty} p_{nk} \leq \sum_{k=N+1}^{\infty} p_k + \left|\sum_{k=N+1}^{\infty} p_{nk} - \sum_{k=N+1}^{\infty} p_k\right|$$

mas

$$\left|\sum_{k=N+1}^{\infty} p_{nk} - \sum_{k=N+1}^{\infty} p_k\right| = \left|\left(1 - \sum_{k=1}^{N} p_{nk}\right) - \left(1 - \sum_{k=1}^{N} p_k\right)\right|$$

$$= \left|\sum_{k=1}^{N} p_{nk} - \sum_{k=1}^{N} p_k\right| \leq \sum_{k=1}^{N} |p_{nk}-p_k| \;.$$

Portanto

$$\sum_{k=1}^{\infty}|p_{nk}-p_k| \leq 2\left[\sum_{k=1}^{N}|p_{nk}-p_k| + \sum_{k=N+1}^{\infty} p_k\right] .$$

Escolhendo primeiro N e depois n grande é fácil ver que o membro esquerdo é arbitràriamente pequeno quando n é grande o que prova o resultado.

Uma conseqüência desta proposição é a seguinte:

Se $A \subseteq \Omega$ temos que

$$|P_n(A) - P(A)| \le \sum_{k=1}^{\infty} |p_{nk} - p_k|$$

e portanto

$$\lim_{n\to\infty} \sup_{A\subseteq\Omega} |P_n(A) - P(A)| = 0 .$$

11. Como F é contínua dado $\epsilon > 0$, existem pontos $x_1 < x_2 < \ldots < x_k$ tal que $F(x_1) < \epsilon$, $1 - F(x_k) < \epsilon$, $F(x_{i+1}) - F(x_i) < \epsilon$ $i=1,2,\ldots,k-1$.

Seja n_o tal que $\forall\, n \ge n_o$, $\forall i$, $1 \le i \le k$

$$|F(x_i) - F_n(x_i)| < \epsilon .$$

Agora é fácil ver que para todo x,

$$|F_n(x) - F(x)| < 2\epsilon$$

analisando o que acontece em cada um dos intervalos.

CAPÍTULO 9

1. [número de trajetórias de $(0,0)$ até $(2n,0)$ que são ≥ 0] $p^n q^n$ = [número de trajetórias de $(0,0)$ até $(2n+1,1)$ que são > 0] $p^n q^n = \binom{2n+1}{n+1} \frac{1}{(2n+1)} p^n q^n = \binom{2n}{n} p^n q^n \frac{1}{n+1} =$

$$= \frac{P_{2n}[S_{2n} = 0]}{n+1} .$$

4. $\sum_{k=0}^{n} \binom{n}{k} \frac{1}{2^n} \binom{n}{k} \frac{1}{2^n} = \sum_{k=0}^{n} \binom{n}{k}^2 \frac{1}{2^{2n}}$.

5. A probabilidade de obter k caras é igual a de obter $(n-k)$ caras. Se indicamos com 1 e -1 cara e coroa e pensamos que os jogadores realizam seus jogos consecutivamente, o experimento pode ser representado por uma sucessão de $2n$ variáveis aleatórias independentes identicamente distribuidas $X_i = \begin{cases} 1 & 1/2 \\ -1 & 1/2 \end{cases}$, $i=1,2,\ldots,2n$. Se S_i, $i=1,2,\ldots,2n$ indica as somas parciais a condição que o primeiro jogador obtenha k caras e o segundo $(n-k)$ caras corresponde ao evento $S_{2n} = 0$. Portanto a probabilidade buscada no exercício 4 e igual a $\binom{2n}{n}\frac{1}{2^{2n}}$.

Do exercício 4 temos então

$$\frac{1}{2^{2n}} \sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}\frac{1}{2^{2n}} .$$

Portanto

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

ANEXO 1

1. Seja $a_{nm} = \begin{cases} \rho^n & 1 \le m \le n \\ 0 & m > n \end{cases}$

$$\sum_{n=1}^{\infty} n\rho^n = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} a_{nm} = \sum_{m=1}^{\infty}\Big[\sum_{n=1}^{\infty} a_{n,m}\Big] = \sum_{m=1}^{\infty}\Big[\sum_{n=m}^{\infty} \rho^n\Big] =$$

$$= \sum_{m=1}^{\infty} \frac{\rho^m}{1-\rho} = \frac{1}{(1-\rho)} \sum_{m=1}^{\infty} \rho^m = \frac{\rho}{(1-\rho)^2} .$$

ERRATA

Para identificar as correções usaremos o seguinte método: nº da página, nº da linha, de cima para baixo e de baixo para cima. Exemplo: pag.20,+5 ou pag.35,-7.

pag. 6,-3:

$$= \sum_{k=1}^{n} (-1)^{k-1} \sum_{1\le i_1<i_2<\ldots<i_k\le n} \prod_{j=1}^{k} I_{A_{i_j}}$$

pag. 7,+1

1.3.7. $I_{\lim_n \sup A_n} = \lim_n \sup I_{A_n}$;

$I_{\lim_n \inf A_n} = \lim_n \inf I_{A_n}$.

pag. 7,-4

... a notação $f|_A$ é a função...

pag. 7,-2

$$f|_A(x) = f(x)$$

pag. 7,-1

$$f|_A$$

pág.11,-7

$$\#(A_{m,n}) = n(n-1)\ldots(n-m+1)$$

pag.13,-2

$$\ldots = \binom{n}{m}.$$

pag.14,+8

$$\phi(f) = (f|_{A_1},\ldots,f|_{A_p}) \ \ldots$$

pag.16,-3

$$\frac{n(n-1)}{2}$$

pag.16,-1

$$\frac{n(n-1)}{2}$$

pag.17,+8

$$A = \{(1,j): 1 \leq j \leq i \leq n\}$$

pag.18,+7

$$f(i) = \begin{cases} g(i)-1, & i = 1 \\ g(i)-g(i-1), & 1 < i \leq m \end{cases}$$

pag.20,+4

$$n! \cong \sqrt{2\pi}\ n^{n+\frac{1}{2}}\ e^{-n}$$

pag.20,-4

$$\log(n!) = \int_{1/2}^{n+1/2}$$

pag.21,+4

$$I(x) = \int_0^x \log y \, dy$$

pag.21,-5

$$a_k = \int_0^{1/2} \log \frac{1}{1-(\frac{y}{k})}\, dy \ ,$$

pag.22,+7

... dos n primeiros...

pag.22,-7

$$g(n) = \sum_{k=0}^{n-1} \binom{n-1}{k} g(k)$$

pag.23,+6

$$\sum_{v=0}^{n} (-1)^v \binom{a}{v} = (-1)^n \binom{a-1}{n} \quad \text{(usar Proposição 2.8.1)}$$

pag.23,+7

(usar Proposição 2.8.1)

pag.23,-2

(Sugestão: $\binom{i}{k}\binom{k}{j} = \binom{i}{j}\binom{i-j}{k-j}$)

pag.35,-12

como $\sum_{j=1}^{\infty} p_{i_j}$ converge

pag.35,-11

$$\sum_{j=k}^{\infty} p_{i_j} < \epsilon$$

pag.35,-8

$$P(A_i) \leq \sum_{j=k}^{\infty} p_{i_j} < \epsilon$$

pag.36,-5

da seção 2.

pag.37,+9

$$p_1 = p_2 = \ldots = p_6 = \frac{1}{6}$$

pag.40,-4

$$= 1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1 - \frac{(r-1)}{n}\right).$$

pag.41,+9

menor

pag.44,+10

$$C^R_{3,52}$$

pag.46,-7

N

pag.53,-6 e -5

Os últimos têrmos são iguais, por indução,a

$$\prod_{i=1}^{n} P(A_i^c) \prod_{i=n+1}^{r} P(A_i) \text{ é } \prod_{i=1}^{n} P(A_i^c) \prod_{i=n+2}^{r} P(A_i)$$

pag.59,+8

A

pag.63,+4

$$\sum_{\omega\in\Omega} p(\omega) = 1$$

pag.67,-10

fechado

pag.67,-8

$$P(B_1) = \frac{1}{2}$$

pag.70,-7 e -6

(Sugestão: Usar a seguinte ...

pag.71,+2

pretas. .... urna

pag.71,+3

urna

pag.86,-1

$$\ldots < p_\lambda(k)\, e^{\frac{k\lambda}{n}} \leq \ldots$$

pag.87,+8

$$E(X) = \int X(\omega)P(d\omega) =$$

pag.88,-3

$$\omega \in \Omega$$

pag.89,+2

menor

pag.89,-5

Proposição 6.1.1

pag.89,-3

$$EX_n = \lim_{m\to\infty} E(X_n \wedge Y_m) \leq$$

pag.91,-2

esperanças

pag.91,-1

esperança

pag.92,+1

... serão usadas como sinônimos.

pag.93,-1

ésperanças

pag.94,+9

esperança

pag.94,-6

$$\{x_1, x_2, \dots\}$$

pag.95,-6

esperanças

pag.97,-4

pares de variáveis independentes

pag.97,-3

X^+Y^+ , X^-Y^- , X^+Y^- e X^-Y^+

pag.100,-2

... distribuidas com lei ...

pag.102,+1

$g(x) = g(|x|) \geq g(\varepsilon)$)

pag.102,+5

$g(x) = |x|^r$

pag.102,-2

$$I_{\{x:x\geq\mu+\varepsilon\}} \leq \frac{(x-b)^2}{(\mu+\varepsilon-b)^2}$$

pag.104,+7

em $\Omega = \{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\}$

pag.104,+8

$f:[0,1] \to \mathbb{R}$ e $A \subseteq [0,1]$

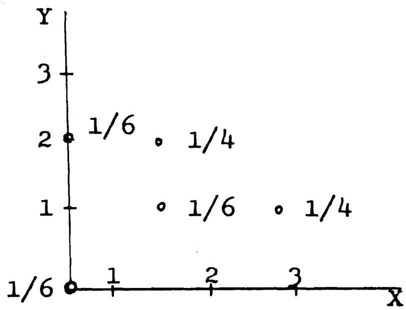
pag.109,+6

... $P[X=(x_{1j_1},\ldots,x_{i-1,j_{i-1}},x_{ij},x_{i+1,j_{i+1}}$ ...

pag.109,+7

... $p_X(x_{1j_1},\ldots,x_{i-1,j_{i-1}},x_{ij},x_{i+1,j_{i+1}}$ ...

pag.109, figura 7.1.1



pag.111,+4

$$\{x_{i1}, x_{i2}, \ldots\}.$$

pag.112,-9

$$Q(\{y_j\}) \text{ se } P[X=x_i] = 0$$

pag.112,-8

$$\bar{\mu}(x_i;A) = \sum_{\{j:y_j \in A\}} \bar{\mu}(x_i;\{y_j\})$$

pag.112,-6

Dada a distribuição condicional de Y dado

pag.113,+5

7.4 - Esperança condicional

pag.113,-9

.... esperança condicional de Y ...

pag.113,-1

é chamada esperança condicional ...

pag.114,+4

DEFINIÇÃO 7.4.3 - A função ...

pag.114,-8 e -7

b) $E(1|X) = 1$

c) $E(cY|X) = cE(Y|X)$

pag.115,+1

... esperanças.

pag.115,+2

a) ... $\frac{1}{P[X=x_i]}$ ...

pag.115,+4

... $(\sum_j y_j$ ...

pag.115,-6

$$\ldots = \sum_{i=1}^{\infty} g(x_i)\, I_{[X=x_i]} = g \circ X \,.$$

pag.115,-5

... $(\int_{[X=x_i]} YdP)$ ...

pag.115,-4

$$\ldots \int_{[X=x_i]} YdP = \int_{\underset{\{i:p_X(x_i)>0\}}{\cup}[X=x_i]} YdP =$$

pag.116,+6

... esperança ...

pag.116,+8

... determinar ...

pag.116,-10

... de mínima distância ...

pag.117,+2

$$+ 2\ E\{(Y - E(Y|X))\ (E(Y|X) - h(X))\}.$$

pag.117,+5

7.4.1, e) e f))

pag.117,+6

$$2E\{[E(Y|X) - h(X)]E[(Y-E(Y|X))|X]\}$$

pag.117,+7

Mais $E[(Y-E(Y|X))|X] = E(Y|X) - E(Y|X) = 0$

pag.117,+8

... Proposição 7.4.1 d) e f))

pag.117,+9

frase nula

pag.118,+2

a) $\operatorname{var}(Y|X) = \sum_i [\sum_j y_j^2 \ldots -(\sum_j y_j\ \mu(x_i;\{y_j\}))^2] \ldots$

pag.121,-4

e $\operatorname{cov}(X+Z,Y) = \operatorname{cov}(X,Y) + \operatorname{cov}(Z,Y))$

pag.122,-3

... $\operatorname{var}(X_i) + 2 \sum_{i<j}$

pag.123,-8

$$\text{é} \quad 1 - \rho^2_{X,Y} \;.$$

pag.124,+4

$$\frac{a\sigma_X}{\sigma_Y} = \rho_{X',Y'} = \rho_{X,Y} \quad \text{(ver Proposição 7.5.3 a))}$$

pag.128,-3

parâmetros $r=2,\ 0 < p < 1$

pag.131,+2

INTRODUÇÃO AO TEOREMA CENTRAL DO LIMITE

pag.131,+5

Teorema Central do Limite.

pag.132,-7

$$\int_{-\infty}^{+\infty} f(x)dx = \ldots.$$

pag.132,-2

$$\left|\frac{\partial(x,y)}{\partial(r,\varphi)}\right| = r$$

pag.133,+8

$$\ldots. \quad \int_x^{+\infty} f(y)dy < \frac{f(x)}{x}$$

pag.135,-8

TEOREMA 8.2.1 (Teorema Central do Limite)

pag.138,+1

... Teorema Central do limite ...

pag.139,+2

$$= P[S_{2n}=n] \prod_{k=0}^{j-1} \frac{1}{(1+\frac{j}{n-k})} .$$

pag.143,-5

... Teorema Central do Limite ...

pag.143,-4

$$\lim_{n\to\infty} \; e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} = \frac{1}{2} .$$

pag.143,-1

... Teorema Central do Limite).

pag.144,+2

... Central do Limite.

pag.147,-8

$$\ldots N_{(n,x)} := \binom{n}{\frac{n+x}{2}}.$$

pag.148,-2

$$= p^a q^b \{ \binom{n-1}{\frac{n+x}{2} -1} - \ldots .$$

pag.149,+1

$$\ldots = \binom{n}{a} \frac{a-b}{n} = \binom{n}{a} \frac{x}{n} \quad (9.1.1)$$

pag.149,+7

$$\ldots = \frac{a-b}{n} = \frac{a-b}{a+b} = \frac{x}{n}$$

pag.150,+3

$$= 2\binom{2n-1}{n} \ldots .$$

pag.150,+8

$$\ldots\ f_{2n} = \frac{\cdots}{n2^{2n}\quad \cdots}$$

pag.151,-4

$$= \binom{\ell+2k}{k} \quad \ldots\ldots$$

pag.152,-7

$$= P_{2n}[S_{2n}=0] = u_{2n}$$

pag.153,+4

$$M_n = \max_{0\leq i\leq n} S_i \ .$$

pag.153,+7

$$\overline{M}_n = \max_{2\leq i\leq n+1} (S_i - S_1)$$

pag.154,+4

$$= \frac{1}{2^n} \binom{n}{[\frac{n}{2}]}$$

pag.154,-4

$$P_n[\frac{M_n}{\sqrt{n}} \leq x] \quad \ldots$$

pag.154,-2

$$P_n[\frac{M_n}{\sqrt{n}} \leq x] = \frac{1}{2^n} \sum_{k\leq x\sqrt{n}} \binom{n}{[\frac{n-k}{2}]} \ .$$

pag.156,+3

$$= \frac{\binom{2k}{k}\binom{2(n-k)}{n-k}}{2^{2n}} \ .$$

pag.157,+3

$$P_{2n}[a < \quad \ldots\ldots$$

pag.157,+5

$$\lim_{n\to\infty} P_{2n}[\ldots.$$

pag.157,+7

$$P_{2n}[\ldots.$$

pag.158,+3

$$P_{2n}[\ldots.] = \frac{P_{2n}[S_{2n}=0]}{n+1}$$

pag.158,+4

$$(P_{2n}[S_{2n}=0] = \ldots$$

pag.159,-9

$$\lim_{n\to\infty} \sum_{i=0}^{n} |a_i| < +\infty .$$

pag.159,-3

$$\sum_{n=0}^{\infty} a_{h(n)}$$

pag.160,-1

$$|q_{i+1}| \to |a_{i+1}|$$

pag.163,-8

$$\varphi(u) \xrightarrow[u\to+\infty]{} +\infty ,$$

pag.167,-3

[8] - PROHOROV, YU V.