

Uma Introdução à Probabilidade

Maria Eulália Vares
Centro Brasileiro de Pesquisas Físicas
Vladas Sidoravicius
Instituto de Matemática Pura e Aplicada

RESUMO

Neste curso, após introduzirmos algumas das noções e resultados básicos de probabilidade, vamos abordar a relação entre o modelo conhecido como gás de Lorentz (em sua versão discreta) e teoria de percolação. Mesmo se tratando de uma simplificação do modelo proposto há aproximadamente cem anos, os resultados rigorosos são extremamente escassos. A relação com percolação nos permite introduzir e discutir várias questões interessantes. Neste contexto, e utilizando fatos de percolação, vamos provar um teorema devido a Harris.

1 Introdução

A Teoria da Probabilidade constitui uma importante área da Matemática, dotada de sólida base axiomática e que experimentou um vertiginoso crescimento desde a segunda metade do século 20. Oferece uma modelagem para uma

classe de fenômenos sujeitos a incerteza, nos quais a idéia de regularidade estatística aparece como ponto importante. Para sermos mais corretos, deveríamos de fato falar em “teorias” de probabilidade, uma vez que ao longo de seu desenvolvimento várias formulações foram propostas, diferindo por vezes em aspectos-chave tais como a própria conceituação de “aleatório”, axiomatização, formulação de resultados, etc. Ficamos entretanto no contexto da axiomatização proposta por Kolmogorov, cujo sucesso manifesta-se no potencial desta área, bem como na enorme aplicabilidade a problemas oriundos das mais variadas ciências, tais como física estatística, biologia, economia, ciências sociais, a conexão com outras áreas da matemática como combinatória e teoria de números, além de fornecer um excelente arcabouço teórico para a estatística.

O conceito fundamental nesta construção matemática é o de espaço de probabilidade. Pressupõe três ingredientes na modelização de um “experimento” (ideal) sujeito a incerteza:

- Um conjunto Ω não-vazio, denominado espaço amostral, cujos elementos representam os possíveis resultados do experimento.
- Uma classe \mathcal{A} de subconjuntos do espaço amostral. Os conjuntos em \mathcal{A} , são chamados de eventos aleatórios, e a eles será atribuída uma probabilidade. Sobre a classe \mathcal{A} colocam-se as seguintes condições:

(A1) $\Omega \in \mathcal{A}$;

(A2) Se $A \in \mathcal{A}$ então seu complementar $A^c = \{\omega \in \Omega: \omega \notin A\}$ também pertence a \mathcal{A} ;

(A3) Se $A_1, A_2, \dots \in \mathcal{A}$, então $A_1 \cup A_2 \cup \dots \in \mathcal{A}$.

Uma classe satisfazendo (A1), (A2) e (A3) é chamada de σ -álgebra de subconjuntos de Ω .

Observação: Uma σ -álgebra é sempre fechada para aplicações das operações de conjuntos \cup, \cap, \setminus , etc... aplicadas uma quantidade enumerável de vezes.

Por trás desta definição fica o seguinte: um subconjunto A de Ω pode ser naturalmente identificado com a afirmação “o resultado do experimento corresponde a um elemento de A ”, a qual podemos dizer ser verdadeira (caso em que dizemos que “ A ocorre”) ou falsa (i.e. “ A não ocorre”) apenas após a realização do experimento. A isto damos o nome de “evento” e uma classe especial de eventos fica identificada em \mathcal{A} .

- Uma função $P: \mathcal{A} \rightarrow \mathbb{R}$ verificando as seguintes propriedades

(P1) $P(\Omega) = 1$

(P2) $P(A) \in [0, 1]$

(P3) Se A_1, A_2, \dots são elementos de \mathcal{A} dois a dois disjuntos, i.e., $A_i \cap A_j = \emptyset$ se $i \neq j$, então $P(A_1 \cup A_2 \cup \dots) = \sum_{i=1}^{\infty} P(A_i)$.

O estudante tem desde já toda razão para se questionar sobre a justificativa de tais axiomas. Uma primeira reflexão nos leva a verificar que um ponto delicado, que escapa a uma justificativa heurística reside na propriedade (P3). Baseados na idéia de frequência assintótica seria bem mais fácil justificar uma versão mais fraca: (P3)' Se $A, B \in \mathcal{A}$ e $A \cap B = \emptyset$ então $P(A \cup B) = P(A) + P(B)$.

A passagem de (P3)' (aditividade finita) para a condição mais forte (P3) (σ -aditividade) justifica-se puramente com base na conveniência matemática. Como veremos em aula, o que ela acrescenta é de fato uma certa propriedade de continuidade.

A partir de outras experiências, podemos já imaginar como isso é importante na análise matemática. Com relação às propriedades exigidas em \mathcal{A} precisamos que seja uma classe flexível, e com boas propriedades de “fechamento”. Se, por um lado a própria natureza do problema pode impor restrições naturais à classe de eventos nos quais *estamos interessados*, a condição (P3) nos levaria ademais a dificuldades de consistência se quiséssemos definir P para todos os subconjuntos de Ω , de modo geral. (Trata-se de questão delicada ligada aos fundamentos da matemática. Breve discussão em aula.)

Para evitar que nos ocupemos desde já dessas questões vamos começar explorando exemplos onde tais dificuldades não aparecem: se o espaço amostral Ω for finito ou infinito enumerável. Neste caso, se tivermos associado a cada $\omega \in \Omega$ um número $p(\omega) \in [0, 1]$ de modo que $\sum_{\omega \in \Omega} p(\omega) = 1$, podemos tomar \mathcal{A} como a classe de todos os subconjuntos de Ω e para cada $A \subseteq \Omega$ definimos $P(A)$ como a soma dos $p(\omega)$ com ω variando em A . Facilmente verificam-se as propriedades (P1), (P2) e (P3).

Exemplos.

1. *n lançamentos independentes de uma moeda honesta.* Por conforto notacional escrevemos 1 para a face *cara* e 0 para a face *coroa*. Estamos com um experimento idealizado, onde apenas as duas possibilidades *cara/coroa* são permitidas. Neste caso temos

$$\Omega_n = \{(\omega_1, \dots, \omega_n) : \omega_i \in \{0, 1\}, i = 1, \dots, n\} = \{0, 1\}^n, \quad (1)$$

ou seja um *resultado possível* é uma sequência ordenada, representando a face

superior nos sucessivos lançamentos. A idéia de lançamentos independentes com uma moeda honesta manifesta-se pela atribuição de exatamente a mesma probabilidade 2^{-n} para cada sequência. Neste caso temos então uma probabilidade $P^{(n)}$ tal que $P^{(n)}(A) = \#A/2^n$ para todo subconjunto A , onde $\#A$ é número de elementos no conjunto A . Vemos que neste caso o cálculo das probabilidades se reduz a uma questão de contagem. Por exemplo, a probabilidade de observarmos exatamente k caras é dada por $\binom{n}{k}2^{-n}$, para $k = 0, 1, \dots, n$.

2. *n extrações (com reposição) de urna com M bolas numeradas de 1 a M.*

Se estivermos interessados no resultado de forma ordenada, vemos que se trata de uma situação análoga à anterior, com uma "moeda" com M faces:

$$\Omega_{M,n} = \{1, \dots, M\}^n \quad p(\omega) = M^{-n} \quad \text{para cada } \omega$$

Se a ordem das extrações for irrelevante fica natural pensarmos nos resultados possíveis dando o número de vezes que cada bola é extraída, ou seja,

$$\tilde{\Omega}_{M,n} = \{(n_1, \dots, n_M) : n_i \in \{0, 1, \dots\}, \sum_{i=1}^M n_i = n\}.$$

Neste caso vemos que $\#\tilde{\Omega}_{M,n} = \binom{M+n-1}{n}$ e $p(\omega) = \binom{M+n-1}{n}^{-1}$

3. *n extrações (sem reposição) de urna com M bolas numeradas de 1 a M.*

Neste caso, obviamente $n \leq M$. Se olharmos o resultado de forma ordenada então teremos

$$\Omega_{M,n} = \{(\omega_1, \dots, \omega_n) : \omega_i \neq \omega_j, \text{ if } i \neq j\}$$

de modo que $\#\Omega_{M,n} = M(M-1) \dots (M-n+1)$. Se não estivermos interessados na ordem, podemos substituir o espaço amostral por $\tilde{\Omega}_{M,n}$ o conjunto de todos os subconjuntos de $\{1, \dots, M\}$, com n elementos, de modo que $\#\tilde{\Omega}_{M,n} = \binom{M}{n}$.

Na descrição acima utilizamos a linguagem de extrações de urnas. Outra forma sugestiva especialmente para física estatística seria a seguinte: temos compartimentos numerados de 1 a M e dispusemos n objetos nesses compartimentos. O caso de objetos distinguíveis (indistinguíveis) corresponde a olhar as extrações de forma ordenada (não-ordenada). Por outro lado, a possibilidade (ou impossibilidade) de cada compartimento receber mais de um objeto

corresponde a extrações com (ou sem) reposição. Em física estatística, a correspondência seria:

a) Partículas indistinguíveis, sujeitas ao princípio de exclusão de Pauli, correspondem à estatística de Fermi-Dirac (extrações sem reposição, não ordenadas); b) Partículas indistinguíveis, não sujeitas ao princípio de exclusão correspondem à estatística de Bose-Einstein (extrações com reposição, não ordenadas); c) Partículas distinguíveis, sem exclusão correspondem à estatística de Maxwell-Boltzman (extrações com reposição, ordenadas), restando o caso de extrações sem reposição e ordenadas sem uma interpretação em termos de partículas.

Exemplos importantes de espaços de probabilidade em que Ω não seja enumerável aparecerão mais tarde.

2 Probabilidade condicional. Independência estocástica.

Seja $(\Omega, \mathcal{A}, \mathcal{P})$ um espaço de probabilidade e $B \in \mathcal{A}$ com $P(B) > 0$. Independentemente de como chegamos ao espaço inicial, se acrescentamos àquela situação a informação extra dando conta que “ B ocorre”, como deveríamos atualizar nosso modelo? Uma breve reflexão nos leva a definir

$$P(A | B) = \frac{P(A \cap B)}{P(B)} \quad (2)$$

que chamamos de probabilidade condicional de A dado B . Se $P(B) = 0$ definimos $P(A | B)$ de modo arbitrário.

Observações:

(a) Segue-se da definição acima que

$$P(A \cap B) = P(A)P(B | A) = P(B)P(A | B) \quad \text{para quaisquer } A, B \in \mathcal{A}.$$

(b) Se $A_1, \dots, A_n \in \mathcal{A}$, então

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

(c) Sejam $A_1, \dots, A_n \in \mathcal{A}$, disjuntos dois a dois e tais que $\Omega = \cup_{i=1}^n A_i$, i.e. uma partição finita de Ω . Usando a propriedade de aditividade temos $P(B) = \sum_{i=1}^n P(B \cap A_i)$, e então por (a) temos a **fórmula da probabilidade total**:

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

(d) Nas condições do item anterior e se $P(B) > 0$ usando (a) novamente obtemos a **fórmula de Bayes**:

$$P(A_i | B) = \frac{P(A_i)P(B | A_i)}{\sum_{j=1}^n P(A_j)P(B | A_j)}. \quad (3)$$

A expressão acima tem interpretação de interesse para a estatística Bayesiana: os A_i representam hipóteses e $P(A_i)$, suas probabilidades *a priori*. O lado esquerdo de (3) fornece a probabilidade *a posteriori*, dada a ocorrência de B . Exemplos de aplicações desta idéia e fórmula serão vistos em aula. O leitor pode verificar que a fórmula se estende sem problemas se a partição de Ω for infinita enumerável, i.e. $\Omega = A_1 \cup A_2 \cup \dots$, com $A_i \in \mathcal{A}$ para todo $i \geq 1$, disjuntos dois a dois.

Exercício. Consideremos um experimento em duas etapas: inicialmente escolhemos uma moeda em uma urna e depois lançamos a moeda escolhida. Suponhamos que a urna possua três moedas, $m_i, i = 1, 2, 3$, onde m_3 é honesta e m_1, m_2 são desequilibradas dando cara com probabilidade $1/3$ e $2/3$, respectivamente. Calcule: (i) a probabilidade de escolher m_3 *a priori*; (ii) sabendo que a face superior da moeda lançada foi cara.

Definição. Se $A, B \in \mathcal{A}$ e $P(A \cap B) = P(A)P(B)$ dizemos que os eventos A e B são (estocasticamente) independentes. Mais geralmente, $A_1, \dots, A_n \in \mathcal{A}$ são ditos (estocasticamente) independentes se e só se

$$P(\cap_{i \in J} A_i) = \prod_{i \in J} P(A_i) \quad \text{para todo } \emptyset \neq J \subseteq \{1, \dots, n\} \quad (4)$$

Observações. (a) Vemos facilmente que se A_1, \dots, A_n são independentes então também B_1, \dots, B_n onde $B_i = A_i$ ou $B_i = A_i^c$.

(b) A condição (4) representa uma independência total, sendo mais forte do que, por exemplo, exigir que cada par de eventos A_i, A_j com $i \neq j$ sejam independentes (quando dizemos que os A_i são independentes aos pares). Sob (4) vale em particular o seguinte: se formarmos novos eventos B_1, \dots, B_k sendo cada um dos B_i obtido a partir de $A_j, j \in J_i$, por operações de uniões, interseções, diferenças, complementares, e os conjuntos de índices J_1, \dots, J_k forem disjuntos, então B_1, \dots, B_k também são independentes.

Exemplos.

Nos exemplos 1 e 2 da seção 1, se A_i é um evento que depende apenas do i -ésimo lançamento (i -ésima extração) então os eventos A_1, \dots, A_n são inde-

pendentes. Claramente, a independência deixa de valer se tivermos extrações sem reposição, como no exemplo 3 da seção 1.

Entretanto, se n for muito pequeno em relação a M , a dependência apresenta-se de forma fraca, algo que pode ser quantificado.

3 Variáveis Aleatórias. Esperança matemática.

Dentro da modelagem que estamos estudando, uma variável aleatória X representa um característico numérico do experimento. Por exemplo, no contexto do exemplo 2 da Seção 1, podemos estar interessados na frequência, ou no número total de observações da bola de um dado número, etc.

Como estaremos interessados em determinar probabilidades de eventos que dizem respeito a variáveis aleatórias, fica natural colocarmos a seguinte:

Definição. Uma variável aleatória (v.a.) é uma função $X: \Omega \rightarrow \mathbb{R}$ tal que $\{\omega: X(\omega) \leq x\} \in \mathcal{A}$, para todo $x \in \mathbb{R}$.

A cada variável aleatória podemos associar uma função chamada função de distribuição de X , definida através da equação:

$$F(x) = P(X \leq x) \quad \text{para todo } x \in \mathbb{R}, \quad (5)$$

onde adotamos a notação $\{X \leq x\} = \{\omega: X(\omega) \leq x\}$ e, ao escrevermos $P(\{X \leq x\})$, eliminamos os colchetes. Analogamente, $\{a < X \leq b\}$, etc.

Claramente, no caso de Ω finito (ou infinito enumerável) em que tipicamente teremos $\mathcal{A} = \mathcal{P}(\Omega)$, a classe de todos os subconjuntos de Ω , qualquer função a valores reais definida em Ω constitui uma variável aleatória. Em consequência das propriedades (P1), (P2) e (P3) vemos que toda função de distribuição goza das seguintes três propriedades: (F1) $F(x) \rightarrow 0$ se $x \rightarrow -\infty$, e $F(x) \rightarrow 1$ se $x \rightarrow +\infty$; (F2) se $x \leq y$ então $F(x) \leq F(y)$; (F3) F é contínua à direita. Escrevendo $F(x^-) = \lim_{y \uparrow x} F(y)$ vemos que $F(x) - F(x^-) = P(X = x)$, para todo x .

Exemplos. Consideremos n lançamentos independentes de uma moeda tal que p seja a probabilidade de obter cara, com $0 < p < 1$ (também chamados de *ensaios de Bernoulli*). Seja S_n o número de caras observadas. Podemos tomar $\Omega = \Omega_n$ como no ex.1 da seção 1 e neste caso $S_n = \sum_{i=1}^n \omega_i$. Temos

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \dots, n \quad (6)$$

conhecida pelo nome de distribuição Binomial com parâmetros n e p . Em termos das probabilidades acima escreva a função de distribuição desta v.a.

Seja $\lambda > 0$. Podemos tomar $\Omega = \mathbb{Z}_+$, o conjunto dos números inteiros não-negativos, e aí considerarmos uma probabilidade dada por

$$p(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad k = 0, 1, \dots, \quad (7)$$

conhecida como distribuição de Poisson com parâmetro λ .

Exercício: Considere S_n com distribuição Binomial com parâmetros n e p_n de modo que $\lim_{n \rightarrow +\infty} np_n = \lambda$, com $0 < \lambda < +\infty$. Mostre que para cada $k = 0, 1, \dots$, $\lim_{n \rightarrow +\infty} P(S_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}$

Por enquanto só vimos exemplos de espaços de probabilidade finitos ou enumeráveis, que claramente só podem suportar variáveis aleatórias que tomam valores em um conjunto também finito ou enumerável. Essas v.a. (e suas correspondentes funções de distribuição) são chamadas *discretas*. Lançando-se mão de espaços maiores como por exemplo $\Omega = [0, 1]$, o intervalo unitário na reta real \mathbb{R} , pode-se verificar que toda função satisfazendo as propriedades (F1),(F2) e (F3) acima é uma função de distribuição (f.d.). Exemplos importantes nas aplicações são as chamadas funções de distribuição *absolutamente contínuas*: são aquelas da forma

$$F(x) = \int_{-\infty}^x f(y) dy \quad (8)$$

onde $f(\cdot) \geq 0$ and $\int_{-\infty}^{+\infty} f(y) dy = 1$. Neste caso f é chamada de função densidade (da distribuição F , ou da v.a. X). Não nos preocupamos aqui com o sentido exato da integral, e para facilitar seria bom concentrar no caso em que f seja contínua por partes e temos uma integral (imprópria) de Riemann.

Exemplos: 1. Tomemos $f(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\}$, para todo $x \in \mathbb{R}$. Trata-se de uma função densidade e a distribuição associada é chamada de distribuição Normal, ou Gaussiana, padrão. A partir de uma v.a. X Gaussiana padrão constrói-se a família Gaussiana com parâmetros $\mu \in \mathbb{R}$ e $\sigma > 0$ fazendo-se a transformação afim $Y = \sigma X + \mu$. A interpretação estatística de μ e σ^2 vemos a seguir.

2. Seja $\lambda > 0$ e considere $f(x) = \lambda e^{-\lambda x}$ para $x > 0$ com $f(x) = 0$ se $x \leq 0$. Temos uma função densidade, e a distribuição correspondente é conhecida como distribuição exponencial. Compare essa distribuição com a de uma v.a. T

que representa, o número de tentativas necessárias até obter a primeira cara em sucessivos lançamentos de uma moeda, com probabilidade p de dar cara. (Proponha um espaço de probabilidade para esse “experimento”.)

3. Seja $f(x) = \frac{1}{\pi(1+x^2)}$. Então temos aqui também um densidade. A distribuição associada é conhecida como distribuição de Cauchy. Note que essa distribuição apresenta caudas $P(|X| > x)$, que decaem muito mais lentamente do que no caso Gaussiano ou exponencial.

Observações. Quando tivermos variáveis aleatórias (possivelmente definidas em espaços diferentes) que têm a mesma distribuição dizemos que são identicamente distribuídas (i.d.). É fácil ver que uma combinação convexa de f.d. é ainda uma f.d. (qual a interpretação probabilística). Sendo assim, cabe a pergunta: é toda f.d. uma mistura de uma f.d. discreta e uma f.d. absolutamente contínua. A resposta é negativa, uma vez que existem ainda as f.d. que são contínuas mas não são absolutamente contínuas. Conjuntos fractais tem aqui um papel importante. A distribuição de Cantor é um exempl ilustrativo. (Será visto em aula.)

Definição. Se X for uma v.a. que toma um número finito de valores, digamos x_1, \dots, x_k , então definimos sua esperança, ou valor esperado, como

$$EX = \sum_{i=1}^k x_i P(X = x_i). \quad (9)$$

Se o experimento puder ser repetido muitas vezes de modo independente (sob mesmas condições) e observarmos os valores sucessivos das observações da X_i (correspondente a v.a. X na i -ésima repetição) verificaremos que a média empírica observada em nossa amostra $\frac{1}{n} \sum_{i=1}^n X_i$ tipicamente não difere muito de EX , para n grande. Tal resultado clássico (lei dos grandes números) de certa forma explica o nome dado acima. (Isso vale em condições mais gerais, sempre que X for integrável, cf. definido a seguir.)

Para estender a definição acima para qualquer v.a. procede-se por um conveniente método de aproximação, chamada “aproximação horizontal”. Reside aí a diferença fundamental entre o modo como se define a integral de Riemann e essa integral chamada integral abstrata (sempre que temos um E.P. $(\Omega, \mathcal{A}, \mathcal{P})$). Tal construção abstrata é importante para garantir que o objeto assim definido, chamado ainda EX , ou integral abstrata $\int_{\Omega} X dP$, possua boas propriedades de convergência. Entretanto, para o cálculo efetivo de exemplos

usa-se sempre uma transformação que faz recair no cálculo de uma integral de Riemann (ou de Riemann-Stieltjes). Evitaremos entrar nos detalhes, mas em aula esboçaremos a construção. O cálculo de EX pode ser feito assim, em termos da f.d. de X :

(a) Se $X \geq 0$ temos $EX = \int_0^{+\infty} P(X > x)dx$;

(b) Se $X \leq 0$ colocamos $EX = -\int_{-\infty}^0 P(X \leq x)dx$

Se a f.d. de X for absolutamente contínua com densidade f vemos que isso se escreve como: $EX = \int_0^{+\infty} xf(x)dx$, quando $X \geq 0$; $EX = \int_{-\infty}^0 xf(x)dx$, se $X \leq 0$.

No caso geral, fazemos a decomposição $X = X^+ - X^-$ onde $X^+ = \max\{X, 0\}$, $X^- = -\min\{X, 0\} = (-X)^+$ e definimos $EX = EX^+ - EX^-$ sempre que não ocorrer $+\infty - +\infty$. Em particular, se X tiver densidade f

$$EX = \int_{-\infty}^{+\infty} xf(x)dx \quad (10)$$

se a integral for absolutamente convergente. No caso de uma das integrais $\int_0^{+\infty}$ ou de $\int_{-\infty}^0$ ser convergente, e a outra não, a esperança será infinita com o sinal correspondente. No caso de ambas as integrais serem divergentes não se define EX . Dizemos que X é integrável se $EX \in \mathbb{R}$.

Como antes não fomos aqui muito precisos em relação ao tipo de integral que temos, em $\int_0^{+\infty}$ ou $\int_{-\infty}^0$. Se f for contínua por partes como antes tratam-se de integrais (impróprias) de Riemann.

Exercício: calcular EX para X como nos exemplos anteriores.

Definição. Se X for uma v.a. integrável definimos a sua variância como $VarX = E(X - EX)^2$. O k -ésimo momento absoluto de X é dado por $E(|X|^k)$.

Definição. Dizemos que as v.a. X_1, \dots, X_n são independentes se para todo x_1, \dots, x_n , os eventos $\{X \leq x_1\}, \dots, \{X \leq x_n\}$ forem independentes.

Propriedades. (a) Se $P(X = c) = 1$ então $EX = c$.

(b) Se $X \leq Y$ então $EX \leq EY$

(c) $E(aX + bY) = aEX + bEY$, se X e Y são v.a. integráveis, $a, b \in \mathbb{R}$.

(d) $E(XY) = EXEY$ se X e Y são v.a. independentes e integráveis.

(e) Se $a > 0$ então $P(|X| \geq a) \leq a^{-k}E(|X|^k)$, para qualquer $k \geq 1$. Em particular

$$P(|X - EX| \geq a) \leq a^{-2}Var(X) \quad (11)$$

(f) Se X_1, \dots, X_n forem independentes e integráveis, então $Var(X_1 + \dots + X_n) = Var(X_1) + \dots + Var(X_n)$.

(f) Se X for absolutamente contínua com densidade f e se g for uma função razoavelmente regular então $g(X)$ também é uma v.a. e $Eg(X) = \int_{-\infty}^{+\infty} g(x)f(x)dx$. Como exercício obtenha a fórmula análoga para X v.a. discreta.

Exercício. Se S_n for uma v.a. com distribuição Binomial com parâmetros n e p , aplique (d) e (e) acima para concluir que $P(|\frac{S_n}{n} - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}$.

4 Lei dos Grandes Números. Teorema do Limite Central

Do exercício acima temos: a probabilidade que a proporção de caras em n lançamentos independentes difira de p (o valor esperado) por mais de $\varepsilon > 0$ tende a zero quando $n \rightarrow +\infty$, para cada $\varepsilon > 0$. Isto pode ser visto como uma manifestação de “regularidade estatística” e constitui um exemplo da chamada lei (fraca) dos grandes números.

Note que nessa argumentação o espaço de probabilidade pode depender de n , e de fato depende na construção feita até aqui (Ω_n do Ex.1 da Sec.1). Uma pergunta natural seria: podemos colocar todos esses experimentos dentro de um só espaço de probabilidade? Isto nos permitiria, talvez, verificar se

$$P\{\omega: \lim_{n \rightarrow +\infty} S_n(\omega)/n \rightarrow p\} = 1. \quad (12)$$

Vamos esboçar a construção do espaço de probabilidade em Ω_∞ , o conjunto das sequências infinitas $\omega = (\omega_1, \omega_2, \dots)$ com $\omega_i \in \{0, 1\}$ para todo i , e obter (12). Para isso começamos por identificar cada resultado de n lançamentos $(\tilde{\omega}_1, \dots, \tilde{\omega}_n)$ com o conjunto das sequências infinitas $\omega \in \Omega_\infty$ tais que $\omega_i = \tilde{\omega}_i$ se $i \leq n$ e nenhuma restrição é feita nos ω_j com $j > n$. Consequentemente, se $A_n \subseteq \Omega_n$ e $C_{A_n} = \{(\omega_1, \omega_2, \dots): (\omega_1, \dots, \omega_n) \in A_n\}$, definimos $P(C_{A_n}) = P^{(n)}(A_n)$. Trata-se de uma boa definição de P , porém restrita a uma classe bem pobre: dos eventos que dependem de um número finito de lançamentos. De fato, são estes os eventos que podemos observar, só que (11) não pode ser verificada neste contexto. Prova-se que tal P pode ser estendida (de modo único) à menor σ -álgebra que contém os eventos do tipo acima (prova omitida). Neste espaço $(\Omega_\infty, \mathcal{A}_\infty, P)$ verificamos (11) a seguir.

Das definições vemos que

$$\{\omega: S_n(\omega)/n \not\rightarrow p\} = \cup_{k \geq 1} \cap_{n \geq 1} \cup_{m \geq n} \left[\left| \frac{S_m}{m} - p \right| \geq \frac{1}{k} \right] \quad (13)$$

Observação. Sob (P1), (P2) e (P3) valem as seguintes propriedades:

- (i) Se $A_1, A_2, \dots \in \mathcal{A}$ tem-se $P(\cup_{n \geq 1} A_n) \leq \sum_{n \geq 1} P(A_n)$
(ii) Se $A_1 \supseteq A_2 \supseteq \dots$ e $A = \cap_{n \geq 1} A_n$ então $P(A_n) \downarrow P(A)$, quando $n \rightarrow +\infty$.
(iii) Se $A_1 \subseteq A_2 \subseteq \dots$ e $A = \cup_{n \geq 1} A_n$ então $P(A_n) \uparrow P(A)$, quando $n \rightarrow +\infty$.

De fato, prova-se que sob (P1), (P2) e (P3)', a condição (ii) para $A = \emptyset$ equivale a (P3).⁶⁾

Dessas propriedades fundamentais vemos que se

$$\sum_{m \geq 1} P \left[\left| \frac{S_m}{m} - p \right| \geq \frac{1}{k} \right] < +\infty \quad (14)$$

então $P(\cup_{m \geq n} [|\frac{S_m}{m} - p| \geq \frac{1}{k}]) \leq \sum_{m \geq n} P(|\frac{S_m}{m} - p| \geq \frac{1}{k})$ tende a zero quando $n \rightarrow +\infty$. E aplicando (ii) acima aos conjuntos $A_n = \cup_{m \geq n} [|\frac{S_m}{m} - p| \geq \frac{1}{k}]$, obtemos que para cada $k \geq 1$, $P(\cap_{n \geq 1} \cup_{m \geq n} [|\frac{S_m}{m} - p| \geq \frac{1}{k}]) = 0$. Aplicando (i) a esses últimos eventos obtemos (11).

Ao olharmos a estimativa dada no exercício vemos que ela não garante (14). Existe um pequeno truque que pode ser feito aqui: olhar a subsequência n^2 , mostrar a convergência ao longo dela, e depois mostrar que as coisas não ficam estragadas entre os termos sucessivos. Uma alternativa, com interesse por se, é verificar que de fato temos uma estimativa muito melhor do que no exercício. Fazemos isso agora.

Para facilitar a escrita tomamos $p = 1/2$, e deixaremos como exercício a extensão para $0 < p < 1$. Se considerarmos $0 < a < b < 1$ quaisquer (porém fixos no argumento abaixo) então vemos que

$$P\left(a \leq \frac{S_n}{n} \leq b\right) = 2^{-n} \sum_{k: k/n \in [a, b]} \binom{n}{k}. \quad (15)$$

Em particular

$$2^{-n} \max_{k: k/n \in [a, b]} \binom{n}{k} \leq P(a \leq S_n/n \leq b) \leq (n+1)2^{-n} \max_{k: k/n \in [a, b]} \binom{n}{k}$$

Vamos aqui usar a fórmula de Stirling³⁾

$$\sqrt{2\pi n} n^{n+\frac{1}{2}} e^{-n} e^{1/(12n+1)} \leq n! \leq \sqrt{2\pi n} n^{n+\frac{1}{2}} e^{-n} e^{1/(12n)}$$

que implica em particular que $n! \sim \sqrt{2\pi n} n^{n+\frac{1}{2}} e^{-n}$ quando $n \rightarrow +\infty$ (onde $a_n \sim b_n$ significa que $a_n/b_n \rightarrow 1$ quando $n \rightarrow +\infty$). Usando isso e (15) obtemos

(detalhes em aula)

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \log P(a \leq S_n/n \leq b) = - \inf_{x \in [a, b]} I(x) \quad (16)$$

onde para cada $x \in (0, 1)$ temos $I(x) = \log 2 + x \log x + (1 - x) \log(1 - x)$. Como a função I é não-negativa, contínua no intervalo $(0, 1)$ e se anula apenas em $x = 1/2$ podemos concluir que para $p = 1/2$ a probabilidade em (14) decai exponencialmente em n , para cada k fixado. Pode-se fazer argumento análogo para cada $0 < p < 1$, com $I_p(x) = x \log(\frac{x}{p}) + (1-x) \log(\frac{1-x}{1-p})$, que é estritamente positiva se $x \neq p$. (16) também vale se $a = 0$ ou $b = 0$, com $0 \log 0 = 0$.)

O resultado contido em (11) é um caso particular da chamada lei forte dos grandes números: se X_1, X_2, \dots são v.a. i.i.d, EX_1 finita e $S_n = \sum_{i=1}^n X_i$ então $S_n(\omega)/n \rightarrow EX_1$ quando $n \rightarrow +\infty$ para todo ω em certo conjunto Ω_0 com $P(\Omega_0) = 1$. (Diz-se que $S_n/n \rightarrow EX_1$ q.c.) Este resultado, que também é um caso particular do teorema ergódico de Birkhoff, tem muitas extensões: pode-se relaxar a independência para independência aos pares²⁾, ou então para variáveis que não sejam identicamente distribuídas.⁹⁾

Exemplo. A distribuição uniforme. Considere o “experimento”: escolher um ponto ao acaso no intervalo $[0, 1]$. Isto nos leva a atribuir probabilidade dada pela soma dos comprimentos das componentes para eventos que sejam uniões finitas de intervalos. Essa classe é o análogo dos eventos que dependem de finitos lançamentos no exemplo anterior. O análogo de \mathcal{A}_∞ constitui a chamada σ -álgebra de Borel. Para completar essa relação é conveniente pensar na expansão binária de $\omega \in [0, 1]$. A distribuição uniforme corresponde ao caso de moeda honeta. Explorar essa conexão, inclusive pensando em $p \neq 1/2$, é bastante instrutivo.

A partir da lei dos grandes números coloca-se naturalmente a seguinte pergunta: para n grande (porém finito) como são as flutuações de S_n/n em torno de EX_1 ? Um dos resultados mais clássicos diz que no contexto acima, e se $\sigma^2 = \text{Var}X_1$ for finita, então tipicamente $|S_n/n - EX_1|$ é de ordem $n^{-1/2}$, e ademais as probabilidades

$$P\left(\sqrt{n} \left(\frac{S_n}{n} - EX_1\right) \in [a, b]\right) \quad (17)$$

são aproximadas pela correspondente probabilidade para uma Gaussiana

$$\frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2\sigma^2}} dx \quad (18)$$

No caso em que as v.a. X_i correspondam a ensaios de Bernoulli independentes, com parâmetro p , esse resultado pode ser rephraseado como uma aproximação da Binomial com parâmetros n e p pela distribuição Gaussiana, e foi provado inicialmente por De Moivre ($p = 1/2$) e por Laplace ($p \neq 1/2$). Vale aqui uma aproximação do tipo seguinte:

$$\sup_{\{k: |k-np| \leq \psi(n)\}} \left| \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}}} - 1 \right| \rightarrow 0 \quad (19)$$

quando $n \rightarrow +\infty$, desde que $\psi(n)/(np(1-p))^{2/3} \rightarrow 0$, mostrando que a probabilidade de obter k sucessos em n ensaios de Bernoulli é aproximada pela função de densidade Gaussiana com esperança np e variância $np(1-p)$ ou seja a mesma esperança e variância que a distribuição Binomial.

Trata-se de caso particular do que atualmente chamamos de teorema *local* do limite central. Ao se somar em k para $|k - np| \leq x\sqrt{np(1-p)}$ obtém-se então a versão integral (ou usual) do mesmo, como acima. ^{6, 9)} A prova desse resultado clássico pode ser obtida utilizando-se a fórmula de Stirling e fazendo aproximações de Taylor para a função $I_p(x)$ usada acima ⁹⁾.

Note que no contexto de v.a. i.i.d. a validade da aproximação gaussiana depende de condições nas caudas da função de distribuição, formuladas quando se supõe que X_1 tem variância finita. A aproximação deixa de ser válida em geral. Para ver um exemplo disso suponha que a distribuição comum seja a de Cauchy. Nesse caso, algo mais dramático acontece: S_n/n tem exatamente a mesma distribuição que cada uma das v.a. X_1 : não só deixa de valer a aproximação Gaussiana, mas também não vale a lei dos grandes números. (Outros exemplos em aula.) Em sua forma clássica mais geral, o teorema do limite central caracteriza os possíveis limites de somas de um grande número v.a.s independentes, cada uma delas pequena (em relação ao total). Eles constituem as chamadas distribuições infinitamente divisíveis, tais que para cada n coincidem com a distribuição de uma soma de n v.a. i.i.d. Alguns exemplos são Poisson, Normal e Cauchy. A caracterização completa dessas distribuições através de sua transformada de Fourier, é devida a Lévy e a Khinchin. Alguns exemplos serão vistos em aula. ⁹⁾

5 Cadeias de Markov e passeios aleatórios.

Normalmente usamos a expressão processo estocástico para indicar uma evolução temporal que tenha componente aleatória ou estocástica. Pensando em observações realizadas de forma discreta no tempo isso nos leva a uma sequência de variáveis aleatórias (X_0, X_1, \dots) em algum espaço de probabilidade. Nesse contexto, o que mais se aproxima de uma evolução determinística é a chamada Propriedade de Markov, que diz basicamente o seguinte: *dado o presente, futuro e passado são independentes*. Supondo que as variáveis tomem valores em um conjunto S finito ou infinito enumerável,¹ isso pode ser assim formulado:

$$P(X_{n+1} = y \mid X_0 = x_0, \dots, X_n = x_n) = P(X_{n+1} = y \mid X_n = x_n). \quad (20)$$

Podemos simplificar, supondo que esse mecanismo de evolução seja homogêneo no tempo, ou seja,

$$P(X_{n+1} = y \mid X_n = x) = P(X_1 = y \mid X_0 = x), \quad (21)$$

e o chamamos de Cadeia de Markov (homogênea). Nessa situação consideramos a matriz de transição $Q = (q_{i,j})_{i,j \in S}$, onde

$$q_{i,j} = P(X_1 = j \mid X_0 = i) \quad i, j \in S, \quad (22)$$

que verifica as seguintes propriedades:

$$q_{i,j} \geq 0, \quad \sum_{k \in S} q_{i,k} = 1, \quad \forall i, j \in S \quad (23)$$

As matrizes que satisfazem (23) são chamadas *estocásticas*. Vemos que se tivermos a distribuição de X_0 , ou distribuição inicial, $\pi_i = P(X_0 = i)$, $i \in S$ então $P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \pi_{i_0} q_{i_0, i_1} \dots q_{i_{n-1}, i_n}$, ou seja toda a distribuição fica determinada a partir de π e da matriz q . Podemos também calcular as probabilidades de transição em n passos, ou seja se colocamos $q_{i,j}(n) = P(X_{m+n} = j \mid X_m = i)$ então vale a equação de Chapman-Kolmogorov

$$q_{i,j}^{(n+n')} = \sum_{k \in S} q_{i,k}^{(n)} q_{k,j}^{(n')} \quad i, j \in S, \quad q_{i,j}^{(1)} = q_{i,j} \quad (24)$$

¹Podemos pensar em $S = \mathbb{N} = \{1, 2, \dots\}$ ou $S = \{1, 2, \dots, N\}$ para algum $N \in \mathbb{N}$.

A verificação de (24) será feita em aula. Decorre imediatamente que $q^{(n)} = q^n$, a potência n -ésima da matriz q . Por consistência notacional é bom colocarmos $q^{(0)}$ como a matriz identidade, i.e. $q_{i,j}^{(0)} = \delta_{i,j}$, o delta de Kronecker.

Com isso vemos que se π é a distribuição inicial então a distribuição $\pi_i^{(n)} = P(X_n = i)$ na etapa n é dada por

$$\pi_i^{(n)} = \sum_{k \in S} \pi_k q_{k,i}^n. \quad (25)$$

Colocam-se as seguintes questões básicas:

a) Existe uma distribuição de equilíbrio (ou estacionária) $\bar{\pi}$? Ou seja uma probabilidade $\bar{\pi}$ em S tal que $\pi^{(n)} = \bar{\pi}$ para todo n . Claramente isso equivale a encontrar solução para a *equação de balanceamento*

$$\pi_j = \sum_{i \in S} \pi_i q_{i,j} \quad \forall j, \quad (26)$$

com $\pi_i \geq 0$ and $\sum_{i \in S} \pi_i = 1$.

b) Em que condições tal distribuição estacionária $\bar{\pi}$ é única?

Se S for finito vemos que isso se transforma em um problema de autovalor da matriz finita q : $\pi = \pi q$ (autovetor à esquerda com autovalor 1), e pode ser tratado algebricamente, constituindo-se um caso do clássico teorema de Perron-Frobenius. Tem-se unicidade da distribuição estacionária sempre que a cadeia for *irredutível*, no sentido que para quaisquer dois pontos distintos $i, j \in S$ existe $n \geq 1$ tal que $q_{i,j}^n > 0$, i.e., há probabilidade positiva de chegar a j em n passos, partindo de i .

No caso de unicidade, põe-se naturalmente a questão de convergência ao equilíbrio. Se começamos de qualquer outra distribuição π temos convergência a $\bar{\pi}$? Ou seja, $\pi_i^{(n)} \rightarrow \bar{\pi}_i$ para todo i ? No caso afirmativo interessa estimar a velocidade de convergência. Aparece aqui uma particularidade proveniente da discretização do tempo: para termos convergência é necessário evitar a periodicidade, presente nos exemplos abaixo (ou reformular ligeiramente a convergência). Dizemos que uma cadeia é aperiódica se $d_i := m.d.c.\{n: q_{i,i}^n > 0\} = 1$ para todo i . Sendo a cadeia irredutível basta considerar um i fixo qualquer (d_i é constante neste caso). Nessas condições pode-se dar um argumento probabilístico, baseado na idéia de acoplamento: se a cadeia for finita, irredutível e aperiódica, e consideramos duas realizações evoluindo independentemente até se encontrarem (permanecendo depois inseparáveis),

então existem $r \in (0, 1)$ e $c < +\infty$ tais que, para todo n , a probabilidade de não se encontrarem até a etapa n é menor ou igual a cr^{-n} . Isso não só prova a existência e unicidade da distribuição estacionária $\bar{\pi}$ mas também fornece verificação da convergência exponencial a $\bar{\pi}$, para qualquer situação inicial. A razão pela qual o argumento acima funciona provém da validade da *condição de Doeblin*: sendo S finito as condições de irreducibilidade e aperiodicidade são equivalentes à existência de $m \geq 1$ e $\varepsilon > 0$ tal que

$$q_{i,j}^m \geq \varepsilon \quad \text{para todo } i, j \in S$$

e nesse caso o argumento (acoplamento) acima funciona com $r = (1 - \varepsilon)^{1/m}$ e $c = (1 - \varepsilon)^{-1}$. Percebe-se a necessidade de melhor operacionalizar uma estimativa de r , principalmente porque as aplicações envolvem situações em que o número de pontos em S é muito grande. (Discussão em aula.) Essa convergência é importante para algoritmos muito usados como o *amostrador de Gibbs* bem como para o *recozido simulado*, (este último é cadeia não-homogênea) de importância na análise de grandes conjuntos de dados, especialmente em questões ligadas a reconhecimento de imagens.

No caso de cadeias periódicas (como os exemplos abaixo) a convergência acontece ao longo de subsequências adequadas, e no limite em Cesaro:

$$\frac{1}{n} \sum_{k=1}^n \pi_i^{(k)} \rightarrow \bar{\pi}_i \quad \forall i, \quad (27)$$

para qualquer distribuição inicial. Em outras palavras, o valor esperado da proporção de tempo gasto em cada estado i tende a $\bar{\pi}_i$, no limite de tempo infinito. (A análise em tempo contínuo evita a questão da periodicidade.)

De fato, o argumento baseado no acoplamento fornece não apenas um resultado sobre o valor médio assintótico do tempo passado em cada estado i . Como na lei forte dos grandes números, tem-se o seguinte: para um conjunto de realizações que tem probabilidade um, a proporção de tempo passado em i converge a $\bar{\pi}_i$ quando o tempo de observação tende a infinito (em particular, a cadeia retorna infinitas vezes a cada estado i). De fato, pela propriedade (forte) de Markov vemos que se $X_0 = i$ e τ_1, τ_2, \dots são os instantes dos sucessivos retornos ao estado inicial:

$$\tau_{n+1} = \min\{j \geq \tau_n + 1: X_j = X_{\tau_n}\}, \quad n \geq 0 \quad (28)$$

onde $\tau_0 = 0$, então as v.a. $\tau_1, \tau_2 - \tau_1, \tau_3 - \tau_2, \dots$ são i.i.d. Aplicando a lei dos grandes números teremos que τ_n/n converge q.c. a $E\tau_1$,² de onde obtemos a citada convergência e, ademais, comparando com o argumento anterior, temos:

$$\mu_i := E(\tau_1 | X_0 = i) = \frac{1}{\bar{\pi}_i} \quad (29)$$

Notação. $P_i(\cdot) = P(\cdot | X_0 = i)$. $P_\pi(\cdot)$ indica que X_0 tem distribuição π .

Para o tratamento análogo se tivermos uma cadeia irredutível, mas com S for infinito, põe-se imediatamente o seguinte conceito:

Definição. Um estado i é dito transitório se $P(\tau_1 < +\infty | X_0 = i) < 1$, i.e., existe uma probabilidade positiva de não retornar a i . Caso contrário dizemos que i é recorrente. No caso de um estado recorrente, o classificamos como recorrente nulo (positivo) de acordo com $\mu_i = +\infty$ ($\mu_i < +\infty$, respectivamente).

Prova-se facilmente que se a cadeia for irredutível então todos os estados são do mesmo tipo, todos transitórios, recorrente nulos ou recorrente positivos, e dizemos que a cadeia é transitória, recorrente nula ou recorrente positiva, respectivamente. Uma cadeia irredutível e recorrente positiva é dita *ergódica*. Se a cadeia for ergódica ela possui exatamente uma probabilidade estacionária $\bar{\pi}$, sendo que $\bar{\pi}_i$ coincide com a proporção assintótica (q.c.) de tempo gasto no estado i , e $\mu_i = 1/\bar{\pi}_i$, para todo i .

Exemplos. 1. *Cadeia de Ehrenfest.* Bolas numeradas de 1 a m são distribuídas em duas urnas, A e B . Em cada etapa sorteia-se ao acaso um número em $\{1, \dots, m\}$ e a bola correspondente é trocada de urna. Seja X_n o número de bolas na urna A após o n -ésimo sorteio, e X_0 a situação inicial. Temos uma cadeia de Markov em $S = \{0, 1, \dots, m\}$ com matriz de transição dada por:

$$q_{i,i+1} = 1 - \frac{i}{m}, \quad q_{i,i-1} = \frac{i}{m}, \quad 1 \leq i \leq m \quad (30)$$

sendo que os demais termos se anulam. Nesse caso verifica-se facilmente que a distribuição Binomial com parâmetros m e $1/2$ é estacionária. De fato verifica-se que se $\pi_i = \binom{m}{i} 2^{-m}$, então

$$\pi_i q_{i,j} = \pi_j q_{j,i} \quad \forall i, j \in S. \quad (31)$$

Somando em i na equação (31) e lembrando que $\sum_i q_{j,i} = 1$ para todo j vemos que (31) implica (26). A equação (31) é também conhecida como *equação de*

²Para v.a. não-negativas a lei forte valeria mesmo se τ_1 não fosse integrável.

balanceamento detalhado. Ela fornece a reversibilidade temporal: se π é uma probabilidade em S e satisfaz (31), e consideramos a cadeia com distribuição inicial π então $P_\pi(X_0 = i_0, \dots, X_n = i_n) = P_\pi(X_n = i_0, \dots, X_0 = i_n)$, para quaisquer $n \geq 1$, i_0, \dots, i_n , ou seja podemos inverter a direção do tempo e a distribuição fica inalterada.

Lembrando então (29) e a fórmula de Stirling podemos ver que para m grande $\mu_{\lfloor m/2 \rfloor} \sim cm^{1/2}$ ($c = \sqrt{\pi/2}$), ao passo que $\mu_m = 2^m$. Com sua extrema simplicidade, este exemplo teve um papel curioso no início do século passado, quando ainda não estava claro o aspecto probabilístico envolvido no Teorema H de Boltzmann, e críticos apontavam a incompatibilidade entre o crescimento da entropia e a reversibilidade do sistema microscópico (leis da mecânica, no caso). Basta olhar a versão “microscópica” da cadeia de Ehrenfest (diz em que urna está cada bola, a cada etapa): temos nesse caso uma cadeia de Markov em $\tilde{S} = \{0, 1\}^m$ e a distribuição estacionária (também reversível) é uniforme em \tilde{S} . O fator entrópico (número de configurações $x \in \tilde{S}$ correspondentes a um dado valor k de X_n) é que faz a diferença acima nos tempos de recorrência.

2. *Passeios aleatórios.* O nome indica cadeias obtidas como somas parciais de v.a. i.i.d.: $X_n = X_0 + \sum_{k=1}^n \xi_k$ onde as ξ_k são i.i.d. Ou seja, em cada etapa faz-se um deslocamento totalmente independente de tudo já ocorrido, com a mesma distribuição de ξ_1 .

(a) $S = \mathbb{Z}$ o conjunto dos números inteiros. Suponha que $P(\xi_1 = +1) = p$ and $P(\xi_1 = -1) = q = 1 - p$ com $0 < p < 1$. Aplicando a lei dos grandes números vemos que $X_n/n \rightarrow 1 - 2p$ q.c., para qualquer valor inicial. Logo, se $p \neq 1/2$ fica evidente que temos uma cadeia transitória, cada estado é visitado apenas um número finito de vezes, com probabilidade um. Por outro lado, se $p = 1/2$ pode-se verificar que $P(\tau_1 < +\infty) = 1$ mas $E\tau_1 = +\infty$ ou seja temos uma situação de recorrência nula, e não existe probabilidade estacionária.

(b) Seja $S = \mathbb{Z}^d$, o reticulado d -dimensional. Em cada etapa, estando na posição x escolhe-se um dos $2d$ vizinhos mais próximos de x , com igual probabilidade $1/(2d)$, independentemente de todo o resto, e esse passa a ser o novo estado do processo. Se $d = 2$ tem-se a mesma situação do caso $d = 1$ (exemplo anterior com $p = 1/2$) de recorrência nula. Entretanto se $d \geq 3$ a cadeia passa a ser transitória. Vamos verificar isso em aula. Uma maneira de se convencer da

diferença seria considerar o seguinte: Fixado o valor de d , temos

$$P_0(X_{2n} = 0) = \sum_{k_1+k_2+\dots+k_d=n} \binom{2k_1}{k_1} \dots \binom{2k_d}{k_d} (2d)^{-2n}; \quad P_0(X_{2n+1} = 0) = 0.$$

Usando novamente a formula de Stirling podemos ver que

$$\sum_{n=1}^{+\infty} P_0(X_n = 0) < +\infty \quad \Leftrightarrow \quad d \geq 3.$$

Observando que $\sum_{n=1}^{+\infty} P_0(X_{2n} = 0) = E_0(\sum_{n \geq 1} \mathbf{1}_{\{X_n=0\}})$ vemos que se $d \geq 3$ o passeio aleatório faz apenas um número finito de retornos a origem, ao longo de toda a trajetória, sendo portanto transitório. (Os casos $d = 1, 2$ serão vistos em aula.)

6 Do gás de Lorenz à percolação.

Nessa seção vamos considerar um problema (clássico) do movimento de uma partícula em um ambiente onde há obstáculos refletores. O problema começou em 1905 com Lorentz ³, e sua motivação era descrever e estudar teoricamente o movimento de um elétron através de um campo de partículas com massas.

Observação. Desde o início de seu trabalho científico, Lorentz envolveu-se na tarefa de estender a teoria de Maxwell da eletricidade e da luz. Seu trabalho fundamental nos campos de ótica e eletricidade revolucionaram as concepções contemporâneas sobre a natureza da matéria. Lorentz ainda fez contribuições fundamentais ao estudo de problemas de corpos em movimento, em especial sobre fenômenos óticos e elétricos neste contexto. Sua visão de que elétrons desempenham um papel ponderável em fenômenos eletro-magnéticos tornou possível aplicar a teoria molecular à teoria da eletricidade, e explicar o comportamento de ondas de luz atravessando corpos transparentes em movimento.

Numa nomenclatura atual, os obstáculos estariam representados por bolas unitárias com massa infinita (os centros das bolas tomados de forma determinística ou aleatória, e.g. Processo de Poisson), e o elétron ficaria representado por uma partícula pontual, que (na ausência de campo elétrico ou magnético) movia-se com velocidade constante entre sucessivas colisões com os obstáculos,

³Hendrik Antoon Lorentz (1853 - 1928), Prêmio Nobel em 1902, juntamente com Pieter Zeeman, por sua teoria de radiação eletromagnética.

e as colisões supostas elásticas (ver Fig 1. (a) e (b)). Este modelo ficou conhecido pelo nome “gás de Lorentz”.⁸⁾ Como o problema apresentou-se muito difícil para dar uma resposta razoável a nível rigoroso, Ehrenfest sugeriu mudar as formas dos obstáculos para “diamantes” de modo que se o elétron começasse com velocidade horizontal, então apenas as direções horizontal e vertical seriam permitidas para o movimento, cf. Fig. 1(c).¹⁾ Entretanto, mesmo essa simplificação não ajudou tanto, e ainda hoje estamos na mesma posição em relação a este problema que quando começamos em 1905: apesar dos enormes esforços de matemáticos e físico-matemáticos, nenhuma propriedade ergódica deste sistema dinâmico é conhecida (exceto no caso periódico, quando os objetos estão colocados em um modo adequado e periódico, e o sistema tem o assim chamado horizonte infinito, i.e., são permitidas algumas direções excepcionais nos quais vôos não terão colisões (resultados de Blecher e Sinai). Nem tampouco sabemos se a partícula escapa ao infinito, e no caso afirmativo, com que velocidade, e muito menos, se sua trajetória re-escalonada (espaço-tempo) deve convergir a um movimento Browniano.

Na comunidade de probabilidade, esse modelo provocou um grande número de problemas, a maioria dos quais ainda grandemente não respondidos. Vamos nos concentrar no chamado “problema de espelhos”.

O problema de espelhos. Seja $0 \leq p \leq 1$. Em cada vértice do reticulado \mathbb{Z}^2 colocamos um *espelho* ou uma *abertura*, com probabilidades p e $1 - p$, respectivamente. As escolhas em vértices distintos são todas independentes entre si. O plano de cada espelho colocado é sempre perpendicular ao plano do reticulado e o ângulo entre o espelho e o eixo horizontal do reticulado será $3\pi/4$ (quando dizemos que o dito vértice é um espelho do tipo NW) ou $\pi/4$ (quando dizemos que o dito vértice é um espelho do tipo NE), com probabilidade $1/2$ cada, independentemente de todo o resto. Os espelhos são pensados como de dupla-face. Imaginamos agora um raio de luz emitido na origem do reticulado em uma das quatro direções no reticulado. Ao alcançar uma abertura, o raio passa através da mesma, sem ser defletido. Ao alcançar um espelho, é defletido através do ângulo e direção apropriados. A seguinte dicotomia é então bastante clara: ou a luz atravessa um caminho infinito a partir da origem (possivelmente contendo algumas auto-interseções), ou descreve um circuito fechado.

Escrevemos $\xi(p)$ para a probabilidade de que a luz alcance um conjunto infinito de vértices. Uma questão básica consiste em saber se $\xi(p) > 0$ para um

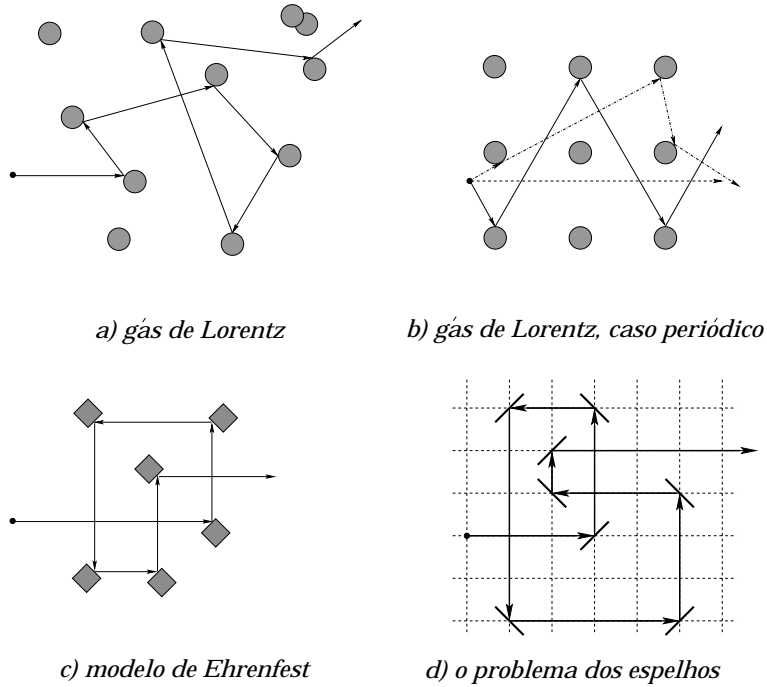


Figura 1:

dado valor de p .

Nessa versão do modelo de Lorentz no reticulado, os espelhos representam as partículas massivas e a luz representa o elétron. Técnicas de percolação são relevantes para tais problemas.

Determinar se $\xi(p) > 0$ para um dado valor de p é um problema notoriamente difícil. O ponto central da dificuldade está no fato de constituir um sistema dinâmico em um ambiente aleatório. Em outras palavras, fixando-se uma realização de espelhos, o comportamento é determinístico; entretanto, a trajetória é muito sensível a pequenas mudanças na configuração dos espelhos. É trivial que $\xi(0) = 1$: na ausência total de espelhos, a luz vai diretamente ao infinito. Por outro lado, o único resultado não-trivial provado rigorosamente sobre esta função é que $\xi(1) = 0$. Simulações numéricas sugerem que $\xi(p) = 0$ para todo $p > 0$.

O fato que $\xi(1) = 0$ depende de um resultado importante devido a Har-

ris: para a percolação de elos em \mathbb{Z}^2 com parâmetro $1/2$, a probabilidade de encontrar um caminho infinito aberto é zero: $\theta(1/2) = 0$.⁵⁾ Assumindo por enquanto esse fato, vamos ver como ele implica que $\xi(1) = 0$.

Prova de que $\xi(\mathbf{1})=0$. Seja $\widehat{\mathbb{Z}}^2$ o “reticulado diagonal”, tendo como vértices os pontos $(m + \frac{1}{2}, n + \frac{1}{2})$ para $m, n \in \mathbb{Z}$ com $m + n$ pares; no qual coloca-se um elo ligando $(m + \frac{1}{2}, n + \frac{1}{2})$ a $(r + \frac{1}{2}, s + \frac{1}{2})$ se e somente se $|m - r| = |n - s| = 1$. $\widehat{\mathbb{Z}}^2$, como na Figura 2.

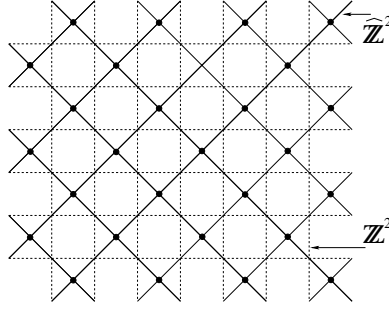


Figura 2: Reticulados \mathbb{Z}^2 (linhas pontilhadas) e $\widehat{\mathbb{Z}}^2$

A partir dos espelhos no reticulado original \mathbb{Z}^2 vamos gerar um processo de percolação em $\widehat{\mathbb{Z}}^2$. Fazemos isso da seguinte forma: um elo em $\widehat{\mathbb{Z}}^2$ ligando $(m - \frac{1}{2}, n - \frac{1}{2})$ a $(m + \frac{1}{2}, n + \frac{1}{2})$ é declarado aberto se o vértice (m, n) of \mathbb{Z}^2 tem um espelho NE; analogamente, um elo em $\widehat{\mathbb{Z}}^2$ ligando $(m - \frac{1}{2}, n + \frac{1}{2})$ a $(m + \frac{1}{2}, n - \frac{1}{2})$ é declarado aberto se (m, n) for um espelho NW. Como $\widehat{\mathbb{Z}}^2$ é isomorfo ao reticulado quadrado, o processo resultante é o modelo de percolação de elos no reticulado quadrado com parâmetro $1/2$. (Ver Fig. 3 e 4.)

Observação 1. Um pouco da notação usada em percolação. Se x e y são vértices em \mathbb{Z}^2 , escrevemos $x \leftrightarrow y$ se existir um caminho aberto ligando x a y , i.e. um caminho formado de elos abertos. Então podemos definir $C(x) = \{y \in \mathbb{Z}^2 : x \leftrightarrow y\}$ o *aglomerado aberto* de x , e a *probabilidade de percolação* $\theta(p) = P_p(|C(0)| = +\infty)$, onde p é a probabilidade de cada elo estar aberto.

Fato fundamental para percolação é a existência de valor crítico $p_c = p_c(d) \equiv p_c(\mathbb{Z}^d)$ tal que

$$\theta(p) = 0 \quad \text{se} \quad p < p_c,$$

and

$$\theta(p) > 0 \quad \text{se} \quad p > p_c;$$

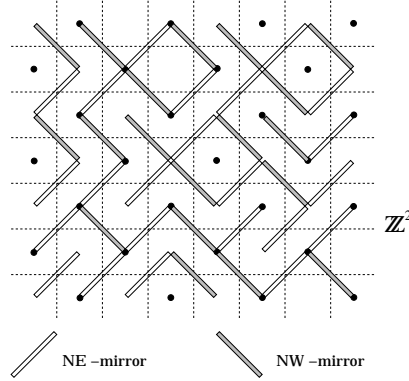


Figura 3: Configuração de espelhos em \mathbb{Z}^2 com $p = 1$: espelho presente em cada vértice.

$p_c(d)$ é chamada de *probabilidade crítica* e formalmente definida como

$$p_c(d) = \sup\{p : \theta(p) = 0\}.$$

sendo que $0 < p_c(d) < 1$, para todo $d \geq 2$.

Vamos utilizar o seguinte lema fundamental:

Lemma. (Harris) Para percolação de elos em \mathbb{Z}^2 temos que $\theta(1/2) = 0$. Em particular $p_c(\mathbb{Z}^2) \geq 1/2$.

Prova. Ver apêndice.

Observação 2. O Lema acima foi provado inicialmente com argumentos de certa complexidade geométrica, no notável artigo ⁵⁾, que também contém muitas das técnicas usadas posteriormente nos desenvolvimentos em percolação. Como sub-produto Harris obteve a unicidade do aglomerado infinito em $d = 2$. No Apêndice apresentamos uma prova simples e elegante de que $\theta(1/2) = 0$. O argumento, devido a Zhang⁴, assume já provado que se $\theta(p) > 0$ então o aglomerado infinito é único em \mathbb{Z}^2 (este último fato vale em \mathbb{Z}^d , $d \geq 2$). ⁴⁾

De fato se $d = 2$ pode-se dizer mais, na forma seguinte.

Teorema. (Kesten) $p_c(\mathbb{Z}^2) = 1/2$.

⁴Comunicação não publicada.

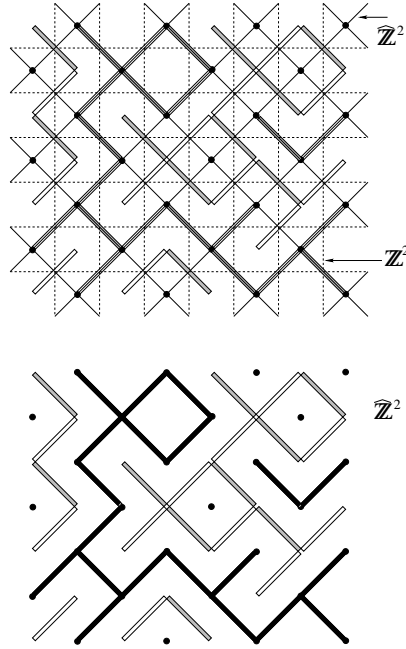


Figura 4: Espelhos no reticulado original \mathbb{Z}^2 (parte superior da figura) geram um processo de percolação de elos em $\tilde{\mathbb{Z}}^2$ (figura inferior). Elos pretos estão abertos, i.e. alinhados com os espelhos e elos ortogonais aos espelhos brancos estão fechados.

Este resultado é um dos mais fundamentais na história da percolação. A prova de Kesten foi o coroamento de quatro artigos publicados em um período de vinte e um anos. Harris ⁵⁾ provou que $\theta(1/2) = 0$, e em 1978 trabalhos independentes (em grande parte equivalentes) de Russo e de Seymour e Welsh forneceram os ingredientes necessários sobre o tamanho médio do aglomerado. Kesten ⁷⁾ mostrou como construir o resultado completo, com base nesses argumentos.

Observação 4. Um conceito útil é a dualidade planar. Para um grafo que pode ser desenhado no plano de modo que os elos se intersectam apenas nos vértices, também chamado de grafo planar, o dual pode ser definido assim: em cada face do grafo colocamos um vértice do dual e ligamos dois desses vértices com um elo quando as faces compartilham um elo no grafo original. No caso particular de \mathbb{Z}^2 isso nos dá simplesmente $\tilde{\mathbb{Z}}^2 = \{(m + \frac{1}{2}, n + \frac{1}{2}) : (m, n) \in \mathbb{Z}^2\}$,

com elos entre os vizinhos, isomorfo ao reticulado original. Cada elo de \mathbb{Z}^2 é cruzado por um único elo no dual (e vice-versa). Para o modelo de percolação dual, dizemos que um dado elo está aberto (fechado) se e só se o elo que cruza em \mathbb{Z}^2 está aberto (fechado) respectivamente.

Voltamos agora à prova de que $\xi(1) = 0$. Como, pelo Lema 1, $\theta(1/2) = 0$, isso implica por dualidade (e simetria entre fechado/aberto) que a origem de \mathbb{Z}^2 fica contida q.c. no interior de um circuito aberto C em $\widehat{\mathbb{Z}^2}$. Cada elo de C corresponde a um espelho (paralelo ao elo) no vértice correspondente do reticulado original, e portanto o circuito C dá origem a um contorno de espelhos em torno da origem. A luz não tem como escapar desse contorno e portanto $\xi(1) = 0$.

Apêndice. Prova do Lema. Seguimos um argumento de Yu Zhang. Suponha que $\theta(1/2) > 0$. Para cada inteiro positivo n , seja $A_{(E,n)}$ ($A_{(D,n)}$, $A_{(S,n)}$, $A_{(I,n)}$, respectivamente) o evento que algum vértice no lado esquerdo (direito, superior, inferior, respectivamente) do quadrado $\Lambda_n = [0, n]^2$ pertença a um caminho infinito aberto em \mathbb{Z}^2 que não usa nenhum outro vértice de Λ_n . Claramente $A_{(E,n)}$, $A_{(D,n)}$, $A_{(S,n)}$, $A_{(I,n)}$ são eventos *crescentes*,⁵ todos com mesma probabilidade e cuja união é o evento que algum vértice de Λ_n pertence a um aglomerado aberto infinito. Devido à hipótese que $\theta(1/2) > 0$ existe um aglomerado aberto infinito com probabilidade 1, de modo que

$$P_{1/2}(A_{(E,n)} \cup A_{(D,n)} \cup A_{(S,n)} \cup A_{(I,n)}) \longrightarrow 1 \quad \text{as } n \rightarrow \infty. \quad (32)$$

Pelo “truque da raiz” segue-se que

$$\lim_{n \rightarrow +\infty} P_{1/2}(A_{(u,n)}) = 1 \quad \text{para } u = E, D, S, I. \quad (33)$$

Observação 4. Truque da raiz. Se A_1, A_2, \dots, A_m forem eventos crescentes, de mesma probabilidade,

$$P_p(A_1) \geq 1 - \{1 - P_p(\cup_{i=1}^m A_i)\}^{1/m}.$$

Para ver isso, note que usando a desigualdade FKG⁵, 4) temos

$$1 - P_p(\cup_{i=1}^m A_i) = P_p(\cap_{i=1}^m A_i^c) \geq \prod_{i=1}^m P_p(A_i^c) = (1 - P_p(A_1))^m.$$

⁵Crescente na configuração de elos.

Escolhemos N de forma que

$$P_{1/2}(A_{(u,N)}) > \frac{7}{8} \quad \text{para } u = E, D, S, I. \quad (34)$$

Indo ao reticulado dual $\tilde{\mathbb{Z}}^2 = \{(m + \frac{1}{2}, n + \frac{1}{2}) : (m, n) \in \mathbb{Z}^2\}$, definimos a caixa dual $\tilde{\Lambda}_n = \{x + (1/2, 1/2) : 0 \leq x_1, x_2 \leq n\}$.

Seja $\tilde{A}_{(E,n)}$ ($\tilde{A}_{(D,n)}$, $\tilde{A}_{(S,n)}$, $\tilde{A}_{(I,n)}$, respectivamente) o evento que algum vértice no lado esquerdo (direito, superior, inferior, respectivamente) de $\tilde{\Lambda}_n$ pertence a um caminho infinito fechado em $\tilde{\mathbb{Z}}^2$ que não usa outro vértice de $\tilde{\Lambda}_n$. Cada elo em $\tilde{\mathbb{Z}}^2$ está fechado com probabilidade $1/2$, de modo que

$$P_{1/2}(\tilde{A}_{(u,N)}) = P_{1/2}(A_{(u,N)}) > \frac{7}{8} \quad \text{for } u = E, D, S, I, \quad (35)$$

por (33). Considere agora o evento $A = A_{(E,N)} \cap A_{(D,N)} \cap \tilde{A}_{(S,N)} \cap \tilde{A}_{(I,N)}$, no qual existem caminhos infinitos abertos em $\mathbb{Z}^2 \setminus \Lambda_N$ tocando os lados esquerdo e direito de Λ_N , e caminhos infinitos fechados em $\mathbb{Z}^2 \setminus \tilde{\Lambda}_N$ tocando os lados superior e inferior de $\tilde{\Lambda}_N$ cf. Fig. 5.

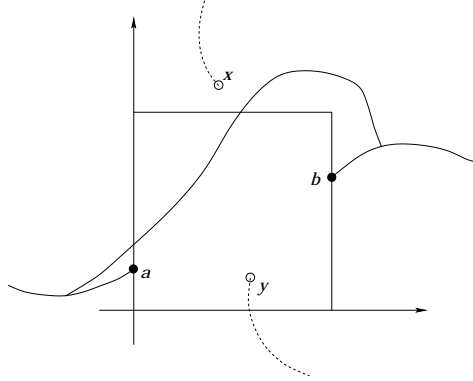


Figura 5: Vértices a e b pertencem a aglomerados abertos infinitos em $\mathbb{Z}^2 \setminus \Lambda_N$; vértices x e y a aglomerados fechados infinitos em $\tilde{\mathbb{Z}}^2 \setminus \tilde{\Lambda}_N$. Se existe um único aglomerado infinito aberto, então existe um caminho aberto π ligando a a b , e portanto os aglomerados infinitos fechados em x e y são disjuntos.

Por (34) e (35), vemos que a probabilidade de A não ocorrer, $P(A^c)$ satisfaz

$$P_{1/2}(A^c) < \frac{1}{2},$$

de modo que $P_{1/2}(A) > \frac{1}{2}$. Se A ocorre, então $\mathbb{Z}^2 \setminus \Lambda_N$ contém dois aglomerados abertos infinitos disjuntos, uma vez que os aglomerados em questão estão fisicamente separados por um caminho fechado infinito do dual; cada caminho aberto em $\mathbb{Z}^2 \setminus T(N)$ ligando esses dois aglomerados conteria um elo que cruza um elo fechado no dual, e um tal elo não pode existir. De forma semelhante, em A , o grafo $\tilde{\mathbb{Z}}^2 \setminus \tilde{\Lambda}_N$ contém dois aglomerados infinitos fechados, separados fisicamente por um caminho aberto infinito em $\mathbb{Z}^2 \setminus \Lambda_N$. Agora, o reticulado total \mathbb{Z}^2 contém (quase certamente) um *único* aglomerado infinito aberto, e segue-se que existe (quase certamente em A) uma conexão aberta do reticulado entre os acima mencionados aglomerados abertos infinitos. Pela geometria da situação (ver Fig. 5 novamente), essa conexão atravessa Λ_N e forma uma barreira para possíveis conexões do dual, ligando os dois aglomerados infinitos fechados. Portanto, quase certamente em A , o reticulado dual contém dois ou mais aglomerados infinitos fechados. Esse último evento tem probabilidade zero, e portanto $P_{1/2}(A) = 0$, em contradição com a dedução acima. A hipótese inicial que $\theta(1/2) > 0$ é portanto incorreta, concluindo a prova do Lema.

Referências

1. Ehrenfest, P. *Collected scientific papers*. (M.J. Klein, ed.), North-Holland, Amsterdam. 1959.
2. Etemadi, N. An elementary proof of the strong law of large numbers. *Z. Wahrsch. Verw. Gebiete* 55, no. 1, 119–122, 1981.
3. Feller, W. *An Introduction to Probability Theory and Its Applications*. 3a. edição, 1967.
4. Grimmett, G. *Percolation*. Second edition. Springer. Heidelberg 1999.
5. Harris, T.E. A lower bound for the critical probability in a certain percolation process, *Proceedings of the Cambridge Philosophical Society* 56, 13-20. 1960.
6. James, B. *Probabilidade: um curso em nível intermediário*. Projeto Euclides. IMPA, segunda edição. 2002.
7. Kesten, H. The critical probability of bond percolation on the square lattice equals $1/2$. *Communications in Mathematical Physics* 74, 41-59. 1980.

8. Lorentz, H.A. The motion of electrons in metallic bodies, I, II, III, Koninklijke Akademie van Wetenschappen de Amsterdam, Section of Sciences 7, 438-453, 585-593, 684-691. 1905.
9. Shiryaev, A. N. *Probability*. Springer-Verlag New York, 1989, 2a. edição. 1a edição - 1980.