

Lições de Teoria Ergódica, Processos Estocásticos e Sistemas Dinâmicos

Nelson Luís Dias

30 de março de 2014

Capítulo 1

Introdução

Em 1895, Osborne Reynolds publicou um artigo sobre escoamentos turbulentos no qual pela primeira vez aparece o que hoje denominamos decomposição de Reynolds: o campo de velocidade $\mathbf{U} = U_i \mathbf{e}_i$ de um fluido foi decomposto em $U_i = \langle U_i \rangle + u_i$ (“média” e “flutuação”), e equações para as médias $\langle U_i \rangle$, e para o segundo momento $\langle u_i u_i \rangle / 2$, foram deduzidas a partir de promediações das equações de Navier-Stokes para U_i (a velocidade do fluido no ponto $\mathbf{x} = x_i \mathbf{e}_i$, e no instante t).

Reynolds deu à promediação ‘ $\langle \cdot \rangle$ ’ que ele usou o significado de uma média espacial, num procedimento claramente inspirado pelo trabalho anterior, e pioneiro, de Maxwell sobre a teoria cinética dos gases (Maxwell, 1867).

Logo, percebeu-se que a dedução das equações para os momentos de ordem 1 e 2 (as equações de Reynolds) a partir das equações de Navier-Stokes requeria um conjunto de postulados, *que não aparece explicitamente no artigo de 1895*:

$$\langle u_i \rangle = 0, \quad (1.1)$$

$$\langle \langle U_i \rangle \rangle = 0, \quad (1.2)$$

$$\langle u_i \langle U_j \rangle \rangle = 0, \quad (1.3)$$

$$\left\langle \frac{\partial U_i}{\partial t} \right\rangle = \frac{\partial \langle U_i \rangle}{\partial t}, \quad (1.4)$$

$$\left\langle \frac{\partial U_i}{\partial x_j} \right\rangle = \frac{\partial \langle U_i \rangle}{\partial x_j}. \quad (1.5)$$

Os “postulados” (1.1)–(1.5) aparecem, de uma forma ou de outra, em todos os livros-texto importantes de turbulência. Por exemplo (e a lista não é abrangente): Hinze (1975), Tennekes e Lumley (1972), Monin e Yaglom (1975), Pope (2000).

O significado original da promediação de Reynolds não se manteve: a analogia entre o movimento das moléculas de um gás e o das partículas materiais de um fluido (debaixo da hipótese do contínuo) é essencialmente insustentável (Tennekes e Lumley, 1972, Capítulo 1).

Curiosamente, em nenhum caso os autores acima (ou quaisquer outros que sejam de nosso conhecimento) procuram deduzir matematicamente (1.1)–(1.5), com a possível exceção de Kundu (1990), que o faz para um *ensemble*, supostamente finito, de realizações do escoamento turbulento.

A definição mais comumente encontrada de ‘ $\langle \cdot \rangle$ ’, e que formalizaremos neste texto, é a de uma média probabilística (*ensemble average*), mas mesmo aqui uma definição precisa não pode ser facilmente encontrada na literatura de turbulência.

De qualquer forma, a abordagem estatística de sistemas complexos era promissora, e logo Reynolds recebeu a companhia ilustre de [Einstein \(1905\)](#) (movimento browniano), [Langevin \(1908\)](#) (idem), [Taylor \(1935a,b,c,d\)](#) (turbulência), e, naturalmente, [Kolmogorov \(1941\)](#) (turbulência).

Entretanto, o *significado* de se tomar médias probabilísticas sobre os resultados de um processo determinístico está longe de ser óbvio. Intuitivamente, a motivação é que o processo é suficientemente “complexo” (as velocidades das moléculas de um gás; um escoamento turbulento; o movimento browniano; etc.) para que seja mais fácil lidar apenas com médias e desvios-padrão, ou seja, *estatísticas*, e não com realizações, trajetórias, moléculas, partículas, etc., *individuais*. Ainda assim, do ponto de vista experimental, a tomada de médias sobre realizações é fortemente restrita pela disponibilidade de tais realizações. A saída é supor, debaixo de hipóteses adicionais tais como estacionariedade (mas que por si só não é suficiente), que médias tomadas sobre um número pequeno de realizações (tão pequeno quanto uma única) são capazes de estimar as médias de *ensemble*. Essa é a *hipótese ergódica*.

De todo modo, os procedimentos adotados pelos pioneiros eram necessariamente intuitivos, e tiveram que aguardar, como acontece tantas vezes na história da Física e da Matemática, por formalização posterior. A própria teoria de probabilidade só seria definitivamente formalizada por [Kolmogorov](#) em 1933; o teorema ergódico é devido a [Birkhoff \(1931\)](#), e a [von Neumann \(1932\)](#).

Finalmente, as conexões entre processos estocásticos e sistemas dinâmicos, que em última análise justificam o procedimento informal dos pioneiros, parecem ser ainda mais recentes (ver [Lebowitz e Penrose, 1973](#); [Collet, 2010](#)).

Neste texto, nós procuramos ao mesmo tempo traçar uma parte da história da abordagem estatística de sistemas dinâmicos, desde [Maxwell \(1867\)](#) até a atualidade, e prover um nível intermediário de formalização.

A formalização aqui não é feita no espírito de “arte pela arte”, mas sim no de embasar mais firmemente as análises de caráter probabilístico de sistemas físicos determinísticos, representados matematicamente por sistemas dinâmicos.

Capítulo 2

2014-02-19: Aditividade finita

Como conceitualizamos e formalizamos a propabilidade?

Existem várias abordagens possíveis:

1. Clássica (teórica ou “a priori”):

Consideramos um processo aleatório com n resultados igualmente prováveis, e um evento A que consiste em m desses resultados. A probabilidade desse evento é então definida por

$$P(A) \equiv \frac{m}{n}.$$

Crítica: no termo “igualmente prováveis”, já há a suposição de que nós “sabemos” o que é probabilidade antes de defini-la. Trata-se portanto de um argumento circular. (COMO PODEMOS MELHORAR ESSE TEXTO?)

2. Empírica (“a posteriori” ou frequentista):

Supõe-se que um determinado experimento é repetido n vezes “nas mesmas condições”. Se A é um evento identificável no experimento, a probabilidade de A é definida como o limite da razão entre número m de ocorrências de A e o número de repetições n quando $n \rightarrow \infty$:

$$P(A) \equiv \lim_{n \rightarrow \infty} \frac{m}{n}.$$

3. Subjetiva:

Aceita-se que podemos atribuir a diversos eventos uma “probabilidade” de ocorrência. Por exemplo, eu *acho* que a probabilidade de que eu encontre petróleo no terreno de minha casa é (ou deve ser) 10^{-12} .

4. Axiomática ([Kolmogorov, 1933](#)).

Uma tripla de propabilidade é uma tripla formada por (Ω, \mathcal{F}, P) , sendo Ω um conjunto não vazio, \mathcal{F} um campo sigma (uma σ -álgebra) de subconjuntos de Ω (PRECISAMOS USAR OS TERMOS CORRETOS), e P uma função, com

$$P : \mathcal{F} \rightarrow [0, 1],$$

$$A \in \mathcal{F} \mapsto P(A).$$

Axiomas:

$$P(A) \geq 0, \quad (2.1)$$

$$P(\Omega) = 1, \quad (2.2)$$

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i), \quad \text{se} \quad A_i \cap A_j = \emptyset. \quad (2.3)$$

Os axiomas funcionam quando Ω é finito. Contudo, há conjuntos maiores/infinitos (?) para os quais a noção de probabilidade não faz sentido. Assim, uma σ -álgebra será um subconjunto de 2^Ω com uma certa estrutura, para o qual deverá fazer sentido especificar probabilidades.

Exemplo: Sabendo que $A_1 \cup A_2 \cup A_3 = (A_1 \cup A_2) \cup A_3$, prove por indução que o axioma (2.3) vale para todo n se ele valer para $n = 2$.

Para $n = 2$,

$$A_1 \cup A_2 = \emptyset \Rightarrow P(A_1 \cup A_2) = \sum_{i=1}^2 P(A_i). \quad (2.4)$$

Suponha agora que (2.3) valha para n , e que

$$A_{n+1} \cap \left[\bigcup_{i=1}^n A_i\right] = \emptyset, \quad i = 1, \dots, n.$$

Então,

$$P(A_1 \cup \dots \cup A_n \cup A_{n+1}) = P(B \cup A_{n+1}), \quad (2.5)$$

fazendo-se

$$B = \bigcup_{i=1}^n A_i.$$

A partir de (2.4),

$$P(B \cup A_{n+1}) = P(B) + P(A_{n+1}). \quad (2.6)$$

Por sua vez, como supusemos a validade de (2.3),

$$P(B) = P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i). \quad (2.7)$$

Logo,

$$P(A_1 \cup \dots \cup A_n \cup A_{n+1}) = P(A_{n+1}) + \sum_{i=1}^n P(A_i) = \sum_{i=1}^{n+1} P(A_i). \quad (2.8)$$

Note entretanto que, para que a prova seja válida, precisamos garantir que

$$A_{n+1} \cap \left[\bigcup_{i=1}^n A_i \right] = \emptyset \Rightarrow A_{n+1} \cap A_i = \emptyset, \forall i = 1, \dots, n.$$

Faça $C = A_{n+1}$, e considere a igualdade:

$$C \cap \left[\bigcup_{i=1}^n A_i \right] = \bigcup_{i=1}^n C \cap A_i. \quad (2.9)$$

Se ela for verdadeira, então:

$$\begin{aligned} A_{n+1} \cap \left[\bigcup_{i=1}^n A_i \right] = \emptyset &\Rightarrow \bigcup_{i=1}^n C \cap A_i = \emptyset \\ &\Rightarrow A_{n+1} \cap A_i = \emptyset, \quad \forall i = 1, \dots, n. \end{aligned}$$

Portanto, se (2.9) for verdadeira, a questão está liquidada.

De fato,

$$\begin{aligned} x \in C \cap \left[\bigcup_{i=1}^n A_i \right] &\Rightarrow (x \in C) \text{ e } \left(x \in \bigcup_{i=1}^n A_i \right), \\ &\Rightarrow \exists j \in \{1, \dots, n\} \mid (x \in C) \text{ e } (x \in A_j) \\ &\Rightarrow x \in C \cap A_j \\ &\Rightarrow x \in \bigcup_{i=1}^n C \cap A_i. \end{aligned}$$

Isso significa que

$$C \cap \left[\bigcup_{i=1}^n A_i \right] \subseteq \bigcup_{i=1}^n C \cap A_i.$$

Por outro lado,

$$\begin{aligned} x \in \bigcup_{i=1}^n C \cap A_i &\Rightarrow \exists j \mid x \in C \cap A_j \\ &\Rightarrow x \in C \cap \bigcup_{i=1}^n A_i. \end{aligned}$$

Isso significa que

$$\bigcup_{i=1}^n C \cap A_i \subseteq C \cap \left[\bigcup_{i=1}^n A_i \right].$$

Com isso, (2.9) está provada, e chegamos ao fim (desta prova).

Notas de Aula Professor Paulo Cezar P. de Carvalho - Ailin

Modelos elementares

$\Omega = \{\omega_1, \dots, \omega_n\} \rightarrow$ espaço amostral

$\mathcal{F} = 2^\Omega$ é o conjunto potência e inclui todos os subconjuntos de Ω , e, em particular, inclui $\{\omega_1\}, \{\dots\}, \{\omega_n\}$, os quais são chamados eventos complementares.

$$P(\{\omega_i\}) = P_i \in [0, 1] \mid \sum_{i=1}^n P_i = 1 \quad (2.10)$$

Caso equiprovável: $P_i = 1/n, \forall i \in \{1, \dots, n\}$.

Exemplo: 3 moedas são lançadas. Qual a probabilidade de saírem 2 caras?

Como defino Ω ? Se considerarmos $\Omega = \{0, 1, 2, 3\}$, as probabilidades são $1/8$ para 0 e 3, e $3/8$ para 1 e 2. Para $i = 0, 1, 2, 3$, temos

$$P(i) = \left(\frac{1}{8}\right)^i \left(1 - \frac{1}{8}\right)^{3-i} \binom{3}{i}, \quad (2.11)$$

distribuição binomial herdada do modelo equiprovável.

Exemplo: Escolher um número no intervalo $[0, 1]$ tal que $P([a, b]) = b - a$ para qualquer intervalo $[a, b] \subset [0, 1]$.

$\Omega = [0, 1]$

A primeira tentativa seria atribuir $P(\{a\})$. Se for equiprovável com $P \neq 0$ já estaria em contradição com a aditividade.

Com um conjunto enumerável (infinito) não é possível ter equiprobabilidade nem atribuindo probabilidade nula, porque não conseguiremos que a “soma” das $P(\{\})$ seja 1.

Capítulo 3

2014-02-24: Aditividade infinita

(Ou σ -Aditividade)

Passamos agora para casos em que o espaço amostral Ω deixa de ser um conjunto finito. Um conjunto infinito pode ser enumerável ou não-enumerável. Um conjunto enumerável é um conjunto cujos elementos possam ser colocados em uma relação biunívoca com os naturais. Os racionais são um conjunto enumerável (Cantor). Os números reais no intervalo fechado $[0, 1]$ são um conjunto não-enumerável.

Quando Ω é finito, todos os elementos de 2^Ω são eventos: a todos e a cada um deles pode ser atribuída uma probabilidade, e os axiomas (2.1)–(2.3) se aplicam.

Exemplo 3.1 Antes de seguir para o infinito, considere o exemplo: n lançamentos de uma moeda, cujos resultados individuais podem ser “cara” (0) ou “coroa” (1). Os eventos elementares com os quais podemos construir um espaço amostral são n -uplas do tipo

$$\begin{aligned} &(0, 0, \dots, 0, 0) \\ &(0, 0, \dots, 0, 1) \\ &(0, 0, \dots, 1, 0) \\ &\vdots \\ &(1, 1, \dots, 1, 1). \end{aligned}$$

Existem 2^n casos. Portanto, o espaço amostral mais “simples” que podemos imaginar aqui é

$$\Omega = \{\omega_k = (x_1, \dots, x_n), x_i = 0 \text{ ou } 1, k = 1, \dots, 2^n\}$$

Observe que

$$P(\{\omega_k\}) = \frac{1}{2^n}.$$

Suponha por exemplo que desejemos calcular a probabilidade de que ocorram k caras (e, conseqüentemente, $n - k$ coroas). Um evento deste tipo (exatamente k caras e $n - k$ coroas) pode ocorrer de $n!$ maneiras. No entanto, a posição das k caras é imaterial: todos os $k!$ casos aparecem da mesma forma. Idem para os $(n - k)!$ casos de permuta das posições das coroas. Concluimos que há

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

possibilidades de ocorrência de k caras. A sua probabilidade é

$$\frac{\binom{n}{k}}{2^n} = \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k}.$$

Isso é um caso particular da distribuição binomial. Se 0 tem probabilidade p , e 1 tem probabilidade $1 - p$, a probabilidade de k zeros em n lançamentos é

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Exercício: Mostre que

$$\sum_{k=0}^n P(k) = 1.$$

Prova:

$$\begin{aligned} (p + (1-p))^n &= 1 \\ &= \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^n P(k). \end{aligned}$$

Agora, se $\Omega = \{x_1, x_2, x_3, \dots\}$ for enumerável, precisamos de

$$\sum_{i=1}^n P(x_i) = 1$$

e fica evidente que os $P(x_i)$ não podem ser todos iguais. Entretanto, ainda é possível aproveitar os x_i 's desde que a soma acima funcione.

Exemplo: Em um jogo, dez bolas numeradas de 0 a 9 podem ser sorteadas. Cada jogador sorteia uma bola, mostra o resultado e retorna a bola. Ganha o primeiro jogador que sortear um 7. O jogo poderia durar para sempre?

Nossa opção para construção do espaço amostral é

$$\Omega = \{0, 1, 2, 3, 4, \dots\}$$

0 significa que o 7 *nunca* é sorteado; 1 significa que o 7 foi sorteado na primeira rodada; 2 na segunda; e assim por diante. As probabilidades desses eventos não são iguais:

$$\begin{aligned} P(0) &= ?, \\ P(1) &= \frac{1}{10}, \\ P(2) &= \frac{9}{10} \times \frac{1}{10}, \\ P(3) &= \frac{9}{10} \times \frac{9}{10} \times \frac{1}{10}, \\ &\vdots \\ P(n) &= \left(\frac{9}{10}\right)^{n-1} \times \frac{1}{10} \end{aligned}$$

É elementar verificar que $P(n)$, $n \geq 1$, é uma série geométrica com soma 1. Portanto, o evento “o 7 nunca é sorteado”, indicado por 0, tem probabilidade complementar à $P(1) + P(2) + \dots = 1$, e sua probabilidade é zero.

Finalmente, considere o caso em que desejamos atribuir probabilidades dentro do conjunto não-enumerável $\Omega = [0, 1]$. Note que faz sentido atribuir probabilidade zero a um ponto qualquer:

$$P(X = a) = 0$$

e que é muito razoável atribuir probabilidades a intervalos:

$$P([a, b]) = b - a.$$

O problema é que se A é um evento, o seu complemento \bar{A} também tem que ser, com $P(\bar{A}) = 1 - P(A)$, pela propriedade de aditividade finita (2.3). Portanto, se $(a, b] \in \mathcal{F}$, devemos também ter $\overline{(a, b]} \in \mathcal{F}$, onde \mathcal{F} será a classe dos eventos cujas probabilidades podem ser quantificadas.

Entretanto, o complemento de um único intervalo $(a, b]$ *não* é um único intervalo. Desconfiamos que uma criatura desse tipo, $[0, a] \cup (b, 1]$, precisa ser definida com as mesmas propriedades genéricas de $(a, b]$, de forma que ambos pertençam a \mathcal{F} . O caminho para \mathcal{F} , entretanto, é longo.

Uma extensão do que vimos para a binomial é a distribuição multinomial:

$$1 = (p_1 + \dots + p_n)^n = \sum_{l_1 + \dots + l_k = n} \binom{n}{l_1 \dots l_k} p_1^{l_1} \dots p_k^{l_k},$$

$$\binom{n}{l_1 \dots l_k} = \frac{n!}{l_1! \dots l_k!}.$$

Capítulo 4

2014-02-26: Semi-anéis, anéis, e outros bichos

Definição

Uma classe \mathcal{S} de conjuntos é um *semi-anel* quando:

$$\begin{aligned}\emptyset &\in \mathcal{S}, \\ A, B \in \mathcal{S} &\Rightarrow A \cap B \in \mathcal{S}, \\ A, B \in \mathcal{S} &\Rightarrow A - B = A \cap \overline{B} = \bigsqcup_{i=1}^n E_i,\end{aligned}$$

onde $E_i \in \mathcal{S}$. O símbolo \bigsqcup significa “uniões disjuntas”.

Seja \mathcal{S} a classe formada por *intervalos* do tipo

$$(a, b] \quad \text{ou} \quad \emptyset.$$

1. $\emptyset \in \mathcal{S}$? Sim.
2. A interseção de dois elementos de \mathcal{S} pertence a \mathcal{S} ?

Sim: as possibilidades para interseção de $(a, b]$ com $(c, d]$ são

$$\begin{aligned}\emptyset &\in \mathcal{S}, \\ (c, b] &\in \mathcal{S}, \\ (c, d] &\in \mathcal{S}, \\ (a, d] &\in \mathcal{S}, \\ (a, b] &\in \mathcal{S}.\end{aligned}$$

Talvez seja possível resumir:

$$(a, b] \cap (c, d] = (\max(a, c), \min(b, d)) \quad \text{ou} \quad \emptyset?$$

3. $A - B$ (A, B intervalos) é exprimível como uma união finita disjunta de intervalos?

Sim:

$$\begin{aligned}(a, b] \cap \overline{(c, d]} &= (a, b] \cap [(0, c] \cup (d, 1]] \\ &= \underbrace{(a, b] \cap (0, c]}_{\in \mathcal{S}} \cup \underbrace{(a, b] \cap (d, 1]}_{\in \mathcal{S}} \quad \blacksquare\end{aligned}$$

Portanto, essa classe \mathcal{S} de intervalos é um semi-anel.

Definição: $\mathcal{R} \subset 2^\Omega$ é um *anel* quando:

$$\begin{aligned}\emptyset &\in \mathcal{R}, \\ A, B \in \mathcal{R} &\Rightarrow A \cap B \in \mathcal{R}, \\ A, B \in \mathcal{R} &\Rightarrow A \Delta B \in \mathcal{R}.\end{aligned}$$

Lembre-se:

$$A \Delta B \equiv (A - B) \cup (B - A).$$

Tivemos uma discussão sobre o motivo de se usar a diferença simétrica nessas definições: alguém gostaria de resumir a discussão em L^AT_EX?

Anéis triviais são:

$$\{\emptyset, \Omega\} \quad \text{e} \quad 2^\Omega.$$

É possível mostrar as seguintes propriedades de anéis: se $A, B \in \mathcal{R}$, então:

$$\begin{aligned}A \cap B &\in \mathcal{R}, \\ A \Delta B &\in \mathcal{R}, \\ A \cup B &\in \mathcal{R}, \\ A - B &\in \mathcal{R}, \\ B - A &\in \mathcal{R}, \\ \emptyset &\in \mathcal{R}.\end{aligned}$$

Mas a última não faz parte da *definição* de \mathcal{R} ????

Note também que o complemento ainda não apareceu na jogada.

Teorema

A partir de um semi-anel \mathcal{S} é possível construir um anel \mathcal{R} por meio somente de uniões disjuntas finitas de elementos de \mathcal{S} , ou seja:

$$\mathcal{R} = \left\{ \bigsqcup_{i=1}^n A_i \right\}, \quad A_i \in \mathcal{S}.$$

Definição

Uma *álgebra* ou um *campo* \mathcal{A} é um anel que contém Ω . Segue-se imediatamente que

$$A \in \mathcal{A} \Rightarrow \overline{A} \in \mathcal{A}.$$

Definição Um σ -anel é um anel fechado por uma união enumerável

Comentário: “fechado” significa que uniões enumeráveis de elementos do σ - \mathcal{R} ainda pertencem a ele.

Segue-se imediatamente que o σ - \mathcal{R} também é fechado por interseções enumeráveis.

Definição Uma σ -álgebra ou σ -campo ou campo de Borel é uma álgebra fechada por uniões enumeráveis.

Segue-se imediatamente que uma σ -álgebra também é fechada por interseções enumeráveis.

Definição Dada uma classe $\mathcal{C} \subseteq 2^\Omega$, uma sequência monótona de elementos $E_i \in \mathcal{C}$, $i \in \mathbb{N}$, é definida por

$$E_i \subseteq E_{i+1}$$

ou

$$E_i \supseteq E_{i+1}.$$

Definição Uma classe \mathcal{M} é dita *monótona* quando todas as suas sequências monótonas atendem a:

$$E_i \subseteq E_{i+1} \Rightarrow \bigcup_{i=1}^n E_i \in \mathcal{M},$$

ou

$$E_i \supseteq E_{i+1} \Rightarrow \bigcap_{i=1}^n E_i \in \mathcal{M}.$$

A questão agora é como construir σ -álgebras a partir de álgebras, anéis ou semi-anéis.

Teorema: Para as famílias \mathcal{R} , \mathcal{A} , $\sigma\text{-}\mathcal{R}$, $\sigma\text{-}\mathcal{A}$, \mathcal{M} de anéis, álgebras, sigma-anéis, sigma-álgebras, e classes monótonas \mathcal{M} , interseções arbitrárias produzem famílias de mesmo tipo.

Em resumo: se \mathcal{F}_i , $i \in I$, são classes de algum dos tipos acima, então

$$\bigcap_{i \in I} \mathcal{F}_i$$

também é.

Por exemplo: se

$$\mathcal{F}_i, i \in I$$

são σ -álgebras, então

$$\bigcap_{i \in I} \mathcal{F}_i$$

também é uma σ -álgebra.

4.1 Geração de σ -álgebras

Seja \mathcal{S} um semi-anel que gera o anel \mathcal{R} . A *menor* σ -álgebra contendo \mathcal{S} ou gerada por \mathcal{R} é denominada “ σ -álgebra de Borel” \mathcal{B} .

No caso de $\Omega = [0, 1]$, a σ -álgebra de Borel é a menor σ -álgebra que contém os intervalos $(a, b]$.

Para ela, é possível definir uma medida de probabilidade

$$\begin{aligned} P &: \mathcal{B} \rightarrow [0, 1], \\ A \in \mathcal{B} &\mapsto P(A) \in [0, 1], \end{aligned}$$

de tal forma que

$$\begin{aligned}P(\emptyset) &= 0, \\P(\Omega) &= 1, \\P\left(\bigsqcup_{i=1}^{\infty} A_i\right) &= \sum_{i=1}^{\infty} P(A_i).\end{aligned}$$

Capítulo 5

2014-03-10: O conjunto de Vitali: existe um conjunto não-mensurável em $(0, 1]$

5.1 A q-soma e resultados preliminares

Uma nova operação será usada *ad nauseam* nesta lição. Para $x, y \in (0, 1]$:

$$x \boxplus y \equiv \begin{cases} x + y, & x + y \leq 1, \\ x + y - 1, & x + y > 1. \end{cases} \quad (5.1)$$

Vamos chamar “ \boxplus ” de q-soma, para distingui-la da soma usual em \mathbb{R} .

O conjunto $(0, 1]$ juntamente com a operação \boxplus constitui um *grupo abeliano*, uma vez que valem as seguintes propriedades:

1) Fechamento:

$$x \boxplus y \in (0, 1]. \quad (5.2)$$

De fato \boxplus configura uma função:

$$\begin{aligned} \boxplus : (0, 1] \times (0, 1] &\rightarrow (0, 1], \\ (x, y) &\mapsto z = x \boxplus y. \end{aligned}$$

2) Comutatividade: para $x, y \in (0, 1]$,

$$x \boxplus y = y \boxplus x. \quad (5.3)$$

3) Associatividade:

$$(x \boxplus y) \boxplus z = x \boxplus (y \boxplus z). \quad (5.4)$$

Uma vez verificada a propriedade associativa, podemos dispensar os parênteses, e escrever $x \boxplus y \boxplus z$.

Precisamos de um voluntário ou voluntária para provar a associatividade.

4) Elemento neutro:

$$x \boxplus 1 = 1 \boxplus x = x. \quad (5.5)$$

5) Elemento inverso da soma:

$$\forall x \in (0, 1], \exists y \in (0, 1] : x \boxplus y = 1. \quad (5.6)$$

Vamos usar a notação $(\boxminus x)$ para indicar o inverso da q-soma. É fácil ver que:

$$(\boxminus x) = \begin{cases} 1, & x = 1, \\ 1 - x, & x < 1. \end{cases}$$

O primeiro caso é trivial, pois $1 \boxplus 1 = 1$. No segundo caso,

$$x \boxplus y = x + y = x + (1 - x) = 1, \text{ (pois } x + y \leq 1) \quad \forall x \in (0, 1].$$

A existência do elemento inverso da soma permite que nós definamos a operação “q-diferença”, \boxminus :

Definição: $x \boxminus y \equiv x \boxplus (\boxminus y)$.

Lema 5.1 Para $x, z \in (0, 1]$:

$$\exists y \in (0, 1] : z = x \boxplus y \Leftrightarrow \exists r \in (-1, 1) : z = x + r. \quad (5.7)$$

Favor verificar se os limites para o intervalo de r estão corretos.

Prova: Se $x = 1$,

$$y = z \boxminus x = z \boxminus 1 = z \in (0, 1] \quad \Leftrightarrow \quad r = y - 1 \in (-1, 1).$$

Se $x < 1$ e $z + 1 - x \leq 1$,

$$y = z \boxminus x = z + 1 - x \quad \Leftrightarrow \quad r = y - 1 \in (-1, 1).$$

Se $x < 1$ e $z + 1 - x > 1$,

$$y = z \boxminus x = z - x \quad \Leftrightarrow \quad r = y \in (-1, 1) \blacksquare$$

Agora, para *qualquer* subconjunto E de $(0, 1]$, defina um novo conjunto $E(x)$

$$E(x) \equiv \{x \boxplus y, y \in E\}. \quad (5.8)$$

$E(x)$ é a *translação* (de uma distância x) do conjunto E .

Dúvida: $E \subseteq (0, 1]$ ou $E \in \mathcal{B}$?

Dúvida: “translação *em* x ” me soa estranho, pois todo $y \in E$ é transladado *de* x .

Suponha que exista uma coisa tal como uma medida “natural” em $(0, 1]$. Para o intervalo $(a, b]$ essa medida é

$$|(a, b]| \equiv b - a. \quad (5.9)$$

Vamos supor, sem entrar em muitos detalhes, que para todo $B \in \mathcal{B}$, existe uma $|B|$ compatível com (5.9), *i.e.*, redutível a (5.9) se B for um intervalo.

A notação $E(x)$ é a usada por Taylor (1973).

Agora, se $E \in \mathcal{B}$, então,

$$|E(x)| = |E|. \quad (5.10)$$

Agora, como sempre, \mathbb{Q} é o conjunto dos racionais. Esse conjunto é enumerável. Consideremos os racionais contidos em $(0, 1]$. Esse segundo conjunto é

$$Q \equiv \mathbb{Q} \cap (0, 1]. \quad (5.11)$$

Valeria a pena provar que

Q é enumerável.

Com essas definições à mão, estudemos então as propriedades dos conjuntos do tipo $Q(x)$.

1.

$$x \in (0, 1] \Rightarrow x \in Q(x). \quad (5.12)$$

Será verdade? Como provar?

$$Q(x) = \{x \boxplus y, y \in Q\}$$

Se 0 pertencesse a Q , (5.12) seria trivial. Mas é quase, porque $1 \in Q$. Faça $y = 1$ acima; então,

$$x = x + 1 - 1 = x \boxplus 1 \in Q(x) \blacksquare$$

2.

$$x \in Q \Rightarrow Q(x) = Q. \quad (5.13)$$

De fato: para começar, $x, y \in Q \Rightarrow x \boxplus y \in Q$. De fato, $x \boxplus y \in (0, 1]$, e tanto $x + y$ quanto $x + y - 1$, conforme for o caso, são números racionais em $(0, 1]$. Isso basta!, pois, *neste caso*, $Q(x) = \{x \boxplus y, y \in Q\} = Q \blacksquare$

3.

$$x_1 - x_2 \notin \mathbb{Q} \Rightarrow Q(x_1) \cap Q(x_2) = \emptyset, \quad (5.14)$$

$$x_1 - x_2 \in \mathbb{Q} \Rightarrow Q(x_1) = Q(x_2). \quad (5.15)$$

Dessa forma, os conjuntos $\{Q(x), x \in (0, 1]\}$ (que constituem uma *classe*, ou *família*) particionam o intervalo $(0, 1]$ em subconjuntos disjuntos cuja união é o próprio $(0, 1]$. Há bastante material aqui. Antes das deduções, vamos escrever formalmente essa última observação:

$$\bigcup_{x \in (0, 1]} Q(x) = (0, 1].$$

Note também que os índices na expressão acima são demasiados, devido a (5.15). Queremos chegar a uma afirmação mais econômica:

$$\bigsqcup_{x \in T} Q(x) = (0, 1]$$

(note a disjunção). Na sequência, precisamos provar (5.14)–(5.15) e prosseguir na obtenção do conjunto T . Esse último se revelará um conjunto interessante, e na verdade o ponto final desta lição: T se revelará um conjunto *não mensurável*.

Para provar (5.14): Vamos tentar *reductio ad absurdum*. Seja $z \in Q(x_1)$ e $z \in Q(x_2)$: nesse caso, a interseção $Q(x_1) \cap Q(x_2)$ não seria o conjunto vazio. Porém, debaixo dessa hipótese:

$$\begin{aligned} z &= x_1 \boxplus y, \quad y \in Q, \\ z &= x_2 \boxplus y, \quad y \in Q. \end{aligned}$$

Agora,

$$\begin{aligned} (x_1 \boxplus y) - (x_2 \boxplus y) &= x_1 - x_2, \text{ ou} \\ &= x_1 - x_2 - 1 \text{ ou} \\ &= x_1 - x_2 + 1. \end{aligned}$$

Portanto, subtraindo (dessa forma) as duas expressões acima,

$$(-1 \text{ ou } 0 \text{ ou } +1) = x_1 - x_2.$$

mas $(-1, 0, +1)$ são racionais, o que contraria a hipótese original sobre $x_1 - x_2$ ■

Para provar (5.15): Volte acima e escreva a expressão geral:

$$(x_1 \boxplus y) - (x_2 \boxplus y) = x_1 - x_2 + s,$$

onde, como vimos, ou $s = 0$ ou $s = -1$ ou $s = +1$. Reescreva:

$$x_1 \boxplus y = x_2 \boxplus y + (x_1 - x_2) + s.$$

Note agora que é *sempre* possível escrever $x_2 \boxplus y = x_2 + r$ (veja (5.7)), onde agora r é racional, e não necessariamente está em Q . Substitua:

$$x_1 \boxplus y = x_2 + r + s + (x_1 - x_2).$$

O lado esquerdo é um número em Q . Portanto, o lado direito é um número em Q :

$$x_2 + p \in Q, \tag{5.16}$$

onde $p = (r + s + (x_1 - x_2))$ e portanto $p \in \mathbb{Q}$ (mas não necessariamente $p \in Q$). Agora, utilizando (5.7), vemos que é possível escrever o lado direito como $x_2 \boxplus z$, onde $z \in Q$. Disso se segue que

$$\{x_1 \boxplus y, \quad y \in Q\} = \{x_2 \boxplus z, \quad z \in Q\}$$

e portanto $Q(x_1) = Q(x_2)$ ■

Relembrando:

$$x \in Q \Rightarrow Q(x) = Q$$

significa, por exemplo:

$$Q(1/2) = Q(1/3) = Q(1).$$

$$x_1 - x_2 \in \mathbb{I} \Rightarrow Q(x_1) \cap Q(x_2) = \emptyset$$

significa, por exemplo:

$$Q(\sqrt{3}/2) \cap Q(3/2) = \emptyset.$$

$$x_1 - x_2 \in \mathbb{Q} \Rightarrow Q(x_1) = Q(x_2)$$

significa, por exemplo:

$$Q(\sqrt{2}/1000 + 1/8) = Q(\sqrt{2}/1000 + 3/8)$$

Ou seja: as criaturas estão ficando estranhas.

Para fechar esta parte bastante cansativa para a mente:

$$\bigcup_{x \in (0,1]} Q(x) = (0, 1]$$

pois $x \in Q(x)$.

5.2 O axioma da seleção e um conjunto estranho

Seja \mathcal{C} a família dos conjuntos $Q(x)$, $x \in (0, 1]$. Escolha um único ponto em $(0, 1]$ pertencente a cada $Q(x)$. O conjunto dos pontos assim escolhidos será chamado T , e temos que $T \subset (0, 1]$. T é um *conjunto de Vitali*. O que fizemos significa que, *por definição*, não há dois pontos em T pertencentes ao mesmo $Q(x)$. Segue-se que

$$\bigsqcup_{t \in T} Q(t) = (0, 1].$$

Estudemos as propriedades dos conjuntos $T(r_i), r_i \in Q$.

1.

$$\bigcup_{i=1}^{\infty} T(r_i) = (0, 1]. \quad (5.17)$$

Só precisamos provar que, se $x \in (0, 1]$, $\exists i \in \mathbb{N} : x \in T(r_i)$. Mas, de (5.12), segue-se que $x \in (0, 1] \Rightarrow x \in Q(x)$. Agora, portanto, se $x \in Q(x)$, escolha $t \in Q(x)$, onde t é o representante de $Q(x)$: $t \in T$.

Veja:

$$Q(x) = x \boxplus y, \quad y \in Q, \\ t \in Q(x) \Rightarrow t = x \boxplus y \text{ para algum } y \in Q.$$

Logo, de (5.7), existe algum $q \in Q$ tal que $t = x + q$; pelo mesmo motivo agora deve existir algum $r \in Q$ tal que $x = t + r$. Mas Q é enumerável (não provamos!), portanto existe um índice i tal que $x = t + r_i$, $i \in \mathbb{N}$, e $Q = \bigsqcup_{i=1}^{\infty} r_i$. Logo, $x \in T(r_i)$. A união dos $T(r_i)$ gera o $(0, 1]$.

2. Os $T(r_i)$ são disjuntos. Lembre-se de que T contém um único representante de cada $Q(x)$. Se houvesse $r_i \neq r_j$ com $y \in [T(r_i) \cap T(r_j)]$, então teríamos

$$\begin{aligned} y &= t_i \boxplus r_i & \text{e} \\ y &= t_j \boxplus r_j. \end{aligned}$$

“Subtraia”:

$$0 = (t_i - t_j) + q$$

onde q é racional (novamente, nós usamos (5.7)). Logo, $t_i - t_j$ é racional, donde $Q(t_i) = Q(t_j)$ (devido a (5.15)). Mas isso não é possível, porque $Q(t_i)$ possui um único representante em T . Portanto, a disjunção dos $T(r_i)$'s transforma (5.17) em

$$\bigsqcup_{i=1}^{\infty} T(r_i) = (0, 1]. \quad (5.18)$$

Finalmente: se T fosse mensurável, haveria $|T| = |T(r_i)|$ para todo i . Agora, pela sigma-aditividade:

$$\sum_{i=1}^{\infty} |T(r_i)| = |(0, 1]| = 1.$$

Isso, nós também já vimos em outra lição, é impossível. T não pode ser mensurável ■

Capítulo 6

2014-03-12, 2014-03-17: Teoria de probabilidade e incursões na física e em geociências

6.1 Variável aleatória

Este início de seção foi escrito por mim, anteriormente. Entretanto, ele se parece bastante com a exposição da Ailin em 2014-03-17. Seria bom portanto unificar as duas aqui. Aceito sugestões.

Em uma visão moderna de probabilidades e estatística, inventada por Kolmogorov, nós precisamos de:

1. um conjunto Ω — responsável por “sorteios”, denominado *Espaço Amostral*;
2. Uma função

$$\begin{aligned} X : \Omega &\rightarrow \mathbb{R} \\ \omega \in \Omega &\mapsto x = X(\omega) \end{aligned}$$

A função X é denominada “variável aleatória”. A função X tem que ser *mensurável*, e nós vamos gastar algum tempo com isso ainda mais à frente.

3. Uma função

$$\begin{aligned} P : \mathcal{F} &\rightarrow [0, 1] \\ A \in \mathcal{F} &\mapsto P(A) \end{aligned}$$

O conjunto \mathcal{F} é uma *classe* (um conjunto de conjuntos). Como vimos no capítulo 3, \mathcal{F} é uma σ -álgebra. Cada *elemento* de \mathcal{F} é um sub-conjunto mensurável $A \in \Omega$, ao qual nós associamos um número $P(A)$. Os elementos de \mathcal{F} gozam de algumas propriedades

importantes em teoria de probabilidades:

$$A \in \mathcal{F} \Rightarrow \bar{A} \in \mathcal{F} \quad (6.1)$$

$$A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F} \quad (6.2)$$

$$A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F} \quad (6.3)$$

$$\emptyset \in \mathcal{F} \quad (6.4)$$

$$\Omega \in \mathcal{F} \quad (6.5)$$

Nós podemos reconhecer nestas propriedades, imediatamente, que os sub-conjuntos $A \in \Omega$ são *eventos*. A função P associa a cada evento uma *probabilidade*. A função P também deve obedecer a algumas regras. Se A_1, \dots, A_n são conjuntos *disjuntos*, $A_i \cap A_j = \emptyset$, $i \neq j$, então

$$P(A_i) \geq 0 \quad (6.6)$$

$$P(\Omega) = 1 \quad (6.7)$$

$$P(A_i \cup A_j) = P(A_i) + P(A_j) \quad (6.8)$$

Esta última é um caso particular da σ -aditividade:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \quad (6.9)$$

É muito importante enfatizar que o espaço amostral é um conceito *abstrato*: ele ajuda a conceituar probabilidades e também ajuda a demonstrar alguns resultados, mas na prática tudo o que vemos no mundo “real”, ou seja, os únicos objetos que podemos manipular em aplicações de teoria de probabilidade, são as *realizações* da variável aleatória: $x = X(\omega)$.

Para avaliar a probabilidade de ocorrência de certos níveis de uma variável aleatória nós usamos eventos do tipo $X(\omega) \leq x$, que pode ser entendido com o auxílio da figura 6.1.

Diferentes maneiras de escrever o evento A são:

$$A = \{\omega \mid X(\omega) \leq x\},$$

$$A = \{X(\omega) \leq x\},$$

$$A = \{X \leq x\}.$$

Note que na última forma desapareceu qualquer referência explícita ao espaço amostral.

6.2 A conveniência de definir funções de Ω em \mathbb{R} .

Em princípio, para cada pergunta que nós podemos formular sobre eventos e probabilidades, é possível construir um espaço amostral Ω “sob medida” para respondê-la. Junto com esse espaço, é possível, também em princípio, definir uma medida de probabilidade adequada. Nesse caso, tudo o que é necessário, sempre, é a formulação de uma tripla de probabilidade (Ω, \mathcal{F}, P) que nos permita medir a probabilidade dos eventos A de nosso interesse.

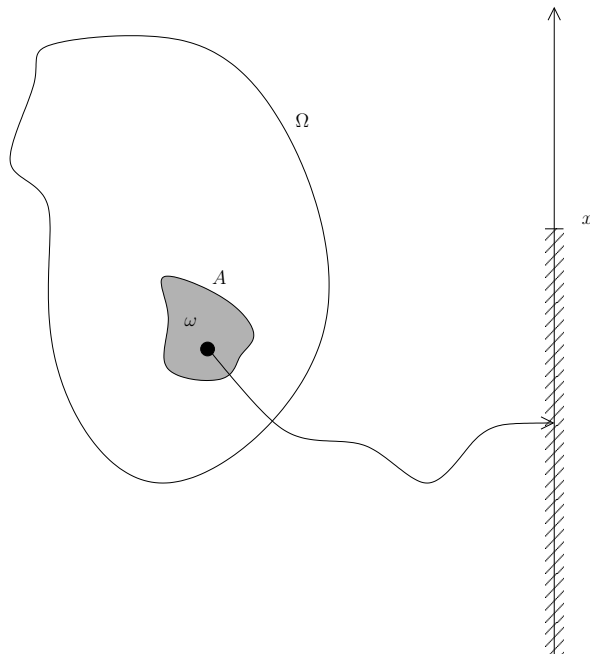


Figura 6.1: Representação gráfica do evento $X(\omega) \leq x$

Entretanto, isso não é *prático*. O motivo é que, *sempre*, as perguntas que podemos formular a partir de uma dada tripla (Ω, \mathcal{F}, P) inicial, podem ser expressas em termos de *funções* cujo domínio é Ω . Rapidamente, portanto, nós nos deparamos com a conveniência, e quase que com a imposição, da definição de funções de Ω em \mathbb{R} . Tais funções são denominadas *variáveis aleatórias*.

Considere, por exemplo, o lançamento de n moedas do exemplo 3.1. Se “cara” significa $x_i = 0$ e “coroa” significa $x_i = 1$, a pergunta: “qual é o número de coroas obtidas em n lançamentos?” pode ser respondida com o auxílio da função

$$X : \Omega \Rightarrow \mathbb{R},$$

$$\omega = (x_1, \dots, x_n) \mapsto \sum_{i=1}^n x_i.$$

Neste exemplo, X é uma função. A imagem de X é $\text{Im}X = \{0, 1, \dots, n\} \subset \mathbb{R}$. X não é sobrejetiva.

Perguntas similares podem ser respondidas com outras funções, sem que seja necessário redefinir uma nova tripla (Ω, \mathcal{F}, P) para cada pergunta! Por exemplo, se desejarmos calcular probabilidades associadas ao número de caras obtidas em n lançamentos, podemos usar:

$$X : \Omega \Rightarrow \mathbb{R},$$

$$\omega = (x_1, \dots, x_n) \mapsto n - \sum_{i=1}^n x_i.$$

Ou ainda poderíamos estar interessados no número médio de coroas em n lançamentos (ou ainda no limite dessa quantidade quando $n \rightarrow \infty$), quando então poderíamos usar a

função

$$X : \Omega \rightarrow \mathbb{R},$$

$$\omega = (x_1, \dots, x_n) \mapsto \frac{1}{n} \left[\sum_{i=1}^n x_i \right].$$

Esses breves exemplos sugerem que é *econômico* e *útil* trabalhar com funções definidas em Ω , em vez de reformular, do zero, um novo Ω para cada pergunta relevante que possamos ter sobre eventos relacionados com o Ω original. Sob esse ponto de vista, variáveis aleatórias não são, em absoluto, uma necessidade. Elas são apenas uma forma cômoda de responder perguntas sobre probabilidades.

A conveniência de trabalhar com variáveis aleatórias, entretanto, cria um novo conjunto de “problemas” técnicos que agora precisam ser resolvidos. Grosso modo, a questão é a seguinte:

Como é que o contradomínio \mathbb{R} “percebe” a tripla de probabilidade (Ω, \mathcal{F}, P) ?

Sucede que o que faz sentido é calcular probabilidades dos subconjuntos de \mathbb{R} pertencentes à classe \mathcal{B} , onde \mathcal{B} é a σ -álgebra de Borel.

Lembremo-nos de que \mathcal{B} é a menor σ -álgebra que contém os intervalos $(a, b]$, $a, b \in \mathbb{R}$, os intervalos $(-\infty, b]$, $(a, +\infty)$, e também os intervalos $[a, b]$, (a, b) , $[a, b)$ e suas uniões e interseções, e todos os conjuntos abertos e fechados.

Não sei o que os conjuntos abertos e fechados estão fazendo aqui. Seria bom explicar sua relação com os intervalos.

Precisamos, portanto, “ligar”, de alguma forma, \mathcal{B} a \mathcal{F} , onde esta última é a σ -álgebra da tripla de probabilidade “original”. No fim, o que vai funcionar é o seguinte: para qualquer $B \in \mathcal{B}$, vamos definir a probabilidade de conjuntos mensuráveis em \mathcal{B} como

$$\mathcal{P}(B) \equiv P(X^{-1}(B)). \quad (6.10)$$

É claro que é preciso “garantir” que $X^{-1}(B) \in \mathcal{F}$. Pode-se mostrar que uma condição suficiente para isso é que

$$X^{-1}((-\infty, b]) \in \mathcal{F}. \quad (6.11)$$

Seria bom enunciar o resultado na forma de um teorema, mesmo que sem demonstração. Dizemos que, se X atende (6.11), X é uma função *mensurável*. Portanto,

Definição: Dada uma tripla de probabilidade (Ω, \mathcal{F}, P) , uma *variável aleatória* é uma função

$$X : \Omega \rightarrow \mathbb{R},$$

$$\omega \mapsto x = X(\omega),$$

tal que

$$\forall b \in \mathbb{R}, X^{-1}((-\infty, b]) \in \mathcal{F}.$$

É interessante observar que definição de mensurabilidade de uma função é feita em termos da relação (não necessariamente uma função) *inversa* X^{-1} . O motivo é técnico: uniões, interseções e complementos *arbitrários* de conjuntos $B \in \mathcal{B}$ *permanecem em* \mathcal{B} . Da mesma forma, é preciso garantir que as pré-imagens dessas uniões, interseções e complementos *permaneçam em* \mathcal{F} . Dadas as propriedades de fechamento de qualquer σ -álgebra sob essas operações, essa garantia é dada pelas propriedades sempre válidas (dada uma função X qualquer, não necessariamente mensurável — isso é outra parte de nossos requisitos!):

$$X^{-1}(A \cup B) = X^{-1}(A) \cup X^{-1}(B), \quad (6.12)$$

$$X^{-1}(A \cap B) = X^{-1}(A) \cap X^{-1}(B), \quad (6.13)$$

$$X^{-1}(\overline{A}) = \overline{X^{-1}(A)}. \quad (6.14)$$

2014-03-28T10:07:30 — aceito agradecidamente verificações se a prova abaixo está certa Por exemplo:

$$\begin{aligned} X^{-1}(A \cap B) &= \{\omega : X(\omega) \in A \cap B\} \\ &= \{\omega : [X(\omega) \in A] \wedge [X(\omega) \in B]\} \\ &= \{\omega : X(\omega) \in A\} \cap \{\omega : X(\omega) \in B\} \\ &= X^{-1}(A) \cap X^{-1}(B) \blacksquare \end{aligned}$$

As duas primeiras relações podem ser estendidas para uniões e interseções arbitrárias (não necessariamente enumeráveis, embora isso baste para nossas sigma-álgebras):

$$\begin{aligned} X^{-1}\left(\bigcup_{\lambda \in \Lambda} A_\lambda\right) &= \bigcup_{\lambda \in \Lambda} X^{-1}(A_\lambda), \\ X^{-1}\left(\bigcap_{\lambda \in \Lambda} A_\lambda\right) &= \bigcap_{\lambda \in \Lambda} X^{-1}(A_\lambda). \end{aligned}$$

A necessidade de definir mensurabilidade usando X^{-1} decorre do fato de que *é falso que*:

$$\begin{aligned} X(A \cap B) &= X(A) \cap X(B), \\ X(\overline{A}) &= \overline{X(A)}. \end{aligned}$$

Para isso, bastam dois contra-exemplos.

Exemplo 6.1 Sejam $A = (-\infty, 0)$ e $B = [0, +\infty)$. Então, $\overline{A} = B$ e $\overline{B} = A$. Dada a função

$$\begin{aligned} f : \mathbb{R} &\rightarrow \mathbb{R}, \\ x &\mapsto 1, \end{aligned}$$

Temos:

$$A \cap B = \emptyset, \quad f(A) = f(B) = \{1\},$$

e

$$f(A) \cap f(B) = \{1\} \neq \emptyset = f(A \cap B).$$

Exemplo 6.2 Sejam $A = (-\infty, 0)$ e $B = [0, +\infty)$. Então, $\overline{A} = B$ e $\overline{B} = A$. Dada a função

$$\begin{aligned} f : \mathbb{R} &\rightarrow \mathbb{R}, \\ x &\mapsto x^2, \end{aligned}$$

Temos:

$$f(\overline{A}) = f(B) = [0, \infty) = B \neq [0, -\infty) = \overline{f(A)}.$$

Uma função mensurável em Ω tem o papel de transferir a estrutura (Ω, \mathcal{F}, P) para uma tripla equivalente $(\mathbb{R}, \mathcal{B}, \mathcal{P})$. A probabilidade de sub-conjuntos de \mathbb{R} que “fazem sentido” está bem definida por (6.10) (na verdade, nos sub-conjuntos que pertencem a \mathcal{B} (a sigma-álgebra de Borel) e também possivelmente mais alguns, que pertencem a \mathcal{L} (a sigma-álgebra de Lebesgue) — mas a diferença entre ambos consiste em conjuntos de medida zero). Seguem-se

$$\begin{aligned} \mathcal{P}(\mathbb{R}) &= 1, \\ 0 &\leq \mathcal{P}(B) \leq 1, \\ \mathcal{P}\left[\bigsqcup_{i=1}^{\infty} B_i\right] &= \sum_{i=1}^{\infty} \mathcal{P}(B_i) \quad \sigma\text{-aditividade} \end{aligned}$$

Cabe observar que nem toda função de Ω em \mathbb{R} é uma variável aleatória. Considere por exemplo a função

$$\begin{aligned} \mathbb{1}_{\overline{T}} : (0, 1] &\rightarrow \mathbb{R}, \\ \omega &\mapsto \begin{cases} 1, & \omega \in \overline{T} \cap (0, 1], \\ 0, & \omega \in T. \end{cases} \end{aligned}$$

onde T é o conjunto de Vitali definido na seção 5.2. Sabemos que T é não-mensurável: $T \notin \mathcal{F}$. Segue-se que a função $\mathbb{1}_{\overline{T}}$ não é mensurável:

$$\begin{aligned} \mathbb{1}_{\overline{T}}^{-1}\left((-\infty, 1/2]\right) &= \left\{\omega \in (0, 1] : \mathbb{1}_{\overline{T}}(\omega) \in (-\infty, 1/2]\right\} \\ &= \left\{\omega : \mathbb{1}_{\overline{T}}(\omega) = 0\right\} \\ &= T \notin \mathcal{F} \blacksquare \end{aligned}$$

6.3 2014-03-19: Propriedades de variáveis aleatórias

Exemplos de variáveis aleatórias construídas a partir de outras:

1. Se $A \in \mathcal{F}$,

$$\begin{aligned} \mathbb{1}_A : \Omega &\rightarrow \mathbb{R}, \\ \omega &\mapsto \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A. \end{cases} \end{aligned}$$

é uma variável aleatória em (Ω, \mathcal{F}, P) .

2. Se X, Y são duas variáveis aleatórias em (Ω, \mathcal{F}, P) , então

- a) $X + Y$ é uma variável aleatória;
- b) cX é uma variável aleatória;
- c) $X + c$ é uma variável aleatória;
- d) X^2 é uma variável aleatória;
- e) XY é uma variável aleatória.

3. Se $(X_1, X_2, \dots, X_n, \dots)$ é uma sequência de variáveis aleatórias e $\lim_{x \rightarrow \infty} X_n(\omega)$ existe para cada ω , então:

$$X : \Omega \rightarrow \mathbb{R},$$

$$\omega \mapsto \lim_{n \rightarrow \infty} X_n(\omega)$$

satisfaz

$$X^{-1}\left((-\infty, b]\right) \in \mathcal{F}.$$

A prova é complicada! Ela se baseia no seguinte fato (Rosenthal **2014-03-25T09:44:36** — **Precisamos da referência completa**):

$$X^{-1}\left((-\infty, b]\right) = \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} X_k^{-1}\left(-\infty, x + \frac{1}{m}\right]. \quad (6.15)$$

4. $\max(X, Y)$, $\min(X, Y)$ e $|X|$ são variáveis aleatórias. Notando que

$$\max(X, Y) = \frac{1}{2} [X + Y + |X - Y|],$$

$$\min(X, Y) = (X + Y) - \max(X, Y),$$

basta provar para o $|X|$.

5. $\sup_n X_n$, $\inf_n X_n$ são variáveis aleatórias, desde que bem definidos. Vale lembrar que

$$\sup_{n \in \mathbb{N}} r_n = \begin{cases} \min \{r : r_n \leq r, \forall n \in \mathbb{N}\}, \\ +\infty, \end{cases} \quad \text{se a sequência não for limitada.}$$

6. $\limsup_n X_n$ e $\liminf_n X_n$ são variáveis aleatórias.

Forme:

$$Y_n \equiv \sup_{N \geq n} X_N. \quad \textbf{Confirmar índices!!!}$$

Note que $Y_n(\omega)$ é uma função monótona (não-crescente) de n , para cada ω . Portanto,

$$\exists \lim_{n \rightarrow \infty} Y_n \equiv \limsup_{n \rightarrow \infty} X_n.$$

Forme:

$$Z_n \equiv \inf_{N \geq n} X_N. \quad \textbf{Confirmar índices!!!}$$

Note que $Z_n(\omega)$ é uma função monótona (não-decrescente) de n , para cada ω . Portanto,

$$\exists \lim_{n \rightarrow \infty} Z_n \equiv \liminf_{n \rightarrow \infty} X_n.$$

7. Seja $X : \Omega \rightarrow \mathbb{R}$ uma variável aleatória, tal que os abertos de Ω estão em \mathcal{F} , ou seja, \mathcal{F} é a σ -álgebra de Borel de Ω ou seu completamento. Então:

a) Se $f : \mathbb{R} \rightarrow \mathbb{R}$ for contínua, $f \circ X : \Omega \rightarrow \mathbb{R}$ é uma variável aleatória

8. Se:

$$\begin{aligned} X : (\Omega, \mathcal{F}, P) &\rightarrow (\mathbb{R}, \mathcal{B}, \mathcal{P}), \\ \omega &\mapsto x = X(\omega), \end{aligned}$$

e

$$\begin{aligned} f : (\mathbb{R}, \mathcal{B}, |\cdot|) &\rightarrow (\mathbb{R}, \mathcal{B}, |\cdot|), \\ x &\mapsto y = f(x) \end{aligned}$$

(ou seja: f é mensurável), então $f \circ X$ é uma variável aleatória.

Finalmente, temos uma anotação um pouco solta que precisa ser esclarecida: $B \in \mathcal{B}$, $|B| = 0$, $D \subset B$, D não necessariamente em \mathcal{B} , $D \in \mathcal{L}$, a qual é a σ -álgebra de Lebesgue, que é o completamento, por $|\cdot|$, da σ -álgebra de Borel.

Como a prova de 2014-03-24 foi muito longa, não vou colocá-la dentro da lista acima, mas depois.

Um resultado auxiliar

Definição: Se a sequência $X_k(\omega)$ converge para todo ω , então, pela definição de limite, $X_k(\omega) \rightarrow X(\omega)$ se, para todo $\epsilon = 1/m$, $\exists K(m, \omega)$ tal que:

$$|X_k(\omega) - X(\omega)| < \frac{1}{m}, \quad \forall k \geq K(m, \omega). \quad (6.16)$$

Ante-sala da prova

Vamos agora provar (6.15). Comece observando que

$$\begin{aligned} \omega \in \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} X_k^{-1} \left(-\infty, b + \frac{1}{m} \right] &\Leftrightarrow \forall m \in \mathbb{N}, \\ \omega &\in \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} X_k^{-1} \left(-\infty, b + \frac{1}{m} \right], \\ &\Leftrightarrow \forall m \in \mathbb{N}, \exists N(m, \omega) : \\ \omega &\in \bigcap_{k=N(m, \omega)}^{\infty} X_k^{-1} \left(-\infty, b + \frac{1}{m} \right], \\ &\Leftrightarrow \forall m \in \mathbb{N}, \exists N(m, \omega) : \\ X_k(\omega) &\leq b + \frac{1}{m}, \quad \forall k \geq N(m, \omega). \end{aligned}$$

É bom parar um pouco e observar que tudo isso aí em cima *ainda é o lado direito* de (6.15)! Em outras palavras,

$$\bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} X_k^{-1} \left(-\infty, b + \frac{1}{m} \right] = \left\{ \omega : \forall m \in \mathbb{N}, \exists N(m, \omega) : X_k(\omega) \leq b + \frac{1}{m}, \forall k \geq N(m, \omega) \right\}.$$

Eu gostaria de escrever essa última afirmativa com único “tal que” (:), em vez de dois.

Novamente, (6.15) é da forma

$$D = E$$

onde D e E são conjuntos. A prova usual da igualdade de dois conjuntos segue a lógica:

$$\begin{aligned} (\omega \in D \Rightarrow \omega \in E) &\Rightarrow D \subseteq E, \\ (\omega \in E \Rightarrow \omega \in D) &\Rightarrow E \subseteq D, \end{aligned}$$

e portanto $D = E$.

Prova

Somente agora nós vamos realmente provar a primeira linha de implicaturas acima:

$$\omega \in X^{-1}(-\infty, b] \Leftrightarrow X(\omega) \leq b.$$

Usando a equação (6.16), $\forall m \in \mathbb{N}, \exists K(m, \omega)$:

$$|X_k(\omega) - X(\omega)| < \frac{1}{m}, \forall k \geq K(m, \omega).$$

Em particular,

$$X_k(\omega) < \frac{1}{m} + X(\omega) \leq \frac{1}{m} + b,$$

donde

$$X^{-1}(-\infty, b] \subseteq \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} X_k^{-1} \left(-\infty, b + \frac{1}{m} \right].$$

OK, mas até aqui eu ainda acho que devíamos ter usado $N(m, \omega) = K(m, \omega)$ (um único símbolo) na prova.

A “segunda linha” que precisa ser provada, ainda, é:

$$\omega \in \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} X_k^{-1} \left(-\infty, b + \frac{1}{m} \right] \Rightarrow \omega \in X^{-1}(-\infty, b]. \quad (6.17)$$

Já sabemos que o lado esquerdo acima significa que, $\forall m \in \mathbb{N}, \exists K(m, \omega)$:

$$X_k(m) \leq b + \frac{1}{m}, \forall k \geq K(m, \omega). \quad (6.18)$$

Por outro lado, suponha, por *reductio ad absurdum*, que $X(\omega) > b$, e faça $d = X(\omega) - b$. Mas $X(\omega)$ é o limite de uma sequência, e portanto a partir de algum $k \geq N(\epsilon, \omega)$ (**preciso de ajuda para arrumar $N(m, \omega)$ e $K(m, \omega)$**) devemos ter

$$X(\omega) - \epsilon < X_k(\omega) < X(\omega) + \epsilon.$$

Escolha

$$\epsilon = \frac{1}{m} < \frac{d}{2} \Leftrightarrow \frac{2}{m} < d;$$

então,

$$\begin{aligned} X(\omega) - \epsilon &< X_k(\omega), \\ X(\omega) - b + \left(b - \frac{1}{m}\right) &< X_k(\omega), \\ d + \left(b - \frac{1}{m}\right) &< X_k(\omega), \\ \frac{2}{m} + \left(b - \frac{1}{m}\right) &< d + \left(b - \frac{1}{m}\right) < X_k(\omega), \\ b + \frac{1}{m} &< X_k(\omega). \end{aligned} \tag{6.19}$$

Mas (6.18) e (6.19) são contraditórios (um absurdo): a hipótese $X(\omega) > b$ levou a um absurdo, e portanto temos, como desejávamos (6.17) ■

6.4 2014-03-28T09:49:31 revisão, e a FDA

Tínhamos

$$\begin{aligned} X : (\Omega, \mathcal{F}) &\rightarrow (\mathbb{R}, \mathcal{B}), \\ \omega &\mapsto X(\omega), \end{aligned}$$

de tal forma que $X^{-1}(-\infty, b] \in \Omega$, o que equivale a $X^{-1}(B) \in \mathcal{F}$, $\forall B \in \mathcal{B}$. Podemos agora definir:

$$\begin{aligned} \mathcal{P}(B) &= P\left(\underbrace{X^{-1}(B)}_{\in \mathcal{F}}\right) \\ \mathcal{P}(B) &\geq 0, \\ \mathcal{P}(\emptyset) &= 0, \\ \mathcal{P}(\mathbb{R}) &= P(X^{-1}(\mathbb{R})) = P(\Omega) = 1, \\ \mathcal{P}\left(\bigsqcup_{i=1}^{\infty} B_i\right) &= P\left(X^{-1}\left(\bigsqcup_{i=1}^{\infty} B_i\right)\right) \\ &= P\left(\bigsqcup_{i=1}^{\infty} X^{-1}(B_i)\right) \\ &= \bigsqcup_{i=1}^{\infty} P(X^{-1}(B_i)) \\ &= \sum_{i=1}^{\infty} \mathcal{P}(B_i). \end{aligned}$$

Portanto, vale a σ -aditividade.

Observe que se os B_i 's são disjuntos em \mathcal{B} , suas pré-imagens são disjuntas em \mathcal{F} (ver (6.13)):

$$\begin{aligned} X^{-1}(B_i) \cap X^{-1}(B_j) &= X^{-1}(B_i \cap B_j) \\ &= \emptyset \blacksquare \end{aligned}$$

\mathcal{P} é uma nova medida de probabilidade. Temos uma nova tripla, $(\mathbb{R}, \mathcal{B}, \mathcal{P})$, diferente de $(\mathbb{R}, \mathcal{B}, |\cdot|)$, onde $|\cdot|$ indica a medida de Lebesgue. **2014-03-28T10:34:14 dúvida: nesta última tripla devemos ter: \mathcal{B} ou \mathcal{L} ?**

Mudemos agora de notação: os valores de $\mathcal{P}((-\infty, x])$ determinam as probabilidades em todos os conjuntos restantes de \mathcal{B} . De fato, note que

$$\mathcal{P}((a, b]) = \mathcal{P}((-\infty, b]) - \mathcal{P}((-\infty, a])$$

trivialmente (por aditividade finita). A partir de agora, basta gerar novos conjuntos B a partir dos intervalos $(a, b]$ usando a estrutura de semi-anel. **2014-03-28T10:37:42 eu acho que isso poderia ser melhor dito**

Defina agora a FDA:

$$F_X(x) \equiv \mathcal{P}((-\infty, x]).$$

Temos que

$$\begin{aligned} F_X : \mathbb{R} &\rightarrow \mathbb{R}, \\ x &\mapsto \mathcal{P}((-\infty, x]) \in [0, 1], \forall x \in \mathbb{R}. \\ F_X(x) &\geq 0, \end{aligned}$$

$F_X(x)$ é monótona não-decrescente.

Exemplo 6.3 A distribuição binomial é

$$P(X = k) = \binom{n}{k} \left(\frac{1}{2}\right)^k.$$

Um espaço amostral possível é o conjunto das (2^n) enuplas $(\omega_1, \dots, \omega_n)$, com $\omega_i = 1$ ou $\omega_i = 0$. A variável aleatória associada à distribuição Bernoulli neste caso será

$$\begin{aligned} X : \{0, 1\} \times \dots \times \{0, 1\} &\rightarrow \mathbb{R}, \\ \omega = (\omega_1, \dots, \omega_n) &\mapsto X(\omega) = \sum_{i=1}^n \omega_i. \end{aligned}$$

A figura 6.2 mostra a $F_X(x)$ para $n = 6$.

São propriedades da FDA:

1. $F_X(x)$ é monótona não-decrescente.
- 2.

$$\lim_{x \rightarrow -\infty} F_X(x) = 0.$$

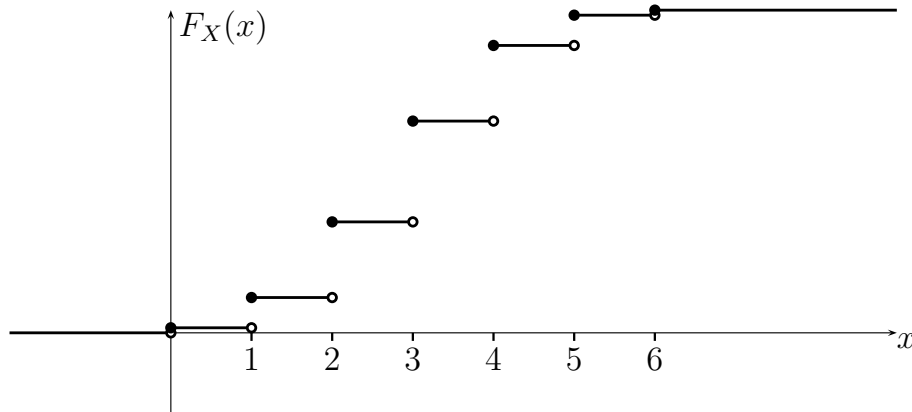


Figura 6.2: Função distribuição acumulada da distribuição binomial, $n = 6$.

3.

$$\lim_{x \rightarrow +\infty} F_X(x) = 1.$$

Reciprocamente, se uma função $F_X(x)$ possui as propriedades acima, então $F_X(x)$ é a FDA de *alguma* variável aleatória. Assim,

Qualquer função $F(x)$ com as propriedades acima define uma medida de probabilidade \mathcal{P} em $(\mathbb{R}, \mathcal{B})$

Exemplo 6.4 Se $\Omega = \{\omega_0, \omega_1, \dots\}$ é um conjunto enumerável e cada evento $\{\omega_i\}$ é um átomo de Ω com $P\{\omega_i\} = p_i$, então

$$F_X(x) = \sum_{X(\omega_i) \leq x} p_i,$$

pois

$$X^{-1}(-\infty, x] = \{\omega_i : X(\omega_i) \leq x\}.$$

Definição: Mas isto é mesmo uma definição??? A tripla $(\mathbb{R}, \mathcal{B}, \mathcal{P})$ é absolutamente contínua em relação a $(\mathbb{R}, \mathcal{B}, |\cdot|)$ quando

$$|B| = 0 \Rightarrow \mathcal{P}(B) = 0, \forall B \in \mathcal{B}.$$

Exemplo 6.5 Se $f : \mathbb{R} \rightarrow \mathbb{R}_+$ for integrável segundo Lebesgue, e

$$\int_{\mathbb{R}} f(x) \, d_L x = 1,$$

então

$$\mathcal{P}(B) = \int_B f(x) \, d_L x = \int_{\mathbb{R}} \mathbb{1}_B(x) f(x) \, d_L x$$

é uma medida de probabilidade, com

$$F_x(x) = \int_{-\infty}^x f(x) \, d_L x,$$

e

$$F_x(b) - F_x(a) = \int_a^b f(x) \, d_L x.$$

2014-03-30T09:04:06 Por que precisamos escrever explicitamente essa última equação?.

$f(x)$ é a *função densidade de probabilidade*.

Teorema 6.1 Seja $F_X(x)$ a FDA da variável aleatória X . O conjunto de descontinuidades de $F_X(x)$ é enumerável (ou finito). As descontinuidades são de tipo “salto”.

Ver seção 2.3 de Bartle — precisamos da referência bibliográfica completa!

Seja agora $\{x_1, \dots, x_n, \dots\} \subset \mathbb{R}$ o conjunto das descontinuidades. Definimos, em cada um desses pontos,

$$p_i \equiv F_X(x_i) - F_X(x_{i-})$$

Definimos agora uma função cuja aparência é similar à que obtivemos no exemplo 6.3, ilustrada na figura 6.2:

$$F_d(x) \equiv \sum_{i: x_i \leq x} p_i.$$

É importante entendermos o que se está fazendo aqui. Estamos “construindo” a parte não-contínua da FDA. Qualquer variável aleatória possui uma FDA que pode ser decomposta em 3 componentes. A parte descontínua é uma delas. O teorema (6.1) e a definição acima provêem a definição da parte descontínua.

Continuando, o que dizer sobre a parte contínua? Temos o seguinte fato (**Um outro teorema?? Creio que sim**)

Toda função monótona possui derivada em quase todos os pontos (existe a derivada exceto em um subconjunto de medida nula, segundo Lebesgue). Podemos portanto, dada uma F com tais características, definir sua FDP:

$$f(x) = \begin{cases} F'(x), & \text{onde ela existe,} \\ 0, & \text{nos demais pontos} \end{cases}$$

Curiosamente (e estranhamente para mim!), isso não é tudo. De fato, se definirmos

$$F_{ac}(x) \equiv \int_{-\infty}^x f(x) \, dx$$

ainda sobra a “parte singular”, que definimos:

$$F_S(x) \equiv F_X(x) - F_d(x) - F_{ac}(x).$$

O final da aula me deixou em dúvida: por que fazemos o seguinte?

Teste discreto \exists um subconjunto enumerável com probabilidade 1 (???)

Teste contínuo

$$|B| = 0 \Rightarrow \mathcal{P}(B) = 0.$$

(???)

Teste singular \mathcal{P} é singular em relação a $|\cdot|$ se existe $B_0 \in \mathcal{B}$, $|B_0| = 0$, mas com $\mathcal{P}(B) = \mathcal{P}(B \cap B_0)$, $\mathcal{P}(B_0) = 1$.

2014-03-30T09:21:14 Vá agora para o final do arquivo, onde eu preparei a aula de **2014-03-31**

6.5 A FDA

Nós agora vamos definir a função distribuição acumulada (FDA) de probabilidade de uma variável aleatória X :

$$F_X(x) \equiv P(\{X \leq x\}). \quad (6.20)$$

A notação da equação (6.20) é normalmente simplificada para $F_X(x) \equiv P\{X \leq x\}$. A FDA nos permite calcular facilmente probabilidade relacionadas a X . Por exemplo, para obter a probabilidade de que X ocorra em um intervalo $[a, b]$:

$$P\{a \leq X \leq b\} = P\{X \leq b\} - P\{X \leq a\} \quad (6.21)$$

$$= F(b) - F(a). \quad (6.22)$$

As propriedades da função distribuição acumulada são:

$$F(-\infty) = 0, \quad (6.23)$$

$$F(+\infty) = 1, \quad (6.24)$$

$$F(b) - F(a) \geq 0, \text{ para } a < b. \quad (6.25)$$

A função densidade de probabilidade (FDP), f_X , é

$$f_X(x) \equiv \frac{dF_X}{dx}, \quad (6.26)$$

com $f_X(x) \geq 0$ e

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1. \quad (6.27)$$

6.6 Média e momentos

A média de uma variável aleatória, ou valor esperado de uma variável aleatória, é

$$\langle X \rangle \equiv \int_{-\infty}^{+\infty} x f_X(x) dx. \quad (6.28)$$

De maneira mais geral, seja $y = q(x)$ uma função de x ; então para cada sorteio de X corresponderá um $Y = q(X)$, ou seja: Y será (também) uma variável aleatória. A média de Y será dada por

$$\langle Y \rangle = \langle q(X) \rangle = \int_{-\infty}^{+\infty} q(x) f_X(x) dx. \quad (6.29)$$

Momentos centrais desempenham um papel importante em Teoria de Probabilidade. A definição do momento central de ordem n é

$$c_{Xn} \equiv \langle (x - \langle X \rangle)^n \rangle = \int_{-\infty}^{+\infty} (x - \langle X \rangle)^n f_X(x) dx. \quad (6.30)$$

6.7 Distribuições conjuntas

Para definirmos distribuições conjuntas de probabilidade, precisamos estender as definições da seção 6.1.

No \mathbb{R}^2 , uma variável aleatória é uma função $\mathbf{X} = (X, Y)$ função de $\omega \in \Omega$:

$$\begin{aligned} \mathbf{X} : \Omega &\rightarrow \mathbb{R}^2, \\ \omega &\mapsto \mathbf{x} = \mathbf{X}(\omega). \end{aligned}$$

A função densidade acumulada conjunta será

$$F_{X,Y}(x, y) \equiv P \{ X \leq x \wedge Y \leq y \}, \quad (6.31)$$

e a função densidade de probabilidade será

$$f_{X,Y}(x, y) \equiv \frac{\partial^2 F_{X,Y}}{\partial x \partial y} = \frac{\partial^2 F_{X,Y}}{\partial y \partial x}. \quad (6.32)$$

As funções densidade de probabilidade *marginais* são

$$f_X(x) \equiv \int_{y=-\infty}^{+\infty} f_{X,Y}(x, y) dy, \quad (6.33)$$

$$f_Y(y) \equiv \int_{x=-\infty}^{+\infty} f_{X,Y}(x, y) dx. \quad (6.34)$$

Naturalmente,

$$\iint_{x,y} f_{X,Y}(x, y) dx dy = 1. \quad (6.35)$$

O *valor esperado* de uma função de X e Y é

$$\langle g(X, Y) \rangle = \iint_{x,y} g(x, y) f_{X,Y}(x, y) dx dy. \quad (6.36)$$

Com os resultados acima, nós podemos agora deduzir o seguinte fato: se X e Y são duas variáveis aleatórias quaisquer,

$$\langle X + Y \rangle = \langle X \rangle + \langle Y \rangle. \quad (6.37)$$

a) *Insight* com matemática finita: seja (x_i, y_i) , $i = 0, \dots, n-1$ uma amostra qualquer de n pares de observações. Então, a *média amostral* da soma é

$$\begin{aligned}\overline{x+y} &= \frac{1}{n} \sum_{i=0}^{n-1} (x_i + y_i) \\ &= \frac{1}{n} \sum_{i=0}^{n-1} x_i + \frac{1}{n} \sum_{i=0}^{n-1} y_i \\ &= \bar{x} + \bar{y}.\end{aligned}$$

Note como não foi necessária nenhuma hipótese adicional sobre a natureza dos x 's e y 's ou sobre qualquer relação entre eles.

b) com Teoria de Probabilidade. Seja $g(x, y) = x + y$. Então,

$$\begin{aligned}\langle g(X, Y) \rangle &= \langle X + Y \rangle = \iint_{x,y} (x + y) f_{X,Y}(x, y) \, dy \, dx \\ &= \iint_{x,y} x f_{X,Y}(x, y) \, dy \, dx + \iint_{x,y} y f_{X,Y}(x, y) \, dy \, dx \\ &= \iint_{x,y} x \underbrace{[f_{X,Y}(x, y) \, dy]}_{f_X(x)} \, dx + \iint_{x,y} y \underbrace{[f_{X,Y}(x, y) \, dx]}_{f_Y(y)} \, dy \\ &= \int_{-\infty}^{+\infty} x f_X(x) \, dx + \int_{-\infty}^{+\infty} y f_Y(y) \, dy \\ &= \langle X \rangle + \langle Y \rangle.\end{aligned}$$

Já que estamos falando de distribuições conjuntas, devemos tocar no conceito fundamental de independência/dependência. Dados dois eventos A e B , a probabilidade de B condicionada a A é

$$P(B|A) \equiv \frac{P(A \cap B)}{P(A)} \quad (6.38)$$

Definição: B e A são independentes quando

$$P(B|A) = P(B). \quad (6.39)$$

Corolário: Se A e B são independentes,

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = P(B) \Rightarrow P(A \cap B) = P(A)P(B). \quad (6.40)$$

Isso nos leva imediatamente à definição de independência de variáveis aleatórias: X e Y são independentes quando os *eventos*

$$A = \{\omega | X(\omega) \leq x\} \text{ e } B = \{\omega | Y(\omega) \leq y\}$$

são independentes, ou seja:

$$\begin{aligned}P(A \cap B) &= P(A)P(B), \\ P(X(\omega) \leq x \wedge Y(\omega) \leq y) &= P(X(\omega) \leq x) P(Y(\omega) \leq y), \\ F_{X,Y}(x, y) &= F_X(x)F_Y(y).\end{aligned}$$

Portanto, se X e Y são independentes, então a FDA conjunto é igual ao produto das FDA's marginais. Fazendo a 2ª derivada cruzada,

$$\begin{aligned}
 f_{X,Y}(x, y) &= \frac{\partial^2 F_{X,Y}(x, y)}{\partial y \partial x} \\
 &= \frac{\partial^2 F_X(x) F_Y(y)}{\partial y \partial x} \\
 &= \frac{\partial}{\partial y} \left[\frac{\partial F_X}{\partial x} F_Y(y) \right] \\
 &= f_X(x) \frac{\partial F_Y}{\partial y} \\
 &= f_X(x) f_Y(y).
 \end{aligned}$$

A próxima definição é a densidade de probabilidade de y *condicionada* à ocorrência de x :

$$f_{Y|x}(y) \equiv \frac{f_{X,Y}(x, y)}{f_X(x)}. \quad (6.41)$$

Embora deva ser possível deduzir rigorosamente (6.41) a partir de (6.38), não vou fazê-lo (ainda). Em vez disto, vou procurar um caso particular muito interessante. Primeiramente, note que, por definição, a densidade marginal de y é dada por

$$\begin{aligned}
 f_Y(y) &= \int_{x=-\infty}^{+\infty} f_{X,Y}(x, y) dx \\
 &= \int_{x=-\infty}^{+\infty} f_{Y|x}(y) f_X(x) dx.
 \end{aligned} \quad (6.42)$$

Considere agora o caso em que $y = g(x)$, ou seja: em que Y está deterministicamente determinado a partir da observação de x . A função densidade de probabilidade condicionada é então, simplesmente,

$$f_{Y|x}(y) = \delta(g(x) - y), \quad (6.43)$$

ou seja: dado x , a probabilidade de que $Y = g(x)$ é 1 (note as letras maiúscula e minúscula). Note também que (6.43) é uma densidade de probabilidade legítima, já que sua integral em y é igual a 1. O primeiro resultado que vamos obter a partir daqui é uma fórmula para $f_Y(y)$:

$$\begin{aligned}
 f_Y(y) &= \int_{x=-\infty}^{+\infty} f_{Y|x}(y) f_X(x) dx \\
 &= \int_{x=-\infty}^{+\infty} \delta(g(x) - y) f_X(x) dx.
 \end{aligned} \quad (6.44)$$

Essa é uma equação totalmente geral, que permite o cálculo da densidade de probabilidade de *qualquer* variável aleatória definida por uma função, $Y = g(X)$, independentemente de ela ser biunívoca ou não! A equação (6.44) pode ser prontamente generalizada para funções de várias variáveis. Em particular, se $Z = g(X, Y)$, tem-se

$$f_Z(z) = \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} \delta(g(x, y) - z) f_{X,Y}(x, y) dy dx. \quad (6.45)$$

Considere agora o seguinte exemplo: $y = x^2$, $-1 \leq X \leq +1$, com $f_X(x) = 1/2$. Essa função *não* é biunívoca. A densidade de probabilidade de y é

$$f_Y(y) = \int_{-1}^{+1} \delta(x^2 - y) \frac{1}{2} dx. \quad (6.46)$$

Para calcular a integral acima, é preciso um pouco de cuidado. Primeiramente, note que

$$y = x^2 \Rightarrow x = \pm\sqrt{y} : \quad (6.47)$$

existem duas raízes, e isto precisa ser levado em consideração pelo ser humano que está resolvendo o problema: a fórmula (6.44) *em si* não vai dizer isso para você! O caminho mais rápido, aparentemente, é recorrer à seguinte propriedade da delta (Butkov, 1988, Cap. 6, p. 231):

$$\delta(x^2 - a^2) = (1/2a) [\delta(x + a) + \delta(x - a)] \quad (a > 0); \quad (6.48)$$

então,

$$f_Y(y) = \int_{-1}^{+1} \frac{1}{2\sqrt{y}} [\delta(x + \sqrt{y}) + \delta(x - \sqrt{y})] \frac{1}{2} dx = \frac{1}{4\sqrt{y}} [1 + 1] = \frac{1}{2\sqrt{y}} \blacksquare \quad (6.49)$$

Esse é um exemplo simples, porém muito rico. O valor esperado de Y , para o qual nós vamos usar a notação $\langle Y \rangle$, agora é facilmente obtido:

$$\langle Y \rangle = \int_0^1 y f_Y(y) dy = \int_0^1 y \frac{1}{2\sqrt{y}} dy = \int_0^1 \frac{\sqrt{y}}{2} dy = \frac{1}{3}, \quad (6.50)$$

enquanto que, devido à simetria de $f_X(x)$, $\langle X \rangle = 0$.

Mais interessante ainda é a seguinte questão: se X é a variável aleatória com distribuição uniforme entre $-1/2$ e $+1/2$, como acima, e $Y = X^2$, qual é a covariância entre X e Y ? Por definição, a covariância entre duas variáveis aleatórias é

$$\text{Cov}\{X, Y\} = \int_{x \in \mathbb{R}} \int_{y \in \mathbb{R}} (x - \langle X \rangle)(y - \langle Y \rangle) f_{X,Y}(x, y) dy dx. \quad (6.51)$$

Quando duas variáveis aleatórias são independentes, sua covariância é nula:

$$\begin{aligned} \text{Cov}\{X, Y\} &= \iint_{x \in \mathbb{R}, y \in \mathbb{R}} (x - \langle X \rangle)(y - \langle Y \rangle) f_{X,Y}(x, y) dy dx \\ &= \iint_{x \in \mathbb{R}, y \in \mathbb{R}} (x - \langle X \rangle)(y - \langle Y \rangle) f_X(x) f_Y(y) dy dx \\ &= \left[\int_{x \in \mathbb{R}} (x - \langle X \rangle) f_X(x) dx \right] \left[\int_{y \in \mathbb{R}} (y - \langle Y \rangle) f_Y(y) dy \right] \\ &= 0. \end{aligned}$$

No nosso caso particular, $y = x^2$, a covariância é

$$\begin{aligned}
\text{Cov}\{X, Y\} &= \int_{x=-1}^1 \int_{y=0}^1 x(y - \frac{1}{3}) \delta(x^2 - y) f_X(x) dy dx \\
&= \int_{x=-1}^1 \int_{y=0}^1 x(x^2 - \frac{1}{3}) \delta(x^2 - y) \frac{1}{2} dy dx \\
&= \int_{x=-1}^1 x(x^2 - \frac{1}{3}) \int_{y=0}^1 \delta(x^2 - y) \frac{1}{2} dy dx \\
&= \frac{1}{2} \int_{x=-1}^1 x(x^2 - \frac{1}{3}) \underbrace{\int_{y=0}^1 [\delta(x^2 - y) dy]}_{=1} dx \\
&= \frac{1}{2} \int_{x=-1}^1 x(x^2 - \frac{1}{3}) dx = 0.
\end{aligned} \tag{6.52}$$

Na linha acima de (6.52), note a utilização simultânea de duas propriedades da delta:

$$\int_{-\infty}^{+\infty} \delta(x - a) dx = 1, \tag{6.53}$$

$$\delta(x) = \delta(-x). \tag{6.54}$$

O resultado (6.52) é admirável: ele mostra que, mesmo que a dependência entre duas variáveis aleatórias seja *total*; mesmo que Y seja *totalmente* dependente de X na forma $Y = g(X)$, é possível que $\text{Cov}\{X, Y\} = 0$. Moral da história: a covariância é uma medida de dependência *linear*, e *não* uma medida universal de dependência.

Além da covariância, é usual encontrar o coeficiente de correlação,

$$\varrho_{XY} = \frac{\text{Cov}\{X, Y\}}{[\text{Var}\{X\} \text{Var}\{Y\}]^{1/2}}.$$

Com a desigualdade de Schwarz,

$$-1 \leq \varrho_{XY} \leq +1$$

Considere agora o cálculo da função densidade de probabilidade da variável aleatória $Z = X + Y$, *soma* de duas variáveis aleatórias X e Y cuja distribuição conjunta de probabilidade, $f_{X,Y}(x, y)$, é conhecida. A função densidade de probabilidade de Z dada a ocorrência de x, y é

$$f_{Z|x,y} = \frac{f_{Z,X,Y}(z, x, y)}{f_{X,Y}(x, y)} = \delta(z - [x + y]) \tag{6.55}$$

Segue-se que a função densidade de probabilidade marginal de Z é

$$\begin{aligned}
f_Z(z) &= \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} f_{Z,X,Y}(z, x, y) dy dx \\
&= \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} \delta(z - x - y) f_{X,Y}(x, y) dy dx
\end{aligned} \tag{6.56}$$

Lembrando das propriedades da Delta de Dirac,

$$\int_{-\infty}^{+\infty} \delta(y - a) f(y) dy = f(a), \tag{6.57}$$

$$\delta(y - a) = \delta(a - y), \tag{6.58}$$

tem-se

$$f_Z(z) = \int_{x=-\infty}^{+\infty} f_{X,Y}(x, z-x) dx = \int_{y=-\infty}^{+\infty} f_{X,Y}(z-y, y) dy \blacksquare \quad (6.59)$$

Finalmente, no caso de variáveis aleatórias X e Y independentes, as integrais acima tornam-se integrais de convolução:

$$f_Z(z) = \int_{x=-\infty}^{+\infty} f_X(x) f_Y(z-x) dx. \quad (6.60)$$

Note que esta é uma convolução no sentido da Teoria de Transformadas de Fourier, e não no sentido da Teoria de Transformadas de Laplace.

Quando, por outro lado, a relação $Y = g(X)$ é biunívoca, as coisas ficam mais fáceis:

$$\begin{aligned} P\{Y \leq y\} &= P\{X \leq x\}, \\ P\{Y \leq g(x)\} &= P\{X \leq x\}, \\ F_Y(g(x)) &= F_X(x), \\ \frac{d}{dx} F_Y(g(x)) &= \frac{d}{dx} F_X(x), \\ \frac{d}{dy} F_Y(g(x)) \frac{dg}{dx} &= f_X(x), \\ f_Y(y) \frac{dg}{dx} &= f_X(x). \end{aligned}$$

Mudando ligeiramente a notação, $y = y(x)$ produz

$$f_Y(y) dy = f_X(x) dx$$

cujas interpretações gráficas são óbvias.

Um caso particularmente útil e interessante é o de geração de uma variável aleatória no computador com uma distribuição dada. A resposta é o uso da própria FDA $F_X(x)$ no lugar de $g(x)$. Seja $U = F_X(X)$ a variável aleatória assim gerada. Nesse caso,

$$\begin{aligned} f_U(u) du &= f_X(x) dx, \\ f_U(u) \frac{du}{dx} &= f_X(x). \end{aligned}$$

mas

$$\frac{du}{dx} = \frac{dF_X(x)}{dx} = f_X(x);$$

portanto,

$$f_U(u) = 1, \quad 0 \leq u \leq 1. \quad (6.61)$$

A “receita” para gerar uma variável aleatória com FDA $F_x(x)$ é a seguinte:

- a) gere uma variável u uniforme em $[0, 1]$.
- b) calcule $x = F_X^{-1}(u)$.

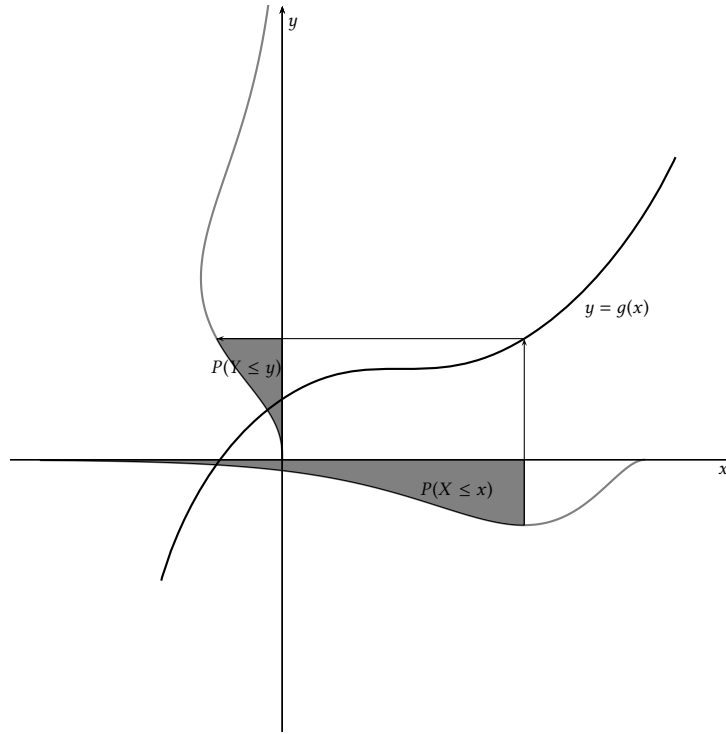


Figura 6.3: Relação entre as distribuições de probabilidade de X e Y quando a relação entre ambos é biunívoca.

Quando existe uma dependência linear, do tipo $Y = aX + b$, a *forma* da distribuição não muda:

$$\begin{aligned} x &= \frac{y - b}{a}, \\ \frac{dy}{dx} &= a, \\ f_Y(y)dy &= f_X(x)dx, \\ f_Y(y)\frac{dy}{dx} &= f_X(x), \\ f_Y(y)a &= f_X(x), \\ f_Y(y) &= \frac{1}{a}f_X\left(\frac{y - b}{a}\right). \end{aligned}$$

Seja Y o estimador de uma variável cujo valor verdadeiro (de população) é y . No caso mais geral,

$$\langle Y \rangle = y' \neq y. \quad (6.62)$$

A diferença entre o valor médio previsto pelo modelo (y') e o valor verdadeiro (y) é denominada o *viés* do modelo.

A figura 6.3 dá uma visão dessas relações.

Exemplo: soma de exponenciais

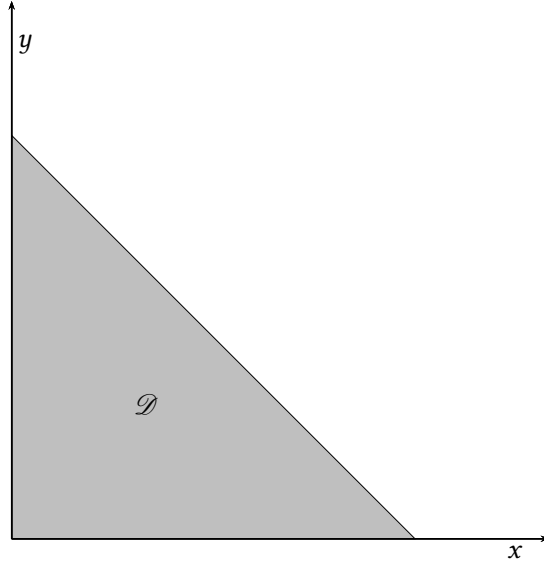


Figura 6.4: Região de integração para cálculo da distribuição da soma de duas variáveis aleatórias exponenciais.

Sejam

$$f_X(x) = \frac{1}{\beta} \exp\left(-\frac{x}{\beta}\right),$$

$$f_Y(y) = \frac{1}{\beta} \exp\left(-\frac{y}{\beta}\right),$$

ou seja: duas variáveis aleatórias X e Y independentes e igualmente distribuídas com distribuição exponencial. Se $Z = X + Y$,

$$\begin{aligned} P\{Z \leq z\} &= P\{X + Y \leq z\} \\ &= P\{(X, Y) \in \mathcal{D}\} \\ &= P(\mathcal{D}). \end{aligned}$$

Veja a figura [6.4](#)

Agora,

$$\begin{aligned} P\{X + Y \leq z\} &= \int_{x=0}^z \int_{y=0}^{z-x} \frac{1}{\beta} \exp\left(-\frac{x}{\beta}\right) \frac{1}{\beta} \exp\left(-\frac{y}{\beta}\right) dy dx \\ &= \int_{x=0}^z \exp\left(-\frac{x}{\beta}\right) \left[\int_{y=0}^{z-x} \exp\left(-\frac{y}{\beta}\right) d\left(\frac{y}{\beta}\right) \right] d\left(\frac{x}{\beta}\right) \\ &= \int_{x=0}^z \exp\left(-\frac{x}{\beta}\right) \left[1 - \exp\left(-\frac{z-x}{\beta}\right) \right] d\left(\frac{x}{\beta}\right) \\ &= \int_0^z \left(e^{-x/\beta} - e^{-z/\beta} \right) d\frac{x}{\beta} \\ &= 1 - e^{-z/\beta} - \frac{z}{\beta} e^{-z/\beta}. \end{aligned}$$

Portanto,

$$F_Z(z) = 1 - e^{-z/\beta} - \frac{z}{\beta} e^{-z/\beta},$$
$$f_Z(z) = \frac{z}{\beta^2} e^{-z/\beta}.$$

A fdp tem a forma de distribuição gama, cuja fórmula geral é

$$f_X(x) = \frac{1}{\beta \Gamma(\alpha)} \left(\frac{x}{\beta} \right)^{\alpha-1} e^{-\frac{x}{\beta}}. \quad (6.63)$$

Para $\alpha = 2$, $\Gamma(\alpha) = 1$, e a fórmula coincide. Esse é uma caso particular do seguinte resultado, mais geral:

$$X_1, \dots, X_n \sim \text{EXP}(\beta) \Rightarrow Z = \sum_{i=1}^n X_i \sim \text{GAMA}(n, \beta).$$

Capítulo 7

Modelos e erros

Exemplo 7.1 Seja M um modelo de previsão da temperatura da água de um rio. O modelo funciona de acordo com o seguinte esquema:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \longrightarrow \boxed{M(\mathbf{x}; \boldsymbol{\theta})} \longrightarrow y' \neq y.$$

Aqui, o vetor $\mathbf{x} = (x_1, x_2, \dots, x_n)$ representa os *dados de entrada*; eles podem ser, por exemplo, a radiação solar incidente, a vazão do rio, a temperatura do ar, etc.. O modelo M em si é composto de elementos tais como a geometria rio (largura, profundidade, comprimento), pelas as equações utilizadas para calcular ou estimar a temperatura y , etc.. A *natureza* de M pode ser qualquer. Por exemplo, M pode ser simplesmente uma regressão linear múltipla

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (7.1)$$

mas também pode ser um modelo físico de balanço de entalpia cuja incógnita seja a temperatura da água (Dias, 2003; Dias et al., 2003). Finalmente note que existe um vetor de *parâmetros* $\boldsymbol{\theta}$ de M . No caso da regressão múltipla (7.1), os parâmetros são a_0, a_1, \dots, a_n .

A partir do esquema geral representado pelo Exemplo 7.1, é fácil identificar que numerosos erros podem ocorrer no processo de modelagem. Em geral, nós classificamos os erros em 3 tipos

- A) Erro dos dados: os dados de entrada do modelo geralmente contêm erros. Só isto já faz com que os dados de entrada devam ser considerados como variáveis aleatórias X_1, X_2, \dots, X_n e consequentemente a saída do modelo, Y , também.
- B) Erro do modelo: todo modelo contém imperfeições e limitações, e muitas vezes é extremamente difícil até mesmo identificar inequivocamente um único modelo (por exemplo, é comum que diversas distribuições diferentes de probabilidade representem “igualmente bem” um determinado conjunto de observações).
- C) Erro dos parâmetros. Mesmo que o modelo seja perfeito, a aleatoriedade seja dos dados de entrada seja das próprias observações y nas quais nos baseamos para “calibrar” M — o processo de estimativa dos parâmetros $\boldsymbol{\theta}$ — faz com que estes últimos também carreguem incerteza.

Algumas estatísticas básicas de erros são definidas a seguir.

O vés do modelo é

$$v \equiv \langle Y \rangle - y = y' - y. \quad (7.2)$$

O erro médio quadrático do modelo é

$$\text{EMQ}\{Y\} \equiv \langle (Y - y)^2 \rangle \quad (7.3)$$

e sua raiz quadrada, que tem as mesmas dimensões de Y , é

$$\text{REMQ}\{Y\} = \sqrt{\text{EMQ}\{Y\}}. \quad (7.4)$$

A variância de Y é

$$\text{Var}\{Y\} = \langle (Y - \langle Y \rangle)^2 \rangle. \quad (7.5)$$

Muito importante: como em geral $\langle Y \rangle \neq y$, $\text{EMQ}\{Y\} \neq \text{Var}\{Y\}$. A relação entre ambas as estatísticas pode ser obtida muito facilmente:

$$\begin{aligned} \text{EMQ}\{Y\} &= \langle (Y - y)^2 \rangle \\ &= \langle [(Y - y') + (y' - y)]^2 \rangle \\ &= \langle (Y - y')^2 + 2(Y - y')(y' - y) + (y' - y)^2 \rangle \\ &= \langle (Y - y')^2 \rangle + 2(y' - y) \underbrace{\langle (Y - y') \rangle}_{=0} + \langle (y' - y)^2 \rangle \\ &= \text{Var}\{Y\} + v^2 \blacksquare \end{aligned} \quad (7.6)$$

Um bom modelo deve combinar duas virtudes:

- ser acurado: $v \rightarrow 0$;
- ser preciso: $\text{Var}\{Y\} \rightarrow 0$.

Capítulo 8

A difusão de um ponto de vista probabilístico

8.1 Um passeio aleatório

Na figura 8.1, uma partícula move-se a partir da origem (com certeza) em passos de tempo n , dando “saltos” de 1 unidade a cada tempo. A posição da partícula no tempo n é $X(n)$. A *probabilidade* de a partícula estar na posição k no instante n é

$$p_k(n) \equiv P\{X(n) = k\}. \quad (8.1)$$

Suponha que uma transição para cima ou para baixo seja igualmente provável, e *independente* do estado atual $X(n)$:

$$P\{X(n+1) = X(n) + 1\} = 1/2, \quad (8.2)$$

$$P\{X(n+1) = X(n) - 1\} = 1/2. \quad (8.3)$$

Então o estado k só pode ser alcançado a partir de $k+1$ ou de $k-1$, e

$$p_k(n+1) = \frac{1}{2}p_{k+1}(n) + \frac{1}{2}p_{k-1}(n). \quad (8.4)$$

Para prosseguirmos, nós vamos necessitar de derivadas numéricas. Veja a figura 8.2: as derivadas numéricas *centradas* em $x = (i+1/2)\Delta x$, e $x = (i-1/2)\Delta x$ são

$$f'_{i+1/2} \approx \frac{f_{i+1} - f_i}{\Delta x}, \quad (8.5)$$

$$f'_{i-1/2} \approx \frac{f_i - f_{i-1}}{\Delta x}. \quad (8.6)$$

A derivada numérica é o coeficiente angular da secante entre 2 pontos; a derivada *exata* é o coeficiente angular da tangente geométrica. Na figura 8.2, ambas são mostradas: se você possuir uma vista suficientemente boa, procure identificar visualmente a diferença entre elas.

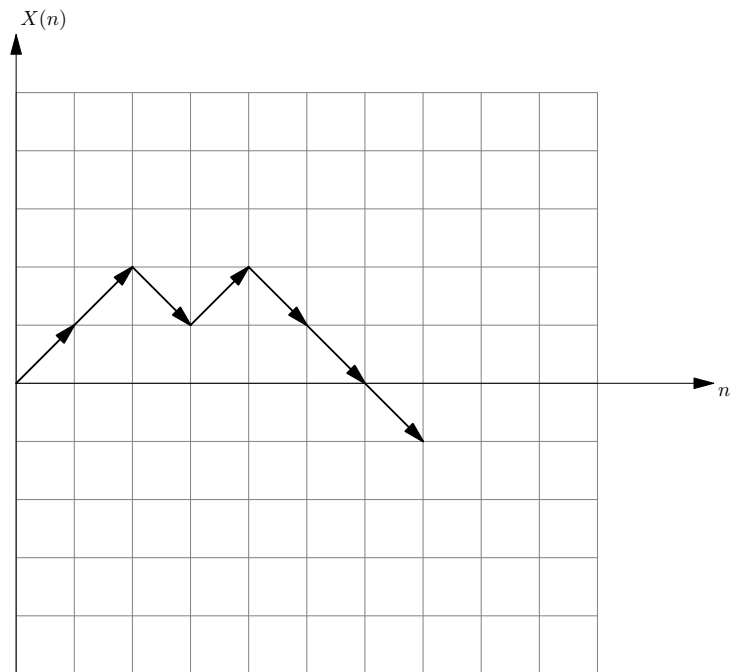


Figura 8.1: Um passeio aleatório simples, começando sempre em $X = 0$.

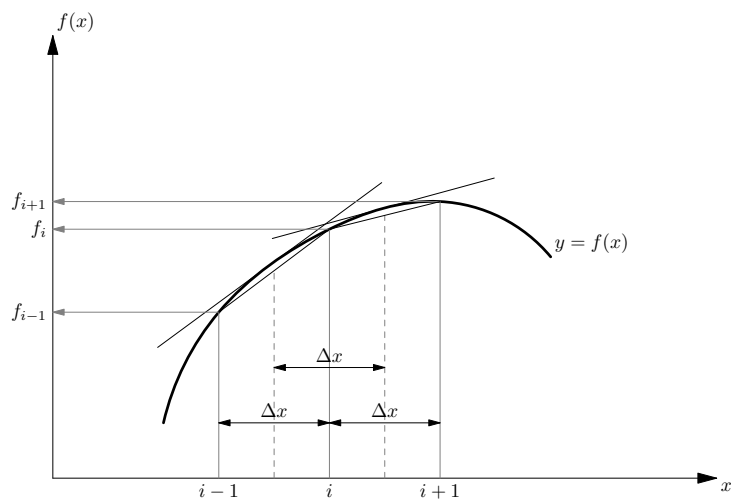


Figura 8.2: Derivadas numéricas de ordem 1 e 2.

A derivada segunda é a derivada da derivada:

$$\begin{aligned}
 f_i'' &\approx \frac{f'_{i+1/2} - f'_{i-1/2}}{\Delta x} \\
 &= \frac{\frac{f_{i+1} - f_i}{\Delta x} - \frac{f_i - f_{i-1}}{\Delta x}}{\Delta x} \\
 &= \frac{f_{i+1} - 2f_i + f_{i-1}}{\Delta x^2}.
 \end{aligned} \tag{8.7}$$

Com isto, nós podemos agora manipular a equação (8.4):

$$\begin{aligned}
 p_k(n+1) &= \frac{1}{2} (p_{k+1}(n) + p_{k-1}(n)) \\
 2p_k(n+1) &= p_{k+1}(n) + p_{k-1}(n) \\
 2(p_k(n+1) - p_k(n)) &= p_{k+1}(n) - 2p_k(n) + p_{k-1}(n) \\
 \frac{p_k(n+1) - p_k(n)}{\Delta x^2} &= \frac{1}{2} \frac{p_{k+1}(n) - 2p_k(n) + p_{k-1}(n)}{\Delta x^2} \\
 \frac{p_k(n+1) - p_k(n)}{\Delta t} &= \frac{\Delta x^2}{2\Delta t} \frac{p_{k+1}(n) - 2p_k(n) + p_{k-1}(n)}{\Delta x^2}
 \end{aligned} \tag{8.8}$$

Faça agora

$$p_k(n) = p(k\Delta x, n\Delta t) = p(x, t)$$

e mantenha

$$\mathcal{D} = \frac{\Delta x^2}{2\Delta t} \tag{8.9}$$

constante, enquanto $\Delta x \rightarrow 0$. (8.8) torna-se

$$\begin{aligned}
 \frac{p(x, t + \Delta t) - p(x, t)}{\Delta t} &= \mathcal{D} \frac{p(x + \Delta x, t) - 2p(x, t) + p(x - \Delta x, t)}{\Delta x^2} \rightarrow \\
 \frac{\partial p}{\partial t} &= \mathcal{D} \frac{\partial^2 p}{\partial x^2}.
 \end{aligned} \tag{8.10}$$

Esta é a mesma equação de difusão que nós resolvemos no capítulo ??! Isto significa que nós podemos dar à concentração de uma substância uma interpretação *probabilística*: ela é (também) a probabilidade de encontrarmos em (x, t) uma partícula emitida em uma certa posição (digamos, ξ) no instante $t = 0$. Mas será $\mathcal{D} = \Delta x^2/(2\Delta t)$ realmente constante? Esta pergunta possui uma resposta *estatística*, encontrada por Albert Einstein em 1905

8.2 A solução de Einstein para o movimento Browniano

Suponha que existem partículas movendo-se aleatoriamente dentro de um fluido de viscosidade dinâmica μ e com temperatura termodinâmica T . A força de resistência ao movimento de uma partícula é dada pela lei de Stokes:

$$F_R = -6\pi\mu a \frac{dX}{dt}, \tag{8.11}$$

onde X é a posição aleatória da partícula, e a é o seu raio. Em (8.11), note que X é um escalar: nossa abordagem aqui será *unidimensional*, porque isto é mais fácil algebricamente. Quando fazemos isto, perdemos um pouco do realismo da situação, e muitas pessoas têm dificuldade em raciocinar com situações que, embora matematicamente mais simples, não possuem uma representação concreta no mundo real. Entretanto, esta é uma *excelente* maneira de abordar aquele mundo real: aos poucos, introduzindo as dificuldades matemáticas apenas quando não é mais possível escamoteá-las.

As velocidades das partículas, V , dependem da temperatura do fluido (e delas mesmas), segundo a Mecânica Estatística, via

$$\left\langle \frac{mV^2}{2} \right\rangle = \frac{1}{2}kT, \quad (8.12)$$

onde k é a constante de Boltzmann.

Além de F_R , as partículas sofrem a ação de forças aleatórias F , que são produzidas pelos choques com as moléculas do fluido. A equação de movimento (unidimensional, é claro) de um partícula é

$$m \frac{d^2 X}{dt^2} = -6\pi\mu a \frac{dX}{dt} + F. \quad (8.13)$$

Volte agora ao coeficiente de difusão dado por (8.9): se partirmos de $x = 0$, $t = 0$, teremos (para cada partícula)

$$X^2 = 2\mathcal{D}t.$$

Suponha que esta relação seja válida na média de todas as partículas; então,

$$\langle X^2 \rangle = 2\mathcal{D}t, \quad (8.14)$$

$$\frac{d}{dt} \langle X^2 \rangle = 2\mathcal{D}, \quad (8.15)$$

$$\frac{d^2}{dt^2} \langle X^2 \rangle = 0. \quad (8.16)$$

Lembre-se de que por enquanto (8.14) é apenas uma suposição, que nós devemos ser capazes de provar. Agora,

$$\frac{d}{dt} \langle X^2 \rangle = \left\langle \frac{d}{dt} X^2 \right\rangle = \left\langle 2X \frac{dX}{dt} \right\rangle, \quad (8.17)$$

$$\frac{d^2}{dt^2} \langle X^2 \rangle = 2 \left[\left\langle \frac{dX}{dt} \right\rangle^2 + \left\langle X \frac{d^2 X}{dt^2} \right\rangle \right]. \quad (8.18)$$

Invocando agora (8.16) (que ainda é apenas uma hipótese que precisa ser provada),

$$\left\langle X \frac{d^2 X}{dt^2} \right\rangle = - \left\langle \frac{dX}{dt} \right\rangle^2. \quad (8.19)$$

A equação (8.19) será útil em (8.13); multiplicando esta última por X e utilizando (8.19):

$$\begin{aligned} m \left\langle X \frac{d^2 X}{dt^2} \right\rangle &= -6\pi a \mu \left\langle X \frac{dX}{dt} \right\rangle + \langle XF \rangle, \\ -m \left\langle \frac{dX}{dt} \right\rangle^2 &= -6\pi a \mu \left\langle X \frac{dX}{dt} \right\rangle + \langle XF \rangle. \end{aligned} \quad (8.20)$$

Raciocine agora probabilisticamente: X é a posição de uma partícula qualquer, com $\langle X \rangle = 0$. F é a força aleatória sobre uma partícula, com $\langle F \rangle = 0$. X e F *devem* ser variáveis aleatórias independentes:

$$\text{Cov}\{X, F\} = \langle (X - \langle X \rangle)(F - \langle F \rangle) \rangle = \langle XF \rangle = 0. \quad (8.21)$$

Segue-se que

$$\begin{aligned} 6\pi a\mu \left\langle X \frac{dX}{dt} \right\rangle &= m \left\langle \frac{dX^2}{dt} \right\rangle = kT, \\ 6\pi a\mu \left\langle 2X \frac{dX}{dt} \right\rangle &= 2kT, \\ 6\pi a\mu \frac{d}{dt} \langle X^2 \rangle &= 2kT, \\ \frac{d}{dt} \langle X^2 \rangle &= \frac{kT}{3\pi a\mu} = 2\mathcal{D}, \end{aligned} \quad (8.22)$$

$$\mathcal{D} = \frac{kT}{6\pi a\mu}. \quad (8.23)$$

Observe que (8.22) é a mesma que (8.15): a hipótese (8.14) é *consistente* com o resultado que obtivemos, e o justifica *a posteriori*.

A relação entre k , o número de Avogadro N_A , e a constante universal dos gases pode ser encontrada nos bons livros do ramo: $R = kN_A$, ou seja:

$$\frac{d}{dt} \langle X^2 \rangle = \frac{RT}{3\pi a\mu N_A}; \quad (8.24)$$

esta é uma forma de determinar experimentalmente o número de Avogadro 🚫**You should re-read Abraham Pais!**

Uma maneira mais formal de deduzir que $\frac{d^2}{dt^2} \langle X^2 \rangle = 0$ é a seguinte. Suponha que X seja um processo estocástico com incrementos estacionários:

$$\begin{aligned} \frac{d}{dt} \langle (X(t) - X(s))^2 \rangle &= 0, \quad \forall t, s, \quad t > s, \\ \frac{d}{dt} \left\langle \frac{(X(t) - X(s))^2}{t - s} \right\rangle &= 0, \\ \frac{d}{dt} \left\langle \frac{(\Delta X)^2}{\Delta t} \right\rangle &= 0, \\ \left\langle \frac{(\Delta X)^2}{\Delta t} \right\rangle &= 2\mathcal{D}. \end{aligned} \quad (8.25)$$

Como vimos, a distribuição de probabilidade da posição de uma partícula no passeio aleatório descrito na seção 8.1 possui equação governante (8.10), com

$$p(x, 0) = \delta(x), \quad (8.26)$$

$$\int_{-\infty}^{+\infty} p(x, t) dx = 1, \quad \forall t. \quad (8.27)$$

É fácil interpretar estas duas últimas equações. (8.26) nos informa que $X(0) = 0$ com probabilidade 1: a partícula *sempre* sai de $x = 0$. (8.27) requer que a densidade de probabilidade possua integral unitária, o que é óbvio. Note que (8.10) juntamente com as condições (8.26)–(8.27) é um problema clássico *que nós já resolvemos*, utilizando transformada de Fourier. A solução que obtivemos é

$$p(x, t) = \frac{1}{\sqrt{4\pi Dt}} \exp \left[-\frac{x^2}{4Dt} \right]. \quad (8.28)$$

Se fizermos, de acordo com a teoria do movimento Browniano de Einstein da seção 8.2

$$\sigma^2 \equiv \langle X^2 \rangle = 2Dt \quad (8.29)$$

e substituirmos em (8.28), teremos

$$p(x, t) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left[-\frac{1}{2} \left(\frac{x}{\sigma} \right)^2 \right]. \quad (8.30)$$

Esta é uma velha conhecida nossa: trata-se da distribuição normal, ou gaussiana, de probabilidade. Neste contexto, o *desvio-padrão* da posição da partícula, σ , cresce com a raiz quadrada do tempo t .

Equações do tipo (8.29) aparecem com frequência em problemas de difusão, e seu surgimento agora deve estar mais claro para você.

8.3 Uma introdução informal a processos estocásticos

Primeiro, alguns exemplos: as condições de tempo (tempo bom — sol, ou tempo ruim — chuva), o total de milímetros de chuva precipitados a cada 24 horas, e a vazão em um rio a cada dia, são todos exemplos de fenômenos que podem ser *modelados* como processos estocásticos.

Note o cuidado com as palavras: estes fenômenos podem ser *modelados* como processos estocásticos; eles não *são* processos estocásticos. Tanto quanto sabemos, a natureza não “sabe” o que ela mesma é; somos nós que lhe atribuímos certas propriedades, e a “enfeitamos” com hipóteses e modelos.

De fato, é totalmente irrelevante o que a natureza realmente “é”: tudo o que devemos nos perguntar é se os modelos que utilizamos são uma boa descrição daqueles aspectos da natureza que consideramos importante modelar, para efeito de compreensão e de previsão

De volta aos fenômenos que identificamos como passíveis de serem modelados como processos estocásticos, observe que

1. O tempo amanhã é muito parecido com o tempo hoje: se hoje o tempo está bom, é grande a chance de o tempo também estar bom amanhã. Uma boa variável para testar isto é a pressão atmosférica.
2. Uma coisa semelhante acontece — não por coincidência, é claro — com a chuva: se hoje está chovendo, é grande a chance de ainda estar chovendo amanhã.

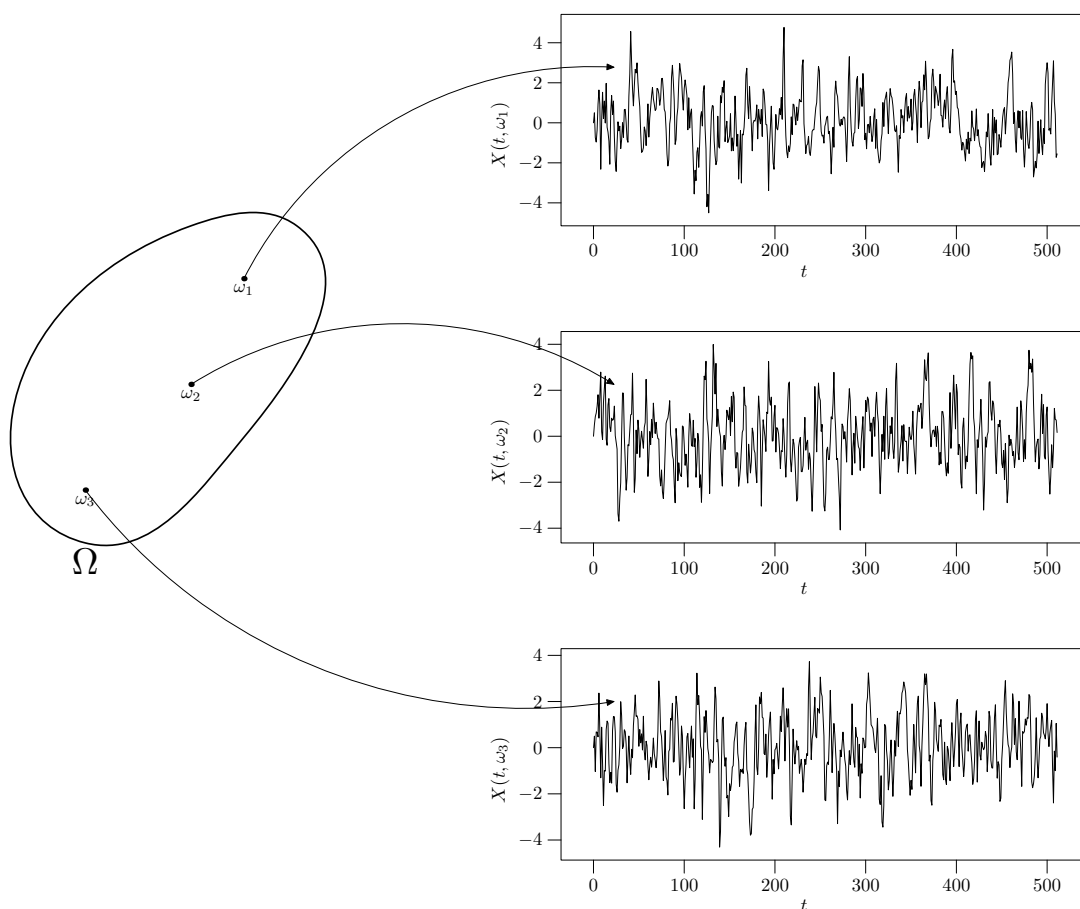


Figura 8.3: Ilustração de um processo estocástico.

3. Idem para a vazão de um rio — de novo você deve notar que estes 3 fenômenos estão intimamente ligados — se hoje a vazão de um rio está baixa (porque não chove na bacia há muitos dias), então é grande a chance de que ela continue baixa amanhã.

Por outro lado, a relação entre o tempo (ou a chuva, ou a vazão) hoje e daqui a 3 meses é praticamente nula: a persistência destes fenômenos não dura — normalmente — mais do alguns dias. Figurativamente, depois de um certo número de dias a natureza “se esquece” de si mesma. A idéia aqui é que muitos fenômenos possuem uma memória “finita”. Nosso trabalho agora é obter uma caracterização matemática razoável do que são processos estocásticos, e do que é a sua “memória”.

Definição: Um processo estocástico $X(t, \omega)$ é uma função que leva cada $\omega \in \Omega$ de um *espaço amostral* em uma função $x(t) = X(t, \omega)$.

Uma figura vale mais do que mil palavras. A figura 8.3 ilustra a definição de processo estocástico. Nesta figura, 3 valores diferentes de ω geram 3 funções $x(t)$ distintas. Mas não se engane! A *estocasticidade* — isto é, a aleatoriedade — do processo está no sorteio dos ω 's, e não na aparência “aleatória” de $x(t)$! De fato, cada um dos 3 gráficos de $X(t, \omega)$ na figura 8.3 poderia ser uma função perfeitamente suave, e ainda assim o “processo” seria estocástico. A aparência de aleatoriedade de cada um dos gráficos está ligada a

propriedades *adicionais* — e muito úteis na prática — de alguns processos estocásticos. Estas propriedades são a memória finita, e a ergodicidade. Nós agora passamos a descrevê-las.

Definição: A função de autocovariância de um processo estocástico $X(t)$ é

$$C(r, s) \equiv \langle (X(r) - \langle X(r) \rangle)(X(s) - \langle X(s) \rangle) \rangle \quad (8.31)$$

Quando $X(t)$ é estacionário, (8.31) simplifica-se consideravelmente. Em processos estacionários, estatísticas calculadas em um instante t qualquer não dependem do mesmo. Então,

$$\langle X(r) \rangle = \langle X(s) \rangle = \mu_X, \quad (8.32)$$

$$\langle (X(t) - \mu_X)^2 \rangle = \sigma_X^2, \quad (8.33)$$

sendo que μ_X e σ_X^2 são *constantes* ao longo do tempo.

Em segundo lugar, se $X(t)$ é estacionário, estatísticas que dependem de *mais de um instante* permanecem inalteradas sob uma translação no tempo. Então,

$$C(r, s) = C(r - s, 0). \quad (8.34)$$

Em outras palavras, *a função de autocovariância de um processo estocástico estacionário depende apenas da diferença entre os instantes r e s* . Isto nos dá, agora, uma definição mais simples (porém também mais restrita, porque só vale para processos estocásticos estacionários) para a função de autocovariância:

$$C(\tau) = \langle (X(t) - \mu_X)(X(t + \tau) - \mu_X) \rangle. \quad (8.35)$$

Com $C(\tau)$ é possível definir uma

Vamos seguir agora para uma introdução informal à Teoria de Difusão Turbulenta de Taylor. Não custa repetir, nossa versão é unidimensional (e portanto demasiadamente simplificada para algumas aplicações).

Vamos agora começar a descrição de uma outra teoria de difusão, que deve muito à de Einstein, e que tentar analisar a difusão turbulenta sob um ponto de vista lagrangeano, ou seja: sob o ponto de vista do passeio aleatório de uma partícula em um fluido em escoamento turbulento.

Da mesma maneira que na seção 8.2, nós vamos estudar o problema em uma dimensão, para simplificar a matemática.

Em primeiro lugar, note que a posição X de uma partícula no instante t pode ser interpretada como uma soma de incrementos, como se segue:

$$X(t) = \int_0^t U(\tau) d\tau = \lim_{\Delta\tau \rightarrow 0} \sum_k U(\tau_k) \Delta\tau.$$

Em (8.36), U é a velocidade da partícula em cada instante, e a integral foi re-interpretada como (o limite de) uma soma de Riemman. É muito razoável agora supor que todos os

$U(\tau_k)$'s são indenticamente distribuídos, e invocar o Teorema Central do Limite. Nós concluímos então, imediatamente, que X é distribuído normalmente.

Nosso ponto de partida — que já foi dado — é a integral (8.36). Vamos supor que o campo de velocidade U que transporta a partícula possui média zero: $\langle U \rangle = 0$. Segue-se de (8.36) que

$$\langle X(t) \rangle = \left\langle \int_0^t U(\tau) d\tau \right\rangle = \int_0^t \langle U(\tau) \rangle d\tau = 0. \quad (8.36)$$

A função de autocovariância simplifica-se ainda mais:

$$C(\tau) = \langle X(t)X(t+\tau) \rangle.$$

Projeto 8.1 Um processo estocástico “clássico” é o modelo AR-2 dado por

$$X_n = a_1 X_{n-1} + a_2 X_{n-2} + E \quad (8.37)$$

onde E é um ruído branco de média zero e variância σ^2 . Os parâmetros a_1 e a_2 devem obedecer a (Bras e Rodríguez-Iturbe, 1993, eq. 2.54):

$$\begin{aligned} a_1 + a_2 &< 1, \\ a_2 - a_1 &< 1, \\ -1 &< a_2 < 1. \end{aligned}$$

Por exemplo, podemos ter $a_1 = 0,8$, e $a_2 = -0,2$.

8.4 2014-03-31 Moving on to stochastic processes and dynamical systems

(Rosenthal, 2008, Teorema 6.1.1): para qualquer função mensurável $g : \mathbb{R} \rightarrow \mathbb{R}$,

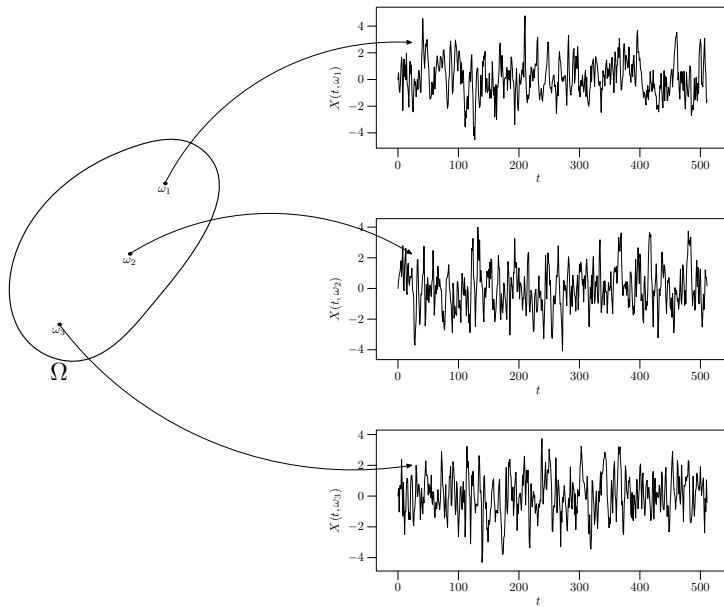
$$\int_{\Omega} g(X(\omega)) dP(\omega) = \int_{\mathbb{R}} g(t) dF_X(t).$$

De Todorovic (1992): Um processo estocástico é uma família de variáveis aleatórias $X(t)$, $t \in T$, definidas em um espaço *comum* de probabilidade (Ω, \mathcal{F}, P) , $T \subset \mathbb{R}$:

$$\begin{aligned} X : \Omega \times T &\rightarrow \mathbb{R}, \\ (\omega, t) &\mapsto x = X(\omega, t) \end{aligned}$$

Um ponto muito importante é o seguinte:

Uma vez que ω tenha sido escolhido, $X(t)$ é *determinístico*, ou seja: é uma função ordinária (de t) como qualquer outra!



A single ω for each $x(t)$!

I will now try to re-write my presentation at Santa Maria using only ω and Ω . The objective is to work on notation to clean it up as much as possible. I believe that notation is important, indeed fundamental: does that make me a formalist?

Vamos construir funções

$$\begin{aligned}\phi_t : \Omega &\rightarrow \Omega, \\ \omega_0 &\mapsto \omega(t) = \phi_t(\omega_0)\end{aligned}$$

O índice t acima é $t \in T \subset \mathbb{R}$. Ele pode representar um sistema discreto ou contínuo.

Já nesse ponto, a similaridade com a definição de processo estocástico é impressionante! Isso não pode ser coincidência!

Defina

$$\phi_{t+s} = \phi_s \circ \phi_t.$$

Um sistema dinâmico agora é a tripla (Ω, \mathcal{F}, P) com ϕ_t . Uma coisa importante pra caramba é que ϕ_t induz uma nova medida de probabilidade em Ω (desde que ϕ_t seja mensurável, é claro). Essa nova medida é definida por

$$P^*(A) \equiv P(\phi_{-t}(A)).$$

O motivo para fazermos a definição “de trás para frente” está, como sabemos, no conjunto de propriedades (6.12)–(6.14).

Definição: Se

$$P(A) = P(\phi_{-t}(A)), \quad \forall A \in \mathcal{F}, \quad (8.38)$$

dizemos que P é *invariante* sob ϕ_t .

From:

http://escuelainvierno.cimfav.cl/documentos/pdf/NotesP_Collet.pdf

Given a probability measure μ (read P) on the phase space Ω and the time evolution map T (or a (semi)-flow) (read ϕ_t), we are **exactly** in the setting of stochastic processes.

This is because, given the dynamical system ϕ_t , and a measurable function g on Ω , we can define a stochastic process $X(t)$ by means of

$$X(\omega, t) \equiv g(\phi_t(\omega)).$$

Minhas observações: nós precisamos entender direitinho por que precisamos, neste ponto, de g para “fazer a ponte” entre sistemas dinâmicos e processos estocásticos

Yes!

Collet (2010):

The fact which may look a little unusual for a Probabilist is that the probability is given on the initial condition, **and there is no randomness appearing in the time evolution** (my emphasis). However the points on the phase space completely characterize the orbits, and we can think of μ (read P) as a probability measure on the orbits, the time evolution being the shift.

In that sense, any stochastic process is a dynamical system (a not very useful remark in practice).

Ergodicity, and Birkhoff's theorem

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T g(\phi_t(\omega_0)) dt = L \stackrel{\text{if ergodic}}{=} \int_{\Omega} g(\omega) d\omega$$

Remark: $X(x_0, t) = g(\phi_t(\omega_0))$ is a single realization of a stochastic process from the initial condition ω_0 . The rightmost expression is *independent of ω_0 , and only then is the process ergodic*.

Moreover, (Lebowitz e Penrose, 1973):

Stated precisely, this means that a system is ergodic on S (read Ω) if and only if *all* regions R of S (read Ω) left invariant by the time evolution, $\phi_t(R) = R$ either have zero area or have an area equal to the area of S (read either have zero probability measure, or have $P(R) = 1$).

Being more accurate,

However, the equality $\phi_t(R) = R$ in Lebowitz e Penrose (1973)'s statement is somewhat *imprecise*! Given the *symmetric difference* between two sets,

$$A \triangle B \equiv (A \cup B) - (A \cap B),$$

then, in terms of probability measure, the corresponding statement is.

$$P(A \triangle \phi_{-t}(A)) = 0 \Leftrightarrow P(A) = 0 \text{ or } P(A) = 1.$$

My simple-minded interpretation: in an ergodic system, the map ϕ_t makes any realization eventually wander over *almost all* of Ω ($P(A) = 1$), or *almost none* of it at all ($P(A) = 0$).

A word of caution on the terminology

(Collet, 2010, p. 40):

Birkhoff's ergodic theorem does not prove ergodicity . It may be a very useful tool in proving ergodicity but as such it does not claim it. In particular it applies in non-ergodic cases. Birkhoff's ergodic applies as soon as you have an invariant measure. You have to work more to decide whether or not the measure is ergodic.

(I want to remove this remark for the final version for Santa Maria)

The mixing condition

implies (is a sufficient condition for) ergodicity:

$$\lim_{\tau \rightarrow \infty} \int_{\Omega} f(x)g(\phi_{\tau}(x)) \, dP(x) = \int_{\Omega} f(x) \, dP(x) \int_{\Omega} g(x) \, dP(x)$$

But

$$C_{fg}(\tau) \equiv \langle f(x)g(\phi_{\tau}(x)) \rangle - \langle f(x) \rangle \langle g(x) \rangle$$

is the τ -covariance function; therefore,

If the τ -covariance function goes to zero, the process is ergodic.

But this does not even guarantee that the process has finite variance. For more regularity, more conditions are needed.

The most common is the existence of integral scales.

Referências Bibliográficas

- Birkhoff, G. D. (1931). Proof of the Ergodic Theorem. *Proceedings of the National Academy of Sciences*, 17(12):656–660.
- Bras, R. L. e Rodríguez-Iturbe, I. (1993). *Random functions and hydrology*. Dover, New York.
- Butkov, E. (1988). *Física matemática*. Guanabara Koogan, Rio de Janeiro.
- Collet, P. (2010). *Dynamical Systems and Stochastic Processes*.
- Dias, N. L. (2003). Obtenção de uma solução analítica da equação de difusão-advecção com decaimento de 1ª ordem pelo método da transformação de similaridade generalizada. *Revista Brasileira de Recursos Hídricos*, 8:181–188.
- Dias, N. L., Gobbi, M. F., e Gobbi, E. F. (2003). Formulação de um modelo matemático do efeito de efluentes térmicos em rios e suas implicações para a legislação ambiental brasileira. *Revista Brasileira de Recursos Hídricos*, 8:169–180.
- Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 322(8):549–560.
- Hinze, J. O. (1975). *Turbulence*. McGraw-Hill Publishing Company, New York.
- Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung (Foundations of the theory of probability)*. Springer, Berlin.
- Kolmogorov, A. N. (1941). The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers (Russian). *Proceedings of the USSR Academy of Sciences*, 30(299–303).
- Kundu, P. K. (1990). *Fluid Mechanics*. Academic Press, San Diego.
- Langevin, P. (1908). Sur la théorie du mouvement brownien. *C. R. Acad. Sci. (Paris)*, 146:530–533.
- Lebowitz, J. L. e Penrose, O. (1973). Modern ergodic theory. *Physics Today*, 26(2):23–29.
- Maxwell, J. C. (1867). On the Dynamical Theory of Gases. *Philosophical Transactions of the Royal Society of London*, 157:pp. 49–88.

- Monin, A. S. e Yaglom, A. M. (1975). *Statistical fluid mechanics: Mechanics of turbulence*, volume 2. MIT Press, Cambridge, Massachusetts.
- Pope, S. B. (2000). *Turbulent Flows*. Cambridge University Press, Cambridge.
- Reynolds, O. (1895). On the dynamical theory of incompressible viscous fluids and the determination of the criterion. *Philos. Trans. R. Soc. Lond. A*, 186:123–164.
- Rosenthal, J. S. (2008). *A first look at rigorous probability theory*. World Scientific.
- Taylor, G. I. (1935a). Statistical theory of turbulence. I. *Proceedings of the Royal Society of London A*, 151:421–444.
- Taylor, G. I. (1935b). Statistical theory of turbulence. II. *Proceedings of the Royal Society of London A*, 151:444–454.
- Taylor, G. I. (1935c). Statistical theory of turbulence. III. *Proceedings of the Royal Society of London A*, 151:455–464.
- Taylor, G. I. (1935d). Statistical theory of turbulence. IV. *Proceedings of the Royal Society of London A*, 151:465–478.
- Taylor, S. J. (1973). *Introduction to measure and integration*. Cambridge University Press.
- Tennekes, H. e Lumley, J. L. (1972). *A first course in turbulence*. The MIT Press, Cambridge, Massachusetts.
- Todorovic, P. (1992). *An introduction to stochastic processes and their applications*. Springer-Verlag, New York.
- von Neumann, J. (1932). Zur Operatorenmethode In Der Klassischen Mechanik. *The Annals of Mathematics*, 33(3):587.