

Processos estocásticos em física:
teoria e fundamentos

Fábio Macêdo Mendes

25 de junho de 2009

Sumário

1	Introdução	10
1.1	Introdução	10
1.2	Estrutura da tese	13
1.3	Probabilidades	13
1.3.1	Uma pequena história	14
1.3.2	Regras para o cálculo de probabilidades	17
1.3.3	Notação	22
1.3.4	Ensembles	24
1.4	Processos estocásticos	26
1.4.1	Condição de compatibilidade de Chapman	27
1.4.2	Definição de Kolmogorov	28
1.4.3	Probabilidades condicionais	30
2	Processos Markovianos	32
2.1	Equação Mestra	35
2.1.1	Equação mestra e processos de saltos	37
2.2	Integrais de trajetória	38
2.2.1	Função Hamiltoniana	42
2.3	Equações dinâmicas	46
2.3.1	Equações de Fokker-Planck	48
2.3.2	Equação mestra e expansão de Kramers-Moyal	49

2.3.3	Ruído de Wiener: formulação Lagrangiana	50
2.4	Conclusão	53
3	Processos de saltos	55
3.1	Processos homogêneos	56
3.1.1	Dinâmica de saltos	58
3.2	Processos analíticos em τ	60
3.2.1	Função de sobrevivência	62
3.2.2	Probabilidade de saltos	65
3.3	Limite de processos de saltos	66
3.4	Conclusão	69
4	Covariância, ergodicidade e difusão	71
4.1	Matrizes de covariância	72
4.1.1	Definição	72
4.1.2	Propriedades matemáticas	74
4.2	Ergodicidade e condição de mistura	84
4.2.1	Ergodicidade fraca	85
4.2.2	Ergodicidade forte	87
4.2.3	Representação da densidade espectral	90
4.3	Difusão	92
4.3.1	Expoentes anômalos	95
4.3.2	Exemplos	99
4.3.3	Difusão em distribuições descontínuas	101
4.4	Conclusão	102
4.4.1	Mistura e ergodicidade	102
4.4.2	Difusão	103
5	Movimento Browniano	104
5.1	Partícula de Stokes	107

5.1.1	Força viscosa	107
5.1.2	Força aleatória	109
5.1.3	Conseqüências matemáticas	112
5.1.4	Relação de flutuação e dissipação	116
5.2	Gás de esferas rígidas	117
5.2.1	Probabilidades $q(u u', n)$	119
5.2.2	Cumulantes	121
5.2.3	Probabilidade de equilíbrio	123
5.2.4	Covariância e difusão	125
5.2.5	Equação de Kramers-Moyal	126
5.2.6	Equação de Langevin	128
5.3	Conclusão	129
6	Equação de Langevin Generalizada	133
6.1	Soluções da equação de Langevin	136
6.1.1	Equação estacionária	136
6.1.2	Equação não-estacionária: Kubo	140
6.1.3	Equação não-estacionária: Volterra	141
6.2	Sistema de osciladores clássicos	143
6.3	Difusão e mistura	148
6.4	Conclusão	151
7	Detecção de tendências em séries temporais	154
7.1	Propriedades de processos Gaussianos	158
7.1.1	Probabilidades condicionais	158
7.1.2	Números aleatórios	161
7.2	Regressão e interpolação	162
7.2.1	Regressão supervisionada	162
7.2.2	Inferência Bayesiana	165

<i>SUMÁRIO</i>	4
7.2.3 Aprendizado não supervisionado	170
7.3 Suavização de curvas	178
7.3.1 Inferência com processos Gaussianos	180
7.3.2 Amostragem contínua: teoria de campos	183
7.4 Conclusão	197
8 Perspectivas e conclusão	198
Apêndice I: Funções especiais	200
Apêndice II: Inferência Bayesiana	203
Apêndice III: Teorema de Doob generalizado	212
Referências Bibliográficas	219

Lista de Figuras

1.1	Diagramas de Venn	17
3.1	Dois processos de saltos com escalas de tempo distintas.	58
7.1	Processo Gaussiano condicional	173
7.2	Otimização do fatores de Bayes.	177
7.3	Fator de Bayes.	178
7.4	Inferência da posição.	190
7.5	Inferência da tendência.	191

Lista de Tabelas

4.2	Funções características de distribuições simétricas	81
4.4	Funções características de distribuições definidas na reta real . .	81

Resumo

Este trabalho formula diversos problemas em física na linguagem de processos estocásticos. Primeiramente, tratamos da representação de processos Markovianos e a relação entre as formulações de equação mestra, integrais funcionais e processos de saltos. Utilizando estes formalismos, se discute a aplicabilidade do modelo de Langevin a sistemas moleculares. Verifica-se que esse modelo é inapropriado para descrever a interação de uma partícula Browniana com um gás de esferas rígidas. A discrepância com relação ao modelo exato é menos acentuada se a massa da partícula Browniana for muito maior que a massa das partículas que formam o gás. A equação de Langevin generalizada, que trata de sistemas não-Markovianos, também é discutida. Novamente, se levantam algumas críticas quanto à sua aplicabilidade a sistemas moleculares. Em especial, mostramos que o teorema de flutuação e dissipação normalmente invocado nesse formalismo pode ser violado em um sistema físico simples.

Finalmente, se formula um método de inferência de tendências em séries temporais utilizando o formalismo de processos Gaussianos. Ao contrário de outros métodos comuns, como as médias móveis, obtemos tanto o valor esperado para tendência quanto o erro associado à esta inferência. O cálculo é feito utilizando o formalismo de integrais de trajetória que evita algumas computações caras normalmente associadas à inferência com processos Gaussianos. Esse procedimento é possível se a taxa de amostragem for alta o suficiente para que seja possível aproximar a série temporal observada por uma trajetória contínua.

Abstract

This work discusses some problems in physics using the language of stochastic processes. Firstly, we discuss the relationships between the formalisms of master equations, path integrals, and jumping processes. Each of them is a different prescription to define a Markovian process. Using these formalisms, we discuss the applicability of the classic Langevin model to molecular systems. We show that this model is inadequate to describe the interaction of a Brownian particle with a gas of hard spheres. The discrepancy with respect to the exact solution is ameliorated if the mass of the Brownian particle is much bigger than the mass of the particles in the gas. The generalized Langevin equation that describes non-Markovian systems is also discussed. Again, we raise a criticism to its relevance in the description of molecular systems. In special, we show a simple physical system that violates the fluctuation dissipation theorem.

Finally, the we present a method of inferring tendencies from time series data. Differently from other common methods, e.g., moving averages, we obtain both the expected value for the tendency and the associated error. The calculation is done using path integrals and avoids some expensive computations commonly associated to other similar methods such as Gaussian process regression. This procedure is feasible if the sampling rate is high enough in order to be possible to approximate the time series data by a continuous function.

Agradecimentos

Aos meus pais, familiares e amigos pelo carinho, inspiração e paciência. À Mari por tudo isso e triplamente pela paciência. Aos mestres e, em especial ao Anníbal, meu orientador, pelos votos de confiança, pela colaboração, pela contribuição nas revisões intermináveis do manuscrito e por todos seus sermões em prol do uso da função característica. Aos tantos outros físicos, matemáticos e probabilistas que desenvolveram e repassaram esse enorme legado cultural.

Um agradecimento especial à Capes pelo apoio financeiro e ao Instituto de Física da UnB, sem os quais esse trabalho não teria sido desenvolvido.

Capítulo 1

Introdução

1.1 Introdução

Este trabalho aborda vários problemas relacionados entre si, mas sem eleger um tema principal e sem possuir um objetivo muito específico. De certo modo, o texto diante do leitor é um esforço de costura: relacionamos vários resultados estabelecidos na teoria de probabilidades com problemas de física e vice-versa. Dado a natureza fragmentada, não há como resumir o objetivo da tese em um único parágrafo. Apresentamos uma lista com os tópicos tratados junto a uma breve descrição da nossa contribuição.

Integral funcional em Processos Markovianos. Toda probabilidade de transição Markoviana possui uma representação como integral funcional, assim como vários outros núcleos que não representam necessariamente uma probabilidade [27]. Nossa contribuição consiste em obter uma condição necessária e suficiente para identificar as integrais funcionais que resultam em probabilidades. Essa condição é uma consequência do teorema de De Finetti para variáveis aleatórias infinitamente divisíveis[45]. Esse tópico é apresentado no *Capítulo 2*.

Processos Markovianos e processos de saltos. Feller mostrou que *todo* processo Markoviano autônomo pode ser representado como um limite a partir de um processo de saltos [16]. Obtemos esse resultado resolvendo explicitamente a equação de Chapman-Kolmogorov. A demonstração apresentada é menos abstrata e fornece melhores indicações sobre quais processos que exibem a propriedade Markoviana. Por outro lado, o resultado é menos geral. No entanto, acreditamos que o método apresentado poderia ser estendido para processos Markovianos de segunda ordem, onde a demonstração de Feller falharia. Esse é o tópico do *Capítulo 3*.

Correlação e difusão. Morgado, et. al, mostraram como caracterizar o expoente de difusão de uma partícula descrita pela equação de Langevin generalizada a partir da densidade espectral da força aleatória[49]. Reformulamos esse problema sem mencionar a força aleatória. Ainda que os resultados se assemelhem, o critério apresentado envolve apenas a densidade espectral do processo, enquanto que o de Morgado utiliza transformadas de Laplace, que geralmente são mais inconvenientes de se manipular. Esse é o tópico do *Capítulo 4*.

Ergodicidade e mistura. Um problema com a literatura de ergodicidade e mistura é a abundância de definições incompatíveis. Ergodicidade e mistura, em suas diferentes acepções, são conceitos relacionados. Demonstramos que certas condições de mistura (uma forte e outra fraca) implicam em ergodicidade (forte e fraca). Os teoremas apresentados são simples e se baseiam em condições fáceis de serem compreendidas. Recentemente, Lee [42]aplicou um teorema devido a Khinchin [33] em, acreditamos, um contexto impróprio e encontrou uma contradição. Posteriormente, comparamos o resultado de Lee com o nosso. Esse tópico também é discutido no *Capítulo 4*.

Equação de Langevin. A equação de Langevin foi formulada para descrever o movimento de partículas mesoscópicas em um fluido. Sua extrapolação para o regime microscópico, talvez seja indevida [17]. Resolvemos um modelo simples que acompanha as colisões em um gás de esferas rígidas. A equação de Langevin é uma aproximação razoável no limite em que a massa da partícula Browniana é muito maior que a das moléculas do fluido. Normalmente considera-se que o coeficiente de difusão é constante. Uma aproximação um pouco melhor, no entanto, seria obtida considerando um coeficiente de difusão que depende da velocidade da partícula Browniana. De qualquer modo, as propriedades de equilíbrio e as correlações lineares do modelo de esferas rígidas e do modelo de Langevin são idênticas. Esse tópico é discutido no *Capítulo 5*.

Equação de Langevin generalizada. A teoria de resposta linear pode ser reescrita como uma equação de Langevin generalizada[38]. Associada a ela, temos o teorema de flutuação e dissipação. Investigamos novamente a aplicabilidade destes resultados a sistemas clássicos microscópicos. Apresentamos um sistema simples onde o teorema de flutuação e dissipação é violado. Esse resultado está baseado em [9], ainda que estes autores não contemplem a possibilidade de violação da relação de flutuação e dissipação no regime clássico. Argumentamos que a condição para que haja violação da flutuação e dissipação é mais comum em sistemas microscópicos que normalmente se assume. Esse é o tópico do *Capítulo 6*.

Processos Gaussianos e detecção de tendências. Processos Gaussianos são utilizados em problemas de interpolação [60] e inferência estatística [57]. Nossa contribuição consiste em um método de detecção de tendências em séries temporais. Os resultados se assemelham ao bem conhecido método de médias móveis, mas com o benefício de existir uma estimativa das incertezas associadas a esta inferência. Consideramos o limite de amostragem contínua, onde o

problema de inferência efetivamente se reduz a uma teoria de campos livre, o que resulta em vários benefícios computacionais. Esse tópico é discutido no *Capítulo 7*.

1.2 Estrutura da tese

A ordem de apresentação foi escolhida de tal maneira que os capítulos se organizam em uma sucessão lógica, onde o problema de um capítulo é sempre abordado de alguma maneira em um capítulo anterior. No entanto, quase todos os capítulos são razoavelmente independentes entre si. As exceções são o capítulo 4 que serve de base para os capítulos 6 e 7, e ainda o capítulo 5, que depende de alguns resultados presentes no capítulo 3.

Iniciamos a tese com uma introdução relativamente longa sobre a teoria de probabilidades e processos estocásticos. Seu objetivo é duplo. Primeiro, queremos uniformizar a notação. Fora isto, existe a intenção de apresentar algumas noções de probabilidades Bayesianas e da teoria de processos estocásticos de Kolmogorov. Estes temas raramente são tratados com o devido cuidado em livros-texto de física. Duas referências particularmente úteis que influenciaram o presente trabalho são o livro de Jaynes sobre probabilidades [26] e o de Van Kampen sobre processos estocásticos [32].

1.3 Probabilidades

As probabilidades não existem
(De Finetti)

Apresentamos uma versão muito resumida da história da teoria de probabilidades e da noção Bayesiana de probabilidade como uma medida de crença em uma possibilidade.

1.3.1 Uma pequena história

Ainda que os jogos de azar normalmente estejam associados a histórias de ruína pessoal, há vários casos em que relação entre matemáticos e jogadores foi bastante profícua. O grande produto desta relação foi a teoria da probabilidades, que é o tema a ser apresentado agora. O primeiro tratamento matemático dado aos jogos de azar foi realizado pelo pensador italiano Girolamo Cardano, essencialmente para uso próprio, já que era um jogador notório. No entanto, uma discussão organizada só se iniciou um século depois, novamente motivada por jogos de azar: um cavalheiro francês, Antoine Gombaud, que entre outras coisas era um matemático amador, jogador e filósofo de botequim, propôs uma discussão sobre como determinar a aposta justa em alguns jogos de azar que chamou a atenção de Pascal e Fermat.

O resultado do diálogo entre os dois matemáticos estabeleceu as bases da teoria de probabilidades e da análise combinatória. A partir daí, os jogos de azar passaram a ser um tema recorrente na discussão de outros matemáticos respeitáveis. Jakob Bernoulli escreveu um livro influente chamado *Ars Conjectandi* (Arte da conjectura) que apresentava importantes contribuições, como por exemplo a lei dos grandes números. Para Bernoulli, a teoria de probabilidades oferece um método geral de inferência.

Esse ponto de vista foi posteriormente desenvolvido pelo marquês de Laplace. Laplace começa seu *Ensaio Filosófico Sobre as Probabilidades* com uma discussão sobre como a noção de azar se concilia com o mundo determinístico implícito nas leis de Newton. Ainda que para os padrões modernos sua discussão sobre determinismo seja obsoleta, a conclusão básica sobre o papel das probabilidades na ciência e vida humana, em geral, se mantêm. Para Laplace, “filosofia sã não vê [no azar] nada mais que a expressão de nossa ignorância a respeito das verdadeiras causas”.

A interpretação de probabilidade como uma medida subjetiva de ignorância

é hoje associada à escola Bayesiana de estatística. Ainda que qualquer menção ao subjetivismo possa causar desconforto em certos círculos, a interpretação Bayesiana é, na realidade, extremamente leniente. Ela se adéqua tanto a usos sérios, como em ciência, quanto a outros usos menos respeitáveis, como a prova estatística da existência de Deus, dada pelo próprio reverendo Thomas Bayes. Deve-se reconhecer que as probabilidades *podem* ser utilizadas para descrever *o grau de crença em hipóteses*, como em “*existe 10% de chance do Brasil ganhar a próxima copa do mundo*”. A questão é se probabilidades deste tipo podem ser admitidas na prática científica.

Após a contribuição fundamental de Maxwell e Boltzmann e posteriormente de Born na mecânica quântica, é inadmissível pensar na física moderna sem o conceito de probabilidade. Mas, exatamente, que papel as probabilidades cumprem numa teoria física? Certamente, o ensemble canônico não é simplesmente o grau de ignorância de Gibbs sobre a posição de um determinado sistema físico no espaço de fase (e, analogamente, o ensemble microcanônico não seria o grau de ignorância de Boltzmann): existe algo mais forte nas atribuições de probabilidade que a simples opinião de um ou outro cientista. O físico americano Edwin T. Jaynes foi pioneiro em trazer os métodos Bayesianos para a física e enfrentou várias destas questões [24]. Aqui fazemos apenas alguns comentários:

- Não queremos que a descrição sobre a natureza esteja sujeita aos capricho das mentes dos cientistas: a ciência se pretende ser uma descrição objetiva da realidade. Existem maneiras bem definidas de escolher atribuições de probabilidade com base na evidência disponível. Subjetividade não significa aceitar qualquer atribuição de probabilidades leviana

¹.

¹Ex.: Jaynes interpreta o método da máxima entropia de Gibbs como a prescrição que gera a distribuição menos enviesada entre todas as possíveis atribuições de probabilidade. Essa interpretação se baseia na noção estatística de entropia devida à Shannon. Mas porque,

- A lei dos grandes números não faz sentido se as probabilidades são interpretadas como frequências: afinal, como é possível falar em convergência para frequências se as probabilidades são *definidas* como frequências? A interpretação Bayesiana eliminaria a circularidade nesse argumento.
- É difícil definir a noção de *aleatoriedade intrínseca*, que é outra alternativa comum à interpretação Bayesiana. Um sistema intrinsecamente aleatório também será epistemologicamente aleatório no sentido que não poderíamos determinar seu estado com certeza. Existe uma questão exotérica sobre se probabilidades que *nós* atribuímos seriam as *verdadeiras probabilidades do mundo*, o que na perspectiva Bayesiana corresponde a um falso problema. No entanto, a maioria dos pretensos exemplos de aleatoriedade intrínseca podem ser caracterizados como o que Jaynes chama de *falácia da projeção mental*: quando “se supõe que as criações da própria imaginação são propriedades reais da Natureza, e que a própria ignorância significa algum tipo de indecisão por parte da Natureza”.

para obter o ensemble canônico, é necessário utilizar somente a energia como vínculo? A resposta não está na dinâmica microscópica do sistema: não importa o caos ou ergodicidade. Fundamentalmente, quem dita os vínculos é a termodinâmica: os sistemas termodinâmicos mais simples são aqueles definidos apenas pela energia e entropia. Segundo Jaynes, devemos utilizar *toda* a informação macroscópica disponível e *somente* esta (por informação macroscópica entenda-se as variáveis termodinâmicas que, de fato, se consegue medir e controlar). Teorias macroscópicas mais complicadas como a hidrodinâmica surgem da aplicação do método de máxima entropia com vínculos mais sofisticados [28].

1.3.2 Regras para o cálculo de probabilidades

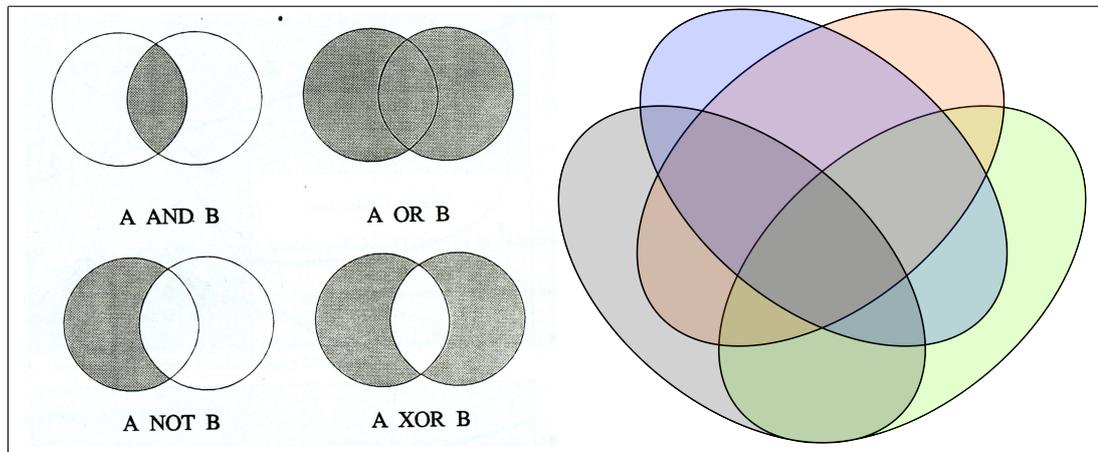


Figura 1.1: Diagramas de Venn: ao lado esquerdo para dois conjuntos e ao lado direito mostra uma representação com quatro conjuntos.

As regras para a atribuição de probabilidades podem ser expressas de maneira intuitiva pelos diagramas de Venn (Fig. [1.1]), pensando nas probabilidades como as “áreas” circunscritas em cada conjunto A , B , $A \text{ AND } B$, $A \text{ OR } B$, etc (AND, OR, XOR, NOT, representam operadores Booleanos.) Os diagramas de Venn ficam rapidamente complicados na medida que se aumenta o número de conjuntos, o que implica que geralmente usaremos uma versão algébrica destas operações. Introduzimos a notação compacta $AB = A \text{ AND } B$, $A + B = A \text{ OR } B$ e $\bar{A} = \text{NOT } A$ que utiliza o fato que as regras para as operações de AND e OR mimetizam respectivamente a multiplicação e a adição. O fato que estas operações sobre conjuntos definem uma álgebra é explorado por matemáticos, que a chamam de álgebra- σ .

Do ponto de vista formal, uma probabilidade é uma mensuração sobre uma álgebra- σ (álgebra de conjuntos). Da Fig. 1.1 podemos inferir as seguintes regras para a manipulação de probabilidades

$$P(A + B) = P(A) + P(B) - P(AB), \quad (1.1)$$

$$P(\bar{A}) = 1 - P(A). \quad (1.2)$$

Adicionalmente, postulamos que a mensuração é finita, atribuindo convenientemente o valor $P(A + B) = 1$, para a probabilidade do conjunto “universo”. O conjunto de todas as possibilidades é denominado de espaço amostral, que possui probabilidade 1. O conjunto de todos os subconjuntos, no nosso caso, $\emptyset, A, B, \bar{A}, \bar{B}, A + B, AB$ e $A + B - AB$ forma a álgebra- σ , que é fechada nas operações Booleanas AND, OR, XOR, NOT, etc.

As idéias anteriores capturam formalmente as propriedades de uma probabilidade, mas nada dizem sobre o que elas significam, ou em que contexto aplicá-las. E assim também ocorre em qualquer outra teoria matemática.

Os diagramas de Venn induzem uma interpretação imediata a partir da noção elementar de conjunto como uma coleção de objetos. Nesse caso, é natural identificar as probabilidades como o número de objetos de cada tipo dentro de cada conjunto. Ou melhor, dada a condição de normalização, a probabilidade seria a fração de objetos em cada conjunto com relação ao número total de elementos. Deste ponto de vista, a probabilidade estaria limitada à lidar com coleções de objetos ou de eventos. Tais eventos poderiam representar os vários resultados de um experimento aleatório repetido inúmeras vezes, o que corresponde à interpretação freqüentista das probabilidades.

A concepção de probabilidades como freqüências é menos útil que aparenta. Normalmente se invoca a lei dos grandes números que diz que em uma série infinitamente longa de resultados, as freqüências estatísticas convergem para a probabilidade, com probabilidade 1. Para evitar a óbvia circularidade no argumento, é necessário um certo contorcionismo epistemológico. Várias

justificativas funcionam mais ou menos assim: a probabilidade representaria a distribuição de elementos sobre um conjunto imaginário, possivelmente infinito, de possíveis realizações. A lei dos grandes números então garantiria a convergência das freqüências do experimento real para aquelas do conjunto imaginário.

Esse argumento, no entanto, deve utilizado com parcimônia: o objetivo da concepção freqüentista é evitar o subjetivismo, ou probabilidades entendidas como “ignorância a respeito das verdadeiras causas”. Isso é violado descaradamente ao invocar a noção de um *conjunto imaginário* de realizações, ou qualquer objeto do gênero. A concepção freqüentista não consegue lidar com problemas simples como, porque atribuímos uma probabilidade *a priori* de $1/6$ a cada uma das faces de um dado: a rigor seria necessário conduzir um experimento de freqüências para determinar essa questão, apesar de quase ninguém o fazer. É imediatamente óbvio, olhando a Fig. 1.1, que o formalismo de probabilidades se presta ao cálculo de freqüências. No entanto, é muito menos claro que esse formalismo é uma ferramenta adequada para lidar com “graus de crença” e ignorância. Essa situação foi contornada com o teorema de Cox, que fundamenta matematicamente a noção subjetivista de probabilidade como sendo a única maneira consistente de manipular “graus de crença” representados por números reais [10].

Cox define “plausibilidade” (ou “grau de crença”) como uma relação entre duas afirmações lógicas:

$$(A|H) = \text{“o quanto } A \text{ é plausível assumindo a hipótese } H\text{.”} \quad (1.3)$$

Outra interpretação possível é: “em que grau a hipótese H implica A ”. Uma vez que se aceite que é possível representar numericamente a noção de plausibilidade, as regras usuais para o cálculo de probabilidades seguem do teorema

de Cox. A interpretação de Cox generaliza a noção elementar de probabilidades como frequências e se aplica a qualquer situação que possa ser formulada como graus de crença atribuídos a conjuntos ou a afirmações lógicas.

Vale observar que, no argumento de Cox, $P(A|H)$ é a probabilidade condicional de A dado H , que normalmente seria definida como

$$P(A|H) = \frac{P(AH)}{P(H)}. \quad (1.4)$$

A interpretação frequentista é que $P(A|H)$ representa a fração dos elementos de H que também pertencem a A . No entanto, para Cox, toda probabilidade é condicional, e ao invés de definir $P(A|H)$ a partir de $P(AH)$, ele faz o contrário. O resultado é a chamada regra de Bayes

$$P(AB) = P(A)P(B|A) = P(B)P(A|B), \quad (1.5)$$

que foi obtida pela primeira vez pelo reverendo Thomas Bayes a partir de um argumento muito menos trivial.

Em vários casos, a hipótese fundamental H é identificada como toda a informação prévia que se têm sobre um determinado modelo estatístico. Por ser essencialmente vaga e raramente aparecer explicitamente nas contas, é comum omitir H (como fizemos em [1.5]). Existe pelo menos um caso onde o reconhecimento que existem hipóteses *a priori* é importante também como ferramenta de cálculo, mas isto será tratado adiante quando discutirmos os chamados fatores de Bayes para a comparação de modelos estatísticos.

O caso de maior interesse para a aplicação da teoria de probabilidades corresponde à situação em que existem várias afirmações A_i ; $i = [1, m]$ em que uma e apenas uma é verdadeira. Um exemplo simples é o sistema de dois

níveis definido abaixo

$$A_1 \equiv \text{“a moeda se encontra no estado } \textit{cara}\text{”} \quad (1.6)$$

$$A_2 \equiv \text{“a moeda se encontra no estado } \textit{coroa}\text{”}. \quad (1.7)$$

Naturalmente, ambas afirmações não podem ser verdadeiras simultaneamente, ou seja,

$$A_1 A_2 = \text{falso} \Rightarrow P(A_1 A_2 | H) = 0. \quad (1.8)$$

Além disto, pelo menos uma das afirmações, A_1 ou A_2 , deve ser verdadeira em qualquer circunstância, de forma que

$$A_1 + A_2 = \text{verdadeiro} \Rightarrow P(A_1 + A_2 | H) = 1. \quad (1.9)$$

Juntando (1.1), com (1.8) e (1.9) obtemos a regra de normalização

$$P(A_1) + P(A_2) = 1. \quad (1.10)$$

A generalização para qualquer conjunto $\{A_i\}$ de afirmações exaustivas e mutualmente exclusivas é imediata.

De maneira geral, considere dois sistemas de afirmações $\{A_i\}$ e $\{B_j\}$ exaustivos e mutualmente exclusivos — ou seja, conjuntos de afirmações em que sempre uma e apenas das afirmações é verdadeira. A condição de normalização é dada abaixo

$$\underbrace{P\left(\sum_i A_i \mid H\right)}_{\text{Verdade}} = \sum_i P(A_i | H) = 1, \quad (1.11)$$

que também se aplica às probabilidades condicionais

$$\sum_i P(A_i|B_jH) = 1, \quad (1.12)$$

qualquer que seja B_j . Destes resultados, deriva-se a regra de normalização para sistemas compostos, ou seja,

$$\sum_{ij} P(A_iB_j|H) = \underbrace{\sum_i P(A_i|H)}_1 \underbrace{\sum_j P(B_j|A_iH)}_1 = 1, \quad (1.13)$$

e a importante regra de marginalização

$$\sum_j P(A_iB_j|H) = P(A_i|H) \sum_j P(B_j|A_iH) = P(A_i|H). \quad (1.14)$$

Estes são os resultados básicos da teoria. De agora em diante nos referimos a *distribuição de probabilidade* toda atribuição feita sobre um conjunto de afirmações exaustiva e mutuamente exclusivas. As regras expostas anteriormente também permitem generalizar o conceito de probabilidade para distribuições contínuas como será feito na próxima seção.

1.3.3 Notação

A notação $P(A|H)$ representa a probabilidade do evento A dada a hipótese H . Normalmente, estes eventos são descritos por índices, que podem ser contínuos ou discretos, como por exemplo a A = “partícula está na posição x no instante t ”. Para deixar a notação mais acessível, permitimos que as afirmações lógicas sejam representadas diretamente por estes índices e, ao invés de escrever A , ou $A(x, t)$ escrevemos diretamente $P(x_t|H)$.

Se o argumento da probabilidade for, por exemplo $AB = A \text{ AND } B$, é útil substituir o operador lógico AND por uma vírgula, para separar mais

claramente as várias hipóteses. Assim, uma probabilidade conjunta destas se escreveria como

$$P(AB|H) = P(A \text{ AND } B|H) = P(A, B|H). \quad (1.15)$$

Em muitos casos, a vírgula torna o argumento mais legível, mas dá a falsa impressão que $P(A, B|H)$ é uma função de três variáveis A , B e H . A probabilidade possui apenas dois argumentos: o condicionante, que no caso é H e argumento/conjunto principal, que aqui é $A \text{ AND } B$.

É lógico que, se as afirmações A e B dependerem de variáveis adicionais (digamos, $A(x)$ e $B(y)$), então a probabilidade $P(A, B|H)$ também será uma função de x e y , que normalmente escrevemos com uma letra maiúscula: $P(AB|H) = F(x, y)$. Caso x e y sejam variáveis contínuas, convém explicitar a densidade de probabilidades escrevendo $P(AB|H) = f(x, y)dx dy$. Essa notação é explícita sobre as variáveis desconhecidas denotadas pelo diferencial $dx dy$ e possui o benefício de cair imediatamente numa integral quando, por exemplo, escrevemos

$$\sum_x P(A|H) = \sum_x P(x|H) = \sum_x dx f(x) = \int dx f(x), \quad (1.16)$$

onde a passagem do segundo para o terceiro termo evoca a definição de integral de Riemann, devidamente descontada a falta de rigor matemático.

Apresentamos algumas notações úteis que serão muito utilizadas no restante do texto. Seja um problema descrito por uma probabilidade contínua $f(x)dx$. Representamos a média de uma função $g(x)$ como

$$\langle g(x) \rangle = \int dx f(x) g(x). \quad (1.17)$$

Também é conveniente introduzir uma notação para a variável centralizada.

Deste modo, definimos

$$\langle\langle g(x) \rangle\rangle = \langle g(x - \langle x \rangle) \rangle = \int dx f(x) g(x - \langle x \rangle). \quad (1.18)$$

A segunda notação é conveniente para representar os momentos centralizados.

A variância, por exemplo, se escreveria simplesmente como

$$\langle\langle x^2 \rangle\rangle = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2. \quad (1.19)$$

Vale ressaltar, novamente, que as médias e desvio padrão são números que capturam informações *parciais* sobre a distribuição de probabilidades. Elas são úteis para embasar decisões que levem em conta o risco em um modelo probabilístico (comprar ou vender um ativo?). No entanto, distribuições muito diferentes entre si ainda podem possuir o mesmo valor para a média, covariância e outras estimativas pontuais.

1.3.4 Ensembles

A lei dos grandes números garante que a frequência estatística converge para a probabilidade no limite de infinitas realizações de um experimento aleatório. (Essa afirmação é bastante mais delicada em um espaço amostral contínuo). Deste modo, é possível representar uma distribuição de probabilidades como um conjunto muito grande (potencialmente infinito) de cópias mentais distribuídas de uma maneira apropriada entre os diferentes estados do espaço amostral. Esse conjunto de elementos é chamado de *ensemble estatístico*.

Ainda que não exista nada de fundamentalmente errado com o raciocínio de ensemble, vale a pena mencionar que ele induz a um erro muito comum, ainda que sutil. Os físicos tratados em física estatística são, via de regra, formados por um conjunto enorme de partículas. É razoável pensar no próprio sistema físico como um *ensemble estatístico*. Esse *ensemble* descreveria a gran-

deza física “densidade de partículas” que, quando devidamente normalizada, é também uma probabilidade. No entanto, o *ensemble* é muitas vezes igualado à probabilidade a uma partícula, $P(q, p)$, que é obtida integrando a distribuição de probabilidades no espaço de fase sobre todas as outras partículas. A rigor, esta identificação é incorreta. Dado que as partículas interagem entre si, não seria razoável considerá-las estatisticamente independentes. Havendo correlações, a lei dos grandes números não pode ser invocada para igualar $P(q, p)$ com a densidade de partículas observada experimentalmente [32].

O dilema apontado por esta questão é que, se por um lado $P(q, p)$ é mais facilmente acessível à teoria (ex.: quebras da hierarquia BBGKY), por outro lado a densidade de partículas é a grandeza experimentalmente mais relevante. De fato, há vários modelos obtêm uma equação cinética assumindo algum tipo de independência estatística entre as diferentes partículas (equação de Boltzmann ou Vlasov). Isto restaura a interpretação correta do *ensemble estatístico* já que lidaríamos com um sistema de partículas que, para efeitos práticos, poderiam ser tratadas como se fossem independentes.

A densidade de partículas ser uma grandeza física bem definida não implica que só existe uma atribuição “correta” de probabilidades $P(q, p)$. Uma atribuição razoável de probabilidade depende do grau de detalhamento que se consegue descrever o sistema. Na situação idealizada onde se conhece as velocidades e posições de todas as partículas, $P(q, p)$ corresponderia a uma mistura de deltas de Dirac. O outro extremo, onde quase nada se sabe sobre o sistema fora seu estado termodinâmico, poderia ser abordado utilizando o método da máxima entropia de Gibbs [24, 25]. Nenhuma das distribuições é mais certa ou errada que a outra: elas apenas refletem um tipo de informação diferente disponível sobre o sistema.

1.4 Processos estocásticos

A grosso modo, um processo estocástico é a generalização do conceito de número aleatório para uma função aleatória. Uma parte considerável da literatura em processos estocásticos lida com funções que dependem de um único parâmetro real, que normalmente identificamos com o tempo. Todos os problemas considerados nesse trabalho são deste tipo, onde a função desconhecida usualmente representa a posição de uma partícula ou sua velocidade como funções do tempo. No entanto, é necessário mencionar que existem situações em que os parâmetros são vetores em \mathbb{R}^n ou mesmo elementos de espaços mais exóticos.

Uma função arbitrária $y(t)$ sobre um parâmetro real se assemelha em diversos aspectos a um vetor. Para enfatizar isto, também utilizamos a notação $y(t) = y_t$ onde o tempo t seria interpretado como um índice contínuo. Atribuímos uma probabilidade a um vetor definindo uma distribuição conjunta sobre cada uma das suas componentes. Por exemplo, a probabilidade de um vetor $\mathbf{v} = (v_1, v_2, \dots, v_N)$ é dada pela distribuição conjunta $P(\mathbf{v}) = P(v_1, v_2, \dots, v_N)$.

É lógico que não é possível seguir o mesmo raciocínio utilizado para vetores para atribuir uma probabilidade a uma função. Trata-se de um objeto com um número infinito e não-enumerável de componentes. É sensato, portanto, começar por um problema mais modesto. Ao invés de falar sobre a distribuição de probabilidade completa para a função $y(t)$ poderíamos definir a probabilidade que o valor de $y(t)$ em t_1 seja igual a y_1 : isto fornece uma informação parcial sobre a função desconhecida. Escrevemos esta probabilidade explicitamente como $P(y_1) = f_{t_1}(y_1)dy_1$.

Esse raciocínio pode ser levado adiante: definimos uma malha temporal com os valores t_1, t_2, \dots, t_N e assim consideramos a probabilidade que a função

$y(t)$ assumo o valor y_i em cada instante t_i como sendo

$$P(y_1, y_2, \dots, y_N) = f_{t_1, t_2, \dots, t_N}(y_1, y_2, \dots, y_N) dy_1 dy_2 \dots dy_N. \quad (1.20)$$

Em algumas situações é mais conveniente utilizar a notação mais explícita,

$$P(y_1, \dots, y_N) = f(y_1, t_1; \dots; y_N, t_N) dy_1 \dots dy_N, \quad (1.21)$$

$$= f(y_1, \dots, y_N | t_1 \dots t_N) dy_1 \dots dy_N, \quad (1.22)$$

mas usualmente favoreceremos a primeira. A probabilidade $P(y_1, \dots, y_N)$ definida sobre a malha temporal t_1, \dots, t_N é uma probabilidade a N -pontos para $y(t)$. Dito isto, poderíamos partir de (1.20) e definir as probabilidades a $N+1$, ou $N-1$ pontos: um processo estocástico é simplesmente uma maneira consistente de definir estas probabilidades.

1.4.1 Condição de compatibilidade de Chapman

É importante garantir que as atribuições de probabilidades sobre diferentes malhas temporais sejam consistentes entre si. Considere, por exemplo, a malha t_1, \dots, t_N e uma segunda malha onde acrescentamos um ponto t_{N+1} . Pela lei de marginalização das probabilidades, é necessário que

$$P(y_1 \dots y_N) = \sum_{y_{N+1}} P(y_1 \dots y_N y_{N+1}). \quad (1.23)$$

Ou seja, a integração sobre um ponto adicional recupera a atribuição sobre a malha original. Podemos generalizar esta equação de diversas maneiras utilizando a idéia que retirar um ponto da malha temporal deve equivaler a integrar sobre a variável y_i na distribuição de probabilidades.

O conjunto de todas estas condições sobre as probabilidades a N -pontos de um processo estocástico é conhecido como as condições de compatibilidade

de Chapman. Podemos construir as condições relevantes na forma de uma hierarquia de equações de compatibilidade,

$$P(y_1) = \sum_{y_2} P(y_1 y_2), \quad (1.24)$$

$$P(y_1 y_2) = \sum_{y_3} P(y_1 y_2 y_3), \quad (1.25)$$

$$\vdots = \quad \vdots \quad (1.26)$$

$$P(y_1 \dots y_N) = \sum_{y_{N+1}} P(y_1 \dots y_N y_{N+1}). \quad (1.27)$$

A condição de compatibilidade é o requerimento trivial de que as distribuições conjuntas sobre uma sub-malha de pontos são calculadas pela integração sobre as variáveis indesejadas.

Ainda que muitas vezes seja razoavelmente simples definir as probabilidades a um 1 ou 2 pontos de maneira consistente, satisfazer a hierarquia (1.27) para qualquer valor arbitrário de N costuma ser um problema matemático complicado: $P(y_1, \dots, y_N)$ ser relaciona de uma maneira não-trivial com $P(y_1, \dots, y_{N+1})$ pela equação integral (1.27). Duas situações comuns onde é possível garantir a condição de compatibilidade serão abordadas nesse trabalho. A primeira, os processos Markovianos, é tratada nos capítulos 2 e 3 enquanto que o capítulo 7 é dedicado aos processos Gaussianos. Vale ressaltar que existem outras maneiras de definir processos estocásticos válidos, tais como as cópulas estocásticas [51] e o funcional característico [39].

1.4.2 Definição de Kolmogorov

Em muitas situações, a aleatoriedade das funções $y(t)$ decorre do desconhecimento de alguns parâmetros que controlam a função $y(t)$. Em muitos problemas de física, os parâmetros desconhecidos correspondem às condições iniciais de um determinado problema. Para ilustrar isto, montamos um processo es-

tocástico extremamente simples baseado nesta idéia.

Considere que as funções desconhecidas $y(t)$ são linhas retas e a aleatoriedade resulta apenas dos parâmetros θ_1 e θ_2 na relação $y(t) = \theta_1 + \theta_2 t$ serem desconhecidos. Seja a probabilidade conjunta destes parâmetros dada por $g(\theta_1, \theta_2)d\theta_1 d\theta_2$. Para esse processo, podemos escrever as probabilidades a N pontos de maneira simples

$$f_t(y) = \int d\theta_1 d\theta_2 g(\theta_1, \theta_2) \delta(y - \theta_1 - \theta_2 t), \quad (1.28)$$

$$f_{tt'}(y, y') = \int d\theta_1 d\theta_2 g(\theta_1, \theta_2) \delta(y - \theta_1 - \theta_2 t) \delta(y' - \theta_1 - \theta_2 t'), \quad (1.29)$$

e assim por diante.

No âmbito da física clássica, é possível afirmar que quase todos os processos estocásticos de interesse são no fundo uma variação mais complicada do exemplo anterior. Após integrar as equações de movimento, qualquer variável dinâmica $y(t)$ seria uma relação determinística (e extremamente complicada) envolvendo o tempo t e o microestado inicial. A aleatoriedade presente nestes sistemas resulta do desconhecimento da condição inicial.

Podemos generalizar o argumento que levou à (1.28-1.29) definindo de maneira mais abstrata

$$y(t) = f(\theta, t), \quad (1.30)$$

onde θ representa um conjunto enumerável (pode ser infinito) de variáveis aleatórias e $f(\theta, t)$ é uma função arbitrária. É óbvio que um processo definido como (1.30) satisfaz as condições de compatibilidade de Chapman. Kolmogorov mostrou que todo processo que satisfaz as condições de compatibilidade também pode ser escrito como (1.30)[32]. Ainda que sua demonstração estabeleça a equivalência entre as duas definições, nem sempre é possível reconstruir a função $f(\theta, t)$ a partir da hierarquia de Chapman, ou mesmo identificar o espaço de parâmetros aleatórios θ .

1.4.3 Probabilidades condicionais

A noção de *processo* surge da maneira como normalmente se fatora a distribuição de probabilidades a N -pontos. Sejam y_1, \dots, y_N os valores de $y(t)$ nos pontos t_1, \dots, t_N . É sempre possível escrever $P(y_1, \dots, y_N)$ como

$$P(y_1, \dots, y_N) = P(y_1)P(y_2|y_1)P(y_3|y_2y_1) \dots P(y_N|y_{N-1} \dots y_1), \quad (1.31)$$

onde as probabilidades $P(y_i|y_{i-1} \dots, y_1)$ são as probabilidades de y_i condicionadas aos valores conhecidos de y_{i-1}, \dots, y_1 . Partindo de $P(y_1)$, define-se $P(y_1, y_2)$ utilizando a condicional $P(y_2|y_1)$; partindo de $P(y_1, y_2)$ define-se $P(y_1, y_2, y_3)$ utilizando $P(y_3|y_2, y_1)$, e assim por diante. O *processo estocástico* pode então ser concebido como uma algoritmo que determina as condicionais $P(y_i|y_{i-1} \dots, y_1)$ que por sua vez definem a probabilidade sobre qualquer malha t_1, \dots, t_N .

Observe que (1.27) deve valer mesmo quando os argumentos y_1, y_2, \dots, y_N não forem ordenados temporalmente. De fato, ainda que a notação utilizada induza a supor um ordenamento temporal, isto não é necessário e as relações acima valem mesmo quando $t_1 \not\prec t_2 \dots \not\prec t_N$. Existem vários processos que se apresentam de maneiras mais simples quando há ordenamento. Por isso, além da clareza de notação, utilizaremos esta convenção a partir de agora. Enfatizamos que o ordenamento é apenas uma convenção que em determinados contextos sequer faz sentido (e.g., o argumento da função pode ser um objeto como um vetor, um número complexo, uma lista, uma palavra, etc).

Satisfazer a hierarquia de Chapman é um problema fundamental na definição de qualquer processo estocástico. Por este motivo, consideraremos processos Markovianos e posteriormente Gaussianos: esses são os dois exemplos mais

imediatos onde é possível mostrar que a hierarquia de Chapman é satisfeita. Muitos autores calculam apenas a evolução da probabilidade marginal e sequer tratam da definição do processo estocástico subjacente. Isto talvez seja satisfatório em algumas aplicações. No entanto, toda informação sobre como as trajetórias estocásticas se comportam é perdida neste tipo de abordagem. Acreditamos que esta é uma parte importante na avaliação da qualidade de qualquer modelo estocástico e por isso tentamos explicitar o processo estocástico completo sempre que isto for possível.

Capítulo 2

Processos Markovianos

Na formulação do processo Markoviano por integrais de trajetória, a probabilidade de transição de x' em t' para x em t é escrita como a integral funcional [27]

$$G_{tt'}(x|x') = \int Dx Dz e^{\int dt [H(z,x,t) - i z \dot{x}]}$$

O resultado desta integral também deve corresponder à uma distribuição de probabilidades, mas isto não é garantido pela expressão acima. A questão principal abordada nesse capítulo é a de determinar quais “Hamiltonianas” $H(z, x, t)$ gerariam processos estocásticos legítimos e quais são as suas principais características. Apresentamos um critério necessário e suficiente a partir do teorema de De Finetti [45] para variáveis aleatórias infinitamente divisíveis. A equação mestra do processo Markoviano é obtida relacionando-a com a representação por integral de trajetórias. A classe de Hamiltonianas mais simples que existe está associada a distribuições estáveis gera uma equação de Fokker-Planck generalizada já conhecida na literatura [67].

Um processo é chamado Markoviano se, dado o estado presente, a probabilidade de um estado posterior for independente do passado. Em linguagem

técnico, isto equivale à imposição sobre as probabilidades de transição:

$$P(x_{N+1}|x_N, x_{N-1}, \dots, x_1) = P(x_{N+1}|x_N), \quad (2.1)$$

onde assumimos que $t_1 < t_2 < \dots < t_N$. Em um processo Markoviano, basta especificar duas distribuições, $P(x_t)$ e $P(x_t|x_{t'})$, para caracterizar qualquer probabilidade a N -pontos, como abaixo,

$$P(x_1, \dots, x_N) = P(x_1)P(x_2|x_1)P(x_3|x_2) \dots P(x_N|x_{N-1}). \quad (2.2)$$

Por simples que aparentem, existe uma dificuldade técnica para se definir um processo Markoviano legítimo: nem toda probabilidade de transição $P(x_{N+1}|x_N)$ respeita automaticamente a condição de compatibilidade. Escrevemos a probabilidade de transição entre dois instantes arbitrários t_1 e t_3 como $P(x_3|x_1) = \sum_{x_2} P(x_3, x_2|x_1)$. Fatorando o lado direito da equação e utilizando a propriedade Markoviana 2.1 chegamos à

$$P(x_3|x_1) = \sum_{x_2} P(x_3|x_2)P(x_2|x_1), \quad (2.3)$$

que é conhecida como equação de Chapman-Kolmogorov (ou equação de Smoluchowski) [32]. Infelizmente, encontrar todas as soluções para a Eq. (2.3) é uma tarefa complicada.

O método mais utilizado para obter um processo Markoviano legítimo é definir a probabilidade de transição como a solução de uma equação mestra. Relacionado a esse método, temos a equação de Fokker-Planck, que é uma forma que equação mestra assume em modelos com ruído contínuo. Apesar da importância desse formalismo, o foco deste capítulo é a caracterização do processo Markoviano por integrais de trajetória.

A integral de trajetória (ou integral funcional) foi introduzida por Wiener,

em 1921, no estudo do movimento Browniano. No entanto, a popularização dos métodos de integral funcional em física se deu após o trabalho de Feynman em mecânica quântica. Posteriormente, estes mesmos métodos voltaram a ser utilizados por físicos para lidar com processos estocásticos. Muitas técnicas desenvolvidas na mecânica quântica se traduzem para o contexto de processos estocásticos devido às semelhanças entre a equação de Schrödinger e a equação de Fokker-Planck. Essa semelhança foi explorada por vários autores para desenvolver a chamada formulação Lagrangiana para a equação de Fokker-Planck [34, 23, 35, 27]. Em um contexto semelhante, Masao Doi utilizou os métodos de segunda quantização para tratar de processos estocásticos com a linguagem de espaços de Fock [12, 55, 46, 11].

Ainda que não seja necessário, o formalismo de integrais de trajetória para tratar de processos estocásticos é normalmente apresentado a partir de uma identificação formal com a física quântica. Adotamos uma abordagem diferente, ressaltando a conexão com a mecânica quântica apenas após desenvolver os conceitos principais em uma linguagem puramente estatística.

Todas funções $P(x_t|x'_t)$ que obedecem à equação de Chapman-Kolmogorov (2.3) possuem uma representação como integral funcional [29, 27]. No entanto, nem toda integral funcional produz um processo estocástico válido pois também é necessário que as soluções para a Eq. (2.3) representem probabilidades válidas. Ao relacionar o formalismo de integrais de trajetória com a teoria de funções características, obtemos as formas mais gerais para uma integral de trajetória que representa um processo estocástico Markoviano. Normalmente a literatura se restringe a dois casos particulares: os processos de Fokker-Planck e os processos de criação e destruição. Essa conexão é, no conhecimento do autor, original e permite apresentar a teoria de processos Markovianos de uma maneira bastante unificada, além de estender o formalismo de integrais de trajetória para outros processos estocásticos como os vôos de Lévy.

2.1 Equação Mestra

Com o intuito de definir a equação mestra, é interessante utilizar uma notação mais explícita para variáveis contínuas: $P(x_t|x_{t'}) = G_{tt'}(x|x')dx$. Na nova notação, equação de Chapman-Kolmogorov se escreve como

$$G_{tt'}(x|x') = \int dy G_{ts}(x|y)G_{st'}(y|x'), \quad (2.4)$$

para $t \geq s \geq t'$. Observe o comportamento no limite $s \rightarrow t$: a equação acima é consistente somente se $G_{tt}(x|x') = \delta(x-x')$. Obtemos um resultado não-trivial derivando (2.4) por t antes de tomar esse limite: isto resulta em uma equação dinâmica para $G_{tt'}(x|x')$ que se escreve como

$$\partial_t G_{tt'}(x|x') = \int dy \bar{W}_t(x|y)G_{tt'}(y|x'), \quad (2.5)$$

onde $\bar{W}_t(x|x') \equiv \partial_t G_{ts}(x|x')|^{s=t}$.

Uma vez fixado $\bar{W}_t(x|x')$, a solução da equação diferencial (2.5) com a condição inicial $G_{t,t'}(x|x')|^{t'=t} = \delta(x-x')$ é unívoca. Observe que $\bar{W}_t(x|x')$ é definido a partir da própria probabilidade de transição, o que faz com que a Eq. (2.5) seja, no fundo, uma definição circular. A utilidade da equação acima está no fato que é possível modelar a probabilidade de uma transição infinitesimal de uma forma relativamente simples. Com isso, a Eq. (2.5) pode ser entendida como uma prescrição para reconstruir $G_{tt'}(x|x')$ a partir do comportamento desta probabilidade no limite $t - t' = \delta t \rightarrow 0$, que é codificado por $W_t(x|x')$.

Normalmente a equação mestra é apresentada de uma forma um pouco diferente, que garante que suas soluções sejam densidades de probabilidades válidas. Para tanto, assumimos que a probabilidade de transição no limite

$\delta t \rightarrow 0$ possui a forma abaixo

$$G_{t'+\delta t,t'}(x|x') = (1 - \lambda_t(x') \delta t) \delta(x - x') + \delta t W_t(x|x'), \quad (2.6)$$

onde $\lambda_t(x') = \int dx W_t(x|x')$ e $W(x|x') > 0$. A Eq. (2.6) representa a aproximação para a probabilidade de transição em primeira ordem em δt , garantindo a sua positividade, normalização e a condição de contorno $G_{tt}(x|x') = \delta(x - x')$.

Substituímos a Eq. (2.6) na Eq. (2.5), obtendo

$$\partial_t G_{tt'}(x|x') = \int dy [W_t(x|y) - \lambda_t(y) \delta(x - y)] G_{tt'}(y|x'), \quad (2.7)$$

$$= \int dy [W_t(x|y) G_{tt'}(y|x') - W_t(y|x) G_{tt'}(x|x')], \quad (2.8)$$

onde a passagem de (2.7) para (2.8) usa o fato que $\lambda_t(x') = \int dx W_t(x|x')$. Essa é a forma canônica da equação mestra, que é interpretada como uma equação de ganhos e perdas para as probabilidades. Num intervalo δt , cada elemento em y possui uma chance $W_t(x|y) \delta t$ de transitar para x . Assim, o acréscimo de probabilidade em x se dá pela soma das transições sobre todos os estados y , como determina o primeiro termo da Eq. (2.8). Por outro lado, os elementos inicialmente em x podem transitar para outros estados diminuindo o valor da probabilidade em x . O segundo termo caracteriza a perda de probabilidade em x devido a estas transições.

Interpretamos $W_t(x|x')$ como um de fluxo de probabilidade de x' para x . Por comparação direta da Eq. (2.8) com a Eq. (2.5) vemos que os dois núcleos W e \bar{W} se relacionam por

$$\bar{W}_t(x|x') = W_t(x|x') - \lambda_t(x') \delta(x - x'). \quad (2.9)$$

O núcleo $W_t(x|x')$ caracteriza univocamente uma probabilidade de transição pela equação mestra. Note que existem processos que não podem ser escritos

por uma equação mestra como (2.8). Nesses casos, a descrição se dá pela equação de Fokker-Planck generalizada a ser apresentada na Sec. 2.3.

2.1.1 Equação mestra e processos de saltos

Existe uma interpretação física interessante para os resultados anteriores. Em vista da Eq. (2.6), a equação mestra descreve um processo que, infinitesimalmente, possui uma probabilidade $1 - \lambda_t(x')\delta t$ de não realizar transição alguma e uma probabilidade $\lambda_t(x')\delta t$ de realizar uma transição de acordo com

$$q_t(x|x') \equiv W_t(x|x')/\lambda_t(x'). \quad (2.10)$$

Assim, espera-se que a dinâmica das funções $x(t)$ de um processo definido pela equação mestra (2.8) seja descontínua. O sistema permaneceria um tempo indeterminado no estado x' e, após um intervalo de espera aleatório, executaria uma transição de acordo com $q_t(x|x')$. Esse processo então se repetiria indefinidamente.

Podemos estimar o tempo médio entre duas transições se $\lambda_t(x')$ não depender do tempo. A probabilidade de não realizar nenhuma transição após um tempo τ pode ser decomposta a partir da probabilidade infinitesimal $1 - \lambda(x')\delta t$ como abaixo

$$S_\tau(x') = \lim_{n \rightarrow \infty} (1 - \lambda(x')\tau/n)^n = e^{-\lambda(x')\tau}. \quad (2.11)$$

O termo $S_\tau(x')$ é conhecido como a função de sobrevivência do processo.

A formulação precisa de um processo de saltos é o tema de um capítulo posterior. Por enquanto, lembramos que os processos descritos pela Eq. (2.6) possuem uma dinâmica descontínua. Isto os distingue de modelos como o processo de Wiener, onde as transições se dão continuamente ao longo do tempo.

2.2 Integrais de trajetória

A equação mestra foi obtida supondo uma forma específica para a probabilidade de transição entre dois instantes infinitamente próximos. No formalismo de integrais de trajetórias, fazemos o mesmo, mas o objetivo não é mais obter uma equação diferencial. Nessa abordagem, quebramos o intervalo $\tau = t - t'$ em pequenas partes de tamanho $\delta t = \tau/N$ para escrever

$$G_{tt'}(x_N|x_0) = \int dx_1 \dots dx_{N-1} \prod_{i=1}^N G_{t_i t_{i-1}}(x_i|x_{i-1}), \quad (2.12)$$

onde $G_{t_i t_{i-1}}(x_i|x_{i-1})$ eventualmente será substituído por uma aproximação como a Eq. (2.6) quando $N \rightarrow \infty$.

A questão central no formalismo de integrais de trajetória é definir o limite de infinitas integrações sucessivas na Eq. (2.15). Esse limite pode ser tratado de maneira *absolutamente* rigorosa em apenas algumas situações, como, por exemplo, o processo de Wiener [58]. Mesmo nos casos onde ainda não existe uma justificativa formal precisa (como nas aplicações à mecânica quântica), o formalismo de integrais de trajetória se mostrou extremamente útil. É possível fazer um paralelo com o cálculo ordinário: Riemann e Cauchy forneceram respectivamente a definição rigorosa do conceito de integral e derivada. Isto não significa que os seus antecessores estavam equivocados, nem que seus cálculos eram fundamentalmente mal definidos. De fato, em boa parte dos casos, a concepção mais intuitiva de Newton e Leibniz reproduz corretamente os resultados da integral de Riemann e a definição de limite segundo Cauchy.

Para efeito de argumento, considere a integral de Riemann, definida como o limite das somas parciais

$$\int_a^b dt f(t) = \lim_{N \rightarrow \infty} \sum_{n=1}^N (t_n - t_{n-1}) f(\bar{t}_n). \quad (2.13)$$

Na Eq. (2.13), particionamos o intervalo $[a, b]$ em $a = t_0 > t_1 > t_2 > \dots > t_N = b$, e os pontos \bar{t}_n são escolhidos arbitrariamente no intervalo $[t_n, t_{n-1}]$. Dizemos que uma função $f(t)$ é integrável no intervalo a, b se o limite ao lado direito de (2.13) existe e não depende da escolha precisa de \bar{t}_n . Informalmente, o resultado de uma integral de Riemann corresponde ao limite da soma de infinitos números infinitamente pequenos. Riemann forneceu uma definição matematicamente rigorosa para o limite ao lado direito da equação, mas que não discutiremos nesta apresentação.

De maneira análoga ao caso anterior, definimos a integral de trajetória como a convolução de infinitas funções infinitesimalmente próximas à delta de Dirac. É o análogo funcional da integração comum, onde multiplicamos entre si infinitos operadores infinitesimalmente próximos da unidade. Vale notar que, ao contrário da integral de Riemann, não existe uma definição matematicamente rigorosa para o processo de limite. Trabalhamos com noções intuitivas como números infinitesimais, que apesar de serem amplamente utilizadas por físicos e engenheiros, não são particularmente bem vistas entre os matemáticos.

A distribuição de Dirac é um objeto bastante singular e, a rigor, só é bem definida caso se encontre dentro do sinal da integral. Por esse motivo, é conveniente trabalhar no espaço de Fourier, onde ela se reduz a uma constante. É possível dar sentido à noção de uma função infinitamente próxima à delta de Dirac escrevendo

$$G_{t+\delta t, t}(x|x') = \frac{1}{2\pi} \int dz e^{\delta t H(z, x', t) - iz(x-x')}, \quad (2.14)$$

onde $H(z, x', t)$ caracteriza a divergência com relação à delta de Dirac, que é recuperada no limite $\delta t \rightarrow 0$. A definição anterior é conveniente porque, ao substituir na Eq. (2.12), o expoente do integrando vira formalmente uma

integral ordinária:

$$P(x_t|x_{t'}) = G_{tt'}(x_N|x_0) dx_N = \int Dx Dz e^{-S[z,x]}, \quad (2.15)$$

onde

$$S = \sum_{i=1}^N [iz_i(x_i - x_{i-1}) - \delta t H(z_i, x_{i-1}, t_{i-1})], \quad (2.16)$$

$$= \int_0^\tau dt [iz\dot{x} - H(z, x, t)], \quad (2.17)$$

e definimos $Dx Dz = \prod_{i=1}^N (dx_i dz_i/2\pi)$ — observe que não integramos sobre o último termo dx_N . Chamamos S na Eq. (2.15) de “ação” e a função $H(z, x, t)$ de “Hamiltoniana” associadas à integração funcional. Essa nomenclatura é emprestada da física quântica, onde estes objetos correspondem, de fato, aos nomes que lhes atribuímos.

Ainda que a definição de S na Eq. (2.16) seja essencialmente o que aparece na Eq. (2.13), é preciso um pouco de cautela porque, via de regra, o termo $H(z, x, t) + i\dot{x}z$ não é Riemann-integrável. Isto é um problema nos processos em que $x(t)$ não é diferenciável, tornando o termo $i\dot{x}z$ divergente. Em alguns casos é possível mitigar esse problema usando uma definição diferente de integral. Considere a integral de Stieltjes definida abaixo

$$\int_a^b dx(t) f(t) = \lim_{N \rightarrow \infty} \sum_{n=1}^N [x(t_n) - x(t_{n-1})] f(\bar{t}_n), \quad (2.18)$$

que coincide com a integral de Riemann $\int_a^b dt \dot{x}(t) f(t)$ se $x(t)$ for uma função diferenciável. Existem algumas funções que não são integráveis no sentido de Riemann, mas o são no sentido de Stieltjes — um exemplo importante é a delta de Dirac. No resto da apresentação, utilizaremos $dx(t)$ ou $dt \dot{x}$ indiscriminadamente, mesmo quando $x(t)$ não for diferenciável — nesse caso,

$\int dt \dot{x}(\dots)$ deve ser entendido como uma notação conveniente para a integral de Stieltjes $\int dx(t)(\dots)$.

Ao estudar as trajetórias do movimento Browniano, o matemático japonês Kiyoshi Itô constatou que a definição de Stieltjes não era apropriada: o resultado da integral depende do ponto \bar{t}_n escolhido dentro de cada partição em (2.18). Para estas funções, é necessário especificar uma prescrição para se escolher os instantes \bar{t}_n para obter uma definição de integral consistente.

A prescrição de Itô é a mesma que adotada até aqui e consiste em utilizar o ponto anterior, $\bar{t}_n = t_{n-1}$. Uma outra possibilidade muito comum é a integral de Stratonovich, que adota o ponto central $\bar{t}_n = \frac{1}{2}(t_n + t_{n-1})$. Ainda que as duas definições coincidam em funções Riemann-integráveis, os resultados são diferentes quando se trata de trajetórias do movimento Browniano e possivelmente em várias outras situações.

Como um exemplo de que é importante se prender atenção a estes detalhes, ressaltamos que o teorema fundamental do cálculo assume uma forma inusitada no cálculo de Itô se $x(t)$ for uma trajetória Browniana. Nesse caso, é possível demonstrar que

$$\int_a^b dx(t) \frac{dF(x)}{dx} = F(x(b)) - F(x(a)) - \frac{1}{2} \int_a^b dt D_2(x) \frac{\partial^2}{\partial x^2} F(x), \quad (2.19)$$

onde $D_2(x)$ é o coeficiente de difusão processo de Browniano a ser definido posteriormente. Ao fazer $D_2 \rightarrow 0$, as trajetórias Brownianas se tornam diferenciáveis e a integral de Itô se iguala à integral de Riemann usual. A prescrição de Stratonovich elimina o segundo termo e assim resulta na mesma regra do cálculo de Riemann — mesmo que o integrando não seja Riemann-integrável.

2.2.1 Função Hamiltoniana

O fato que as funções $G_{t+\delta t, t'}(x|x')$ são probabilidades implica em certas restrições sobre a Hamiltoniana que define a integração funcional. De fato, identificamos o termo

$$\phi_t(z|x') = e^{\delta t H(z, x', t) + izx'}, \quad (2.20)$$

como a função característica da distribuição de probabilidades para $G_{t+\delta t, t'}(x|x')$. Uma das dificuldades com a expressão anterior está em saber quando ela, de fato, representa uma função característica válida. Que propriedades a função $H(z, x', t)$ deve ter para que, ao obter $G_{t+\delta t, t}(x|x')$ a partir de (2.20), a resposta seja uma distribuição de probabilidades positiva e normalizada?

A forma geral (2.20) é bem estudada na literatura de funções características e representa a função característica de uma variável infinitamente divisível. Dizemos que uma variável aleatória é *divisível* se puder ser escrita como a soma de duas outras variáveis aleatórias independentes. Fica claro pela Eq. (2.20), que $x_{t+\delta t}$ corresponde à soma de um número arbitrário de variáveis independentes com a função característica dada por $\phi_t(z|x')^{1/n}$. Fisicamente, isto significa que o incremento δx no intervalo δt corresponde à soma de n incrementos independentes em um intervalo $\delta t/n$. Uma vez que n pode ser tão grande quanto se queira, esta propriedade define uma variável conhecida como *infinitamente divisível*.

De Finetti demonstrou as condições necessárias e suficientes para que a Eq. (3.4), que representa variáveis infinitamente divisíveis, seja uma função característica válida. Citamos sem demonstração:

Theorem 2.1 (De Finetti). *Seja $\psi(z)$ uma função característica. Então $e^{\alpha(\psi(z)-1)}$ é a função característica de uma variável infinitamente divisível para $\alpha > 0$.*

De modo semelhante, a função característica de qualquer distribuição infinita-

mente divisível pode ser escrita como $\lim_{n \rightarrow \infty} e^{\alpha_n(\psi_n(z)-1)}$, para uma seqüência convergente de $\alpha_n > 0$ e de funções características $\psi_n(z)$.

Aqui vale uma observação. É possível que o expoente de uma função característica infinitamente divisível não seja fatorado na forma $\alpha(\psi(z) - 1)$. Esse é o caso da Gaussiana, cuja função característica é dada por $e^{-\frac{1}{2}\sigma^2 z^2}$. Nesse caso, identificamos $\alpha = \frac{1}{2}\sigma^2$ e $\psi_G(z) = 1 - z^2$, sendo que $\psi_G(z)$ não representa uma função característica válida. No entanto, uma Gaussiana de variância σ^2 é infinitamente divisível pois corresponde à soma de outras n variáveis Gaussianas com variância σ^2/n . De todo modo, é fácil verificar que

$$\lim_{n \rightarrow \infty} e^{n\left(e^{-\frac{1}{2}n^{-1}\sigma^2 z^2} - 1\right)} = e^{-\frac{1}{2}\sigma^2 z^2}, \quad (2.21)$$

onde o argumento anterior ilustra o significado do limite no teorema de De Finetti.

Comparando a representação de De Finetti com a Eq. (2.20), vemos que

$$H(z, x', t) = \lambda_t(x') [\psi_t(z|x') - 1], \quad (2.22)$$

onde $\lambda_t(x')$ é um parâmetro com dimensão de freqüência que deixa o expoente adimensional e $\psi_t(z|x')$ é uma função característica arbitrária. O teorema de De Finetti também contempla situações limites como aquela ilustrada na Eq. (2.21). A expressão (2.22) garante que as probabilidades de transição infinitesimais $G_{t+\delta t, t}(x|x')$ são positivas e normalizadas.

Existem outras representações úteis para uma função característica. Invocamos um teorema devido a Lévy e Khinchin [44] que diz que toda função característica pode ser escrita como

$$\phi(z) = e^{-|sz|^\alpha(1-i\beta\text{sgn}(z)\Phi)(1-\omega(sz))+iz\ell}, \quad (2.23)$$

onde ℓ e s são os parâmetros de escala e localização da distribuição¹. Defina-se $\Phi = -\frac{2}{\pi} \ln |z|$ para $\alpha = 1$ e $\Phi = \tan\left(\frac{\pi\alpha}{2}\right)$, caso contrário. O parâmetro $-1 < \beta < 1$ é um coeficiente de assimetria: para $\beta = \pm 1$ a variável aleatória resultante assume somente valores positivos ou negativos. A forma da distribuição de probabilidades é controlada em grande parte pela função $\omega(sz)$, que se for identicamente nula corresponde a uma distribuição estável de Levy (é um atrator para a soma de variáveis independentes). Para toda função característica válida, é verdade que $\omega(0) = 0$, além de $\omega(z)$ ser uma função contínua em uma região próxima da origem. O expoente α possui valores no intervalo $0 < \alpha \leq 2$ e, somente quando $\alpha = 2$, a distribuição resultante possui desvio padrão e média finitos².

Ao comparar a Eq. (2.20) com a Eq. (2.23), lembramos que os parâmetros s e ℓ podem, em princípio, depender de x' , t e δt . Além disto, é necessário que a Eq. (2.23) recupere a função característica da delta de Dirac, dada por $e^{-iz(x-x')}$, quando $\delta t \rightarrow 0$. Deste modo, identificamos o parâmetro de localização como $\ell = x' + \delta t D_1$, onde D_1 é o chamado coeficiente de drift, que pode depender tanto de x' quanto de t . Pelo mesmo raciocínio, vemos que o parâmetro de escala é $s = (\delta t D_2)^{1/\alpha}$, onde $D_2 = D_2(x', t) \geq 0$ é o coeficiente de difusão.

A função $\omega(\cdot)$ pode depender explicitamente de x' , t e δt , e assim escrevemos $\omega(sz|x', t, \delta t)$. Considere o limite

$$\bar{\omega}(z, x', t) \equiv \lim_{\delta t \rightarrow 0} \omega\left((\delta t D_2)^{1/\alpha} z \mid x', t, \delta t\right). \quad (2.24)$$

Em vários casos de interesse, $\bar{\omega} = 0$, porque o argumento sz da função ω se anula e $\omega(0|x', t, \delta t) = 0$. É claro que a dependência em δt pode impedir a

¹Estes parâmetros controlam a dispersão e a localização da distribuição. No caso de uma probabilidade Gaussiana, por exemplo, eles correspondem respectivamente ao desvio padrão multiplicado por $\sqrt{2}$ e à média.

²Não é verdade, no entanto, que toda função deste tipo represente uma função característica válida.

convergência para zero, já que estamos modificando simultaneamente s e δt . De qualquer modo, a restrição $\bar{\omega}(0, x', t) = 0$ persiste para que tenhamos uma função característica consistente.

Após estas considerações, estamos prontos para escrever a Hamiltoniana como

$$H(z, x, t) = D_2(x, t) |z|^\alpha (1 - i\beta \operatorname{sgn}(z)\Phi(z)) (1 - \bar{\omega}(z, x, t)) - iD_1(x, t). \quad (2.25)$$

Observe que a forma (2.23) não garante uma função característica válida e isto depende essencialmente da função $\bar{\omega}(z, x, t)$. Para $\bar{\omega} = 0$, teríamos uma distribuição estável e o processo estocástico correspondente seria descrito por uma equação de Fokker-Planck generalizada dada na Sec. 2.3.1. Nesse caso em particular, é possível garantir a validade de $H(z, x, t)$ e o processo estocástico fica especificado apenas pelos coeficientes de drift e de difusão. De um modo geral, é necessário garantir que $H(z, x, t)$ também possa ser escrito como a Eq. (2.22) ou uma forma limite a partir desta expressão.

É possível relacionar as duas expressões (2.22) e (2.25). Para isto, é necessário determinar o expoente α e as constantes s , ℓ e β da representação de Lévy a partir da função característica $\psi(z)$ que aparece em (2.22). Observamos que $\psi(z)$ também possui uma representação de Lévy com seus coeficientes característicos α_0, β_0, s_0 e ℓ_0 . A relação entre ambos pode ser facilmente estabelecida considerando o limite para $z \rightarrow 0$ da função característica.

Todas funções características se comportam da mesma maneira nesse limite: $\phi(z) \sim 1 - |s z|^\alpha (1 - i\beta \operatorname{sgn}(z)\Phi) + i\ell z$, onde lembramos que $\omega(z \rightarrow 0)$. Cada termo na expressão anterior controla um aspecto diferente da função característica e podemos determiná-los a partir de operações simples. Deste modo, verificamos que os termos α e β de $\phi(z)$ são os mesmos que os de $\psi(z)$,

mas os coeficientes de drift e difusão ficam dados por

$$D_1 = \lambda \ell_0; \quad D_2 = \lambda s_0^\alpha. \quad (2.26)$$

Obtemos o valor de $\bar{\omega}$,

$$\bar{\omega}(z) = 1 - \frac{\psi(z) - 1}{|s_0 z|^\alpha (1 - i\beta \operatorname{sgn}(z)\Phi)}, \quad (2.27)$$

onde é possível constatar que $\bar{\omega}(z)$ se aproxima continuamente de zero no limite $z \rightarrow 0$.

2.3 Equações dinâmicas

A probabilidade de transição infinitesimal pode ser escrita como

$$G_{t+\delta t, t}(x|x') = \frac{1}{2\pi} \int_0^\infty dz e^{-\delta t H(z, x', t) - iz(x-x')}, \quad (2.28)$$

onde a Hamiltoniana foi definida anteriormente em (2.22) e (2.25). Vamos utilizar esse resultado para definir uma equação diferencial para $G_{t, t'}(x|x')$, avaliando $[G_{t+\delta t, t'}(x|x') - G_{t, t'}(x|x')]/\delta t$ a partir da equação de Chapman-Kolmogorov (2.4). O resultado disto é mostrado abaixo

$$\partial_t G_{t, t'}(x|x') = \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} \int dy [G_{t+\delta t}(x|y) - \delta(x-y)] G_{t, t'}(y|x'). \quad (2.29)$$

Ao substituir a representação de Fourier dos termos entre colchetes, o lado direito da equação se reduz a

$$\begin{aligned} \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} \int dy \int dz [e^{\delta t H(z, y, t)} - 1] \frac{e^{-iz(x-y)}}{2\pi} G_{t, t'}(y|x') \\ = \int dy \int dz H(z, y, t) \frac{e^{-iz(x-y)}}{2\pi} G_{t, t'}(y|x'). \end{aligned} \quad (2.30)$$

Daí obtemos uma relação do tipo equação de Schrödinger,

$$\partial_t G_{tt'}(x|x') = \hat{W}_t G_{tt'}(x|x'), \quad (2.31)$$

$$= \int dy \bar{W}_t(x, y) G_{tt'}(y|x'), \quad (2.32)$$

onde o operador Hamiltoniano \hat{W}_t que aparece na Eq. (2.31) possui o núcleo

$$\bar{W}_t(x, x') = \int dz H(z, x', t) \frac{e^{-iz(x-x')}}{2\pi}. \quad (2.33)$$

A “Hamiltoniana” definida na Eq. (2.25) recebeu esse nome em analogia ao mesmo objeto que aparece na representação de integral de trajetórias em mecânica quântica. E, de fato cumpre, ela cumpre um papel análogo ao da Hamiltoniana tradicional no sentido de ser o operador de evolução temporal. É claro que esta “Hamiltoniana” não representa a energia do sistema e, inclusive, possui autovalores complexos: a analogia com a Hamiltoniana tradicional cessa por aqui.

Utilizando as propriedades das transformadas de Fourier, é possível reescrever (2.31) como uma equação diferencial. No cálculo de $\bar{W}_t(x, x')$ na Eq. (2.33), cada potência $f(x')z^n$ na Hamiltoniana vira uma vira um termo do tipo $f(x') (i\partial_x)^n \delta(x - x')$. Ao aplicar cada um destes termos à $G_t(x|x')$ como na Eq. (2.32), ficamos com $(i\partial_x)^n [f(x)G_{tt'}(x|x')]$. Esse resultado pode ser entendido como uma regra de conversão de uma Hamiltoniana escalar em um operador, ou uma “regra de quantização” na linguagem da teoria quântica. Segundo o argumento anterior, um termo do tipo $x^n z^m$ na Hamiltoniana viraria um operador $(i\partial_x)^n [x^n \cdot]$ atuando em $G_{tt'}(x|x')$. Diferentemente da prescrição de simetrização normalmente utilizada em mecânica quântica, que é dada por $xp \mapsto \frac{1}{2}(\hat{x}\hat{p} + \hat{p}\hat{x})$, a prescrição correta no nosso caso é colocar os operadores de “momentum” $\hat{z} = i\partial_x$ à esquerda dos operadores de posição x .

É possível mostrar que a discrepância com relação à regra utilizada na

mecânica quântica se deve à utilização da integral estocástica de Itô ou a de Stratonovich. O primeiro leva à regra de colocar os operadores de momentum à esquerda dos operadores de posição enquanto o segundo está associado à regra de simetrização.

2.3.1 Equações de Fokker-Planck

Consideramos o caso específico em que $\bar{\omega} = 0$ na Hamiltoniana (2.25), caracterizando uma distribuição infinitesimal estável. Para começar, explicitamos a equação mestra para $\alpha = 2$, que corresponde ao processo de Wiener com ruído Gaussiano. De acordo com a regra de “quantização” discutida anteriormente, a equação dinâmica poderia ser escrita como

$$\partial_t G_{tt'}(x|x') = \left\{ -\partial_x [D_1(x, t) \cdot] + \partial_x^2 [D_2(x, t) \cdot] \right\} G_{tt'}(x|x'), \quad (2.34)$$

que é a bem conhecida equação de Fokker-Planck.

Caso $\alpha \neq 2$, existe uma complicação adicional pois o termo de momentum z torna-se uma lei de potência fracionária. Nesse caso, é necessário definir formalmente um operador

$$\Delta_x^{\alpha, \beta} f(x) = - \int dz dy |z|^\alpha (1 - i\beta \text{sgn}(z)\Phi) \frac{e^{-iz(x-y)}}{2\pi} f(y), \quad (2.35)$$

que chamamos de uma derivada fracionária³. Desta definição, resulta a equação de Fokker-Planck generalizada

$$\partial_t G_{tt'}(x|x') = \left\{ -\partial_x [D_1(x, t) \cdot] + \Delta_x^{\alpha, \beta} [D_2(x, t) \cdot] \right\} G_{tt'}(x|x'). \quad (2.36)$$

Valores entre $0 < \alpha < 2$ definem os chamados processos de difusão fractal,

³Em que pese, existem inúmeras definições de derivada fracionária na literatura. Na nossa definição, recuperamos as derivadas ordinárias se $\beta = 0$ e se α for um número par.

associados a distribuições de Lévy com variância infinita. É fácil verificar que, para $\alpha = 2$, a derivada fracionária coincide com ∂_x^2 e a equação de Fokker-Planck usual é restaurada.

2.3.2 Equação mestra e expansão de Kramers-Moyal

Ao avaliar a Eq. (2.33) utilizando a Eq. (2.22), obtemos

$$\bar{W}_t(x, x') = \lambda_t(x') [q_t(x|x') - \delta(x - x')],$$

tendo definido $q_t(x|x') = \int dz \psi_t(z|x') \frac{e^{-iz(x-x')}}{2\pi}$. Esse resultado também foi obtido em (2.9), onde identificamos $W_t(x|x') = \lambda_t(x')q_t(x|x')$ e é equivalente à equação mestra (2.8).

É possível escrever a mesma equação na forma diferencial expandindo $\psi_t(z|x')$ em série de potências. Caso $\psi_t(z|x')$ seja uma função analítica, teríamos

$$\psi_t(z|x') = 1 + \sum_{n=1}^{\infty} \frac{\mu_n(x', t)}{n!} (iz)^n. \quad (2.37)$$

A Hamiltoniana corresponde portanto à

$$H = \sum_{n=1}^{\infty} \frac{\mu_n(x', t)}{n!} \lambda_t(x') (iz)^n \equiv \sum_{n=1}^{\infty} D_n(x', t) (iz)^n, \quad (2.38)$$

onde $D_n(x', t) \equiv \lambda_t(x')\mu_n(x', t)/n!$ e $\mu_n(x', t)$ são os momentos estatísticos de $q(x|x')$ em torno de x' . A regra de “quantização”, que corresponde a colocar os operadores de momentum à esquerda resulta na conhecida expansão de Kramers-Moyal

$$\partial_\tau G_\tau(x|x') = \sum_{n=1}^{\infty} (-1)^n \frac{\partial^n}{\partial x^n} [D_n(x, t) G_\tau(x|x')]. \quad (2.39)$$

A expansão de Kramers-Moyal geralmente é escrita com o intuito de truncá-

la até uma ordem finita. No entanto, de acordo com o teorema de Pawula, qualquer truncagem da Eq. (2.39) até o termo $n \geq 3$ gera distribuições negativas, geralmente em tempos transientes. Ainda que o comportamento seja indecifrável, muitas vezes é interessante resolver a série truncada como aproximação para o problema mais complicado de resolver a equação mestra completa.

2.3.3 Ruído de Wiener: formulação Lagrangiana

Wiener introduziu a noção de integral de trajetória ao estudar o movimento Browniano. Esse formalismo assume que $\bar{\omega} = 0$ e $\alpha = 2$, correspondendo à Hamiltoniana

$$H(z, x, t) = -D_2(x, t) z^2 + i D_1(x, t) z. \quad (2.40)$$

A interpretação física desta escolha é que a probabilidade de realizar uma transição infinitesimal é uma Gaussiana com variância $2D_2\delta t$ e média $D_1\delta t$. O coeficiente de drift D_1 representa uma tendência enquanto D_2 caracteriza a flutuação aleatória: esta interpretação se aplica a qualquer distribuição, mesmo quando a média e o desvio padrão não são bem definidos.

O processo de Wiener corresponde a um dos poucos casos onde é possível realizar as integrações sobre z na Eq. (2.15). O resultado é a expressão

$$P(x_t|x_{t'}) = G_{tt'}(x_N|x_0) dx_N = \int Dx e^{-S'[x, \dot{x}]}, \quad (2.41)$$

onde a nova “ação” $S' = \int_0^T dt L(x, \dot{x}, t)$ é escrita para a “Lagrangiana”

$$L(x, \dot{x}, t) \equiv \frac{1}{4D_2(x, t)} (\dot{x} - D_1(x, t))^2, \quad (2.42)$$

tendo definido a derivada de $x(t)$ como uma diferença finita $\dot{x}(t_i) = (x_i - x_{i-1})/\delta t$.

⁴A versão discretizada da Lagrangiana seria escrita como $L_i = [(x_i - x_{i-1})/\delta t - D_1(x_{i-1}, t_{i-1})]^2/4D_2(x_{i-1}, t_{i-1})$, onde a ação corresponde à soma $S = \sum_{i=0}^N \delta t L_i$.

O elemento de integração é $Dx = \prod_{n=1}^N \frac{dx_n}{\sqrt{2\pi D_2(x_{n-1}, t_i)}}$ e adotamos a prescrição de não integrar sobre dx_N

Ainda que os nomes “ação” e “Lagrangiana” sejam sugestivos, é necessário interpretá-los com um pouco de cautela. Ao tentar encontrar a trajetória mais provável, a que minimiza a ação definida para a Lagrangiana (2.42), é tentador utilizar a experiência prévia com as equações de Euler-Lagrange. No entanto, isto não funciona.

Considere um processo em que D_2 é constante e D_1 depende apenas de x . A ação desse processo estocástico se escreveria explicitamente como

$$S' = \frac{1}{4D_2} \int_0^\tau dt [\dot{x}^2 + D_1(x)^2 - 2D_1(x)\dot{x}], \quad (2.43)$$

onde o último termo representa uma derivada total em t : definimos $D_1 = dG/dx$, e assim $dG/dt = D_1\dot{x}$. O termo com a derivada total não pode ser desprezado como acontece na física clássica. Na integral de Itô, o teorema fundamental do cálculo para trajetórias do movimento Browniano se escreve como

$$\int_0^\tau dx(t) D_1(x) = G(x(\tau)) - G(x(0)) - \frac{D_2}{2} \int_0^\tau dt \frac{dD_1(x)}{dx}. \quad (2.44)$$

Usando estas regras de integração de maneira consistente, a ação (2.43) seria equivalente à expressão abaixo

$$S' = \frac{1}{4D_2} \int_0^\tau dt \left[\dot{x}^2 + D_1^2 + D_2 \frac{dD_1}{dx} \right] + \text{cte}, \quad (2.45)$$

onde a diferença com relação ao resultado clássico decorre das trajetórias $\dot{x}(t)$ não serem Riemann-integráveis.

Como proceder a partir daí?

Os termos D_1^2 e dD_2/dx são Riemann-integráveis e portanto não apresentam

grandes problemas. O termo cinético, no entanto, é problemático: a derivada das trajetórias do movimento Browniano diverge em quase todos os pontos, e possivelmente algo pior acontece com a derivada ao quadrado. Aqui é necessário voltarmos para a definição da integral (2.45) como uma soma discreta, ou seja,

$$S' = \lim_{N \rightarrow \infty} \frac{1}{4D_2} \sum_{n=0}^{N-1} \left[\frac{(x_{n+1} - x_n)^2}{\delta t} + \delta t D_1(x_n)^2 + D_2 \frac{dD_1(x_n)}{dx_n} \right]. \quad (2.46)$$

Minimizamos a ação impondo que $\frac{\partial}{\partial x_i} S'(x_0, \dots, x_N) = 0$. O resultado, é

$$\frac{2x_i - x_{i+1} - x_{i-1}}{\delta t^2} = D_1(x_i) \frac{dD_1(x_i)}{dx_i} + D_2 \frac{d^2 D_1(x_i)}{dx_i^2}, \quad (2.47)$$

que corresponde à versão em diferenças finitas da equação diferencial

$$\ddot{x}(t) = D_1(x) \frac{dD_1}{dx} + D_2 \frac{d^2 D}{dx^2}. \quad (2.48)$$

Incidentalmente, a Eq. (2.48) também resulta de se utilizar as equações de Euler-Lagrange com a seguinte Lagrangiana

$$L = \frac{1}{4D_1} \left[\dot{x}^2 + D_1^2 + D_2 \frac{dD_1}{dx} \right], \quad (2.49)$$

que é o mesmo termo que aparece no integrando da Eq. (2.45). Essa Lagrangiana também aparece quando reescrevemos a integral (2.43) para o cálculo de Stratonovich. Deste modo, alguns autores argumentam que a Eq. (2.49) fornece a “verdadeira Lagrangiana” do sistema enquanto que Eq. (2.42) estaria errada. Talvez esta afirmação seja muito forte: utilizando as regras do cálculo de Itô de maneira consistente, a Eq. (2.42) é equivalente à Eq. (2.49). Na realidade, é possível mostrar que a Eq. (2.49) é uma representação válida da Lagrangiana em qualquer cálculo [23]. Na regra de Itô, a Eq. (2.49) também é

equivalente à Eq. (2.42). Na regra de Stratonovich construímos Lagrangianas equivalentes adicionando termos de derivada total no tempo — outras regras correspondem a diferentes critérios de invariância.

2.4 Conclusão

Historicamente, a integração funcional surgiu com Wiener no contexto de processos estocásticos. No entanto, à exceção da equação de Fokker-Planck, as técnicas de integração funcional são invariavelmente apresentadas em conexão formal com a mecânica quântica e seus operadores, regras de comutação, etc. Um dos objetivos desta exposição foi explicitar que esta incursão pela mecânica quântica não é necessária.

Uma dificuldade na formulação do processo estocástico que não aparece em mecânica quântica é o problema de representação de modelos. A única exigência formal sobre a Hamiltoniana quântica é que a mesma seja Hermitiana: é trivial definir um modelo quântico (mais difícil é definir um modelo com uma física interessante). A situação na teoria de processos Markovianos é menos encorajadora: são conhecidas algumas classes de Hamiltonianas que geram probabilidades de transição legítimas. Mas, estes modelos esgotam todos os casos? Dada uma Hamiltoniana arbitrária, como descobrimos se ela representa corretamente um processo Markoviano?

Apresentamos o critério geral que determina quais são as Hamiltonianas legítimas para um processo estocástico. Com isto, é possível formular modelos Markovianos inteiramente na linguagem de integração funcional. Acreditamos que a existência de um critério deste tipo pode encorajar pesquisas futuras na utilização do formalismo de integrais de trajetória para analisar processos Markovianos.

A formulação Hamiltoniana permite mapear todos processos estocásticos

Markovianos que existem. Essa generalidade vem com um preço: nem sempre é possível escrever a ação na forma Lagrangiana. Por esse motivo, alguns métodos importantes como a aproximação de WKB não funcionariam. Um tema natural para a pesquisa futura é considerar métodos aproximativos que também funcionem nestes casos.

Capítulo 3

Processos de saltos

Feller mostrou que todo processo Markoviano autônomo é um processo de saltos ou se aproxima de um por um limite [16]. Obtemos uma demonstração alternativa deste resultado resolvendo explicitamente a equação de Chapman-Kolmogorov. A solução é obtida em etapas, relaxando progressivamente as restrições sobre um processo simples, cuja solução é trivial, até contemplar processos autônomos arbitrários. A solução apresentada é menos abstrata e pode ser interpretada de maneira mais direta que os resultados anteriores. A desvantagem é que o resultado obtido é um pouco menos geral.

A equação mestra e a integral de trajetórias são duas formulações para determinar a probabilidade de transição de um processo Markoviano que essencialmente evitam atacar diretamente a equação de Chapman-Kolmogorov. O motivo para isto é óbvio: se trata de uma equação integral não-linear extremamente difícil de se resolver.

Neste capítulo nos limitamos a processos autônomos definidos pela probabilidade de transição $G_{t'}(x|x')$ depender apenas da diferença entre os tempos $\tau = t - t'$. Nesse caso, mostramos que é possível resolver a equação de Chapman-Kolmogorov explicitamente.

Para processos autônomos, a equação de Chapman-Kolmogorov se simplifica como

$$G_{\tau+\tau'}(x|x') = \int dy G_{\tau}(x|y)G_{\tau'}(y|x'), \quad (3.1)$$

onde $G_{\tau}(x|x')$ representa a probabilidade de transição de x' para x passado um tempo τ .

Iniciamos a exposição com um exemplo simples, que generalizamos em etapas. Na Sec. 3.1, consideramos que a probabilidade de transição é homogênea na variável de estado, ou seja, $G_{\tau}(x|x') = G_{\tau}(x - x')$. Essa simplificação permite obter uma solução analítica exata e servirá como apresentação de várias idéias importantes. Depois generalizamos para a situação onde a probabilidade de transição não é homogênea em x , mas é analítica em τ (Sec. 3.2). Finalmente, na Sec. 3.3, invocamos um resultado previamente conhecido devido à Feller [16] que diz que todo processo Markoviano autônomo pode ser aproximado por um processo de saltos.

3.1 Processos homogêneos

Considere processos Markovianos invariantes por translação tanto em x quanto em t . A probabilidade de transição se escreve como $G_{tt'}(x|x') = G_{\tau}(x - x')$, onde $\tau \equiv t - t'$. Assim, a equação de Chapman-Kolmogorov se torna uma convolução

$$G_{\tau'+\tau}(x) = \int dy G_{\tau'}(x - y)G_{\tau}(y), \quad (3.2)$$

que pode ser facilmente resolvida no espaço de Fourier. A convolução de distribuições de probabilidade descreve um processo de soma de variáveis aleatórias. Assim, $G_{\tau+\tau'}(x)$ é a distribuição da soma de duas variáveis aleatórias cujas distribuições são $G_{\tau}(x)$ e $G_{\tau'}(x)$.

Definimos a função característica (FC) de $G_{\tau}(x)$ como $\phi_{\tau}(z) = \int dx G_{\tau}(x)e^{izx}$.

Escrita para as funções características, a convolução (3.2) se simplifica,

$$\phi_{\tau'+\tau}(z) = \phi_{\tau'}(z)\phi_{\tau}(z). \quad (3.3)$$

A solução geral para a equação anterior é

$$\phi_{\tau}(z) = e^{\tau\gamma(z)}. \quad (3.4)$$

No entanto, a Eq. (3.4) ainda não representa a resposta final: somente as soluções que correspondem às distribuições de probabilidade nos interessam. É necessário garantir que a Eq. (3.4) é uma FC válida.

A Eq. (3.4) representa a FC de uma distribuição infinitamente divisível, já que equivale à soma de um número arbitrário de variáveis distribuídas de acordo com $\phi_{\tau/n}(z)$. Deste modo, a forma geral da FC em (3.4) é prescrita pelo teorema de De Finetti apresentado no capítulo anterior (página 2.1). Usando esse resultado, vemos que

$$\phi(z) = e^{\tau\lambda(\psi(z)-1)}, \quad (3.5)$$

onde λ é uma constante com unidade de frequência que corrige a dimensionalidade do expoente e $\psi(z)$ é uma FC arbitrária. Essa expressão garante que a solução da equação de Chapman-Kolmogorov de fato representa uma distribuição. Mais ainda, segundo o teorema de De Finetti, qualquer solução da Eq. (3.2) pode ser escrita como a Eq. (3.5) ou como um limite de expressões deste tipo.

Convém introduzir uma nomenclatura adicional. A Eq. (3.5) representa o que chamamos de uma *forma canônica de De Finetti*. Já as FC's que resultam de um limite a partir da Eq. (3.5) são o que chamamos de *formas limite de De Finetti*. A Gaussiana é um exemplo típico do segundo caso.

3.1.1 Dinâmica de saltos

O movimento Browniano de Einstein e os vôos de Lévy são dois exemplos de processos decorrentes de uma forma limite de De Finetti. Em ambos os casos, as características das trajetórias resultantes são muito bem compreendidas: ambas são contínuas e não diferenciáveis em quase todos os pontos. A pergunta que se coloca aqui é a seguinte: existe uma assinatura comum às trajetórias que decorrem de processos com a forma canônica de De Finetti?

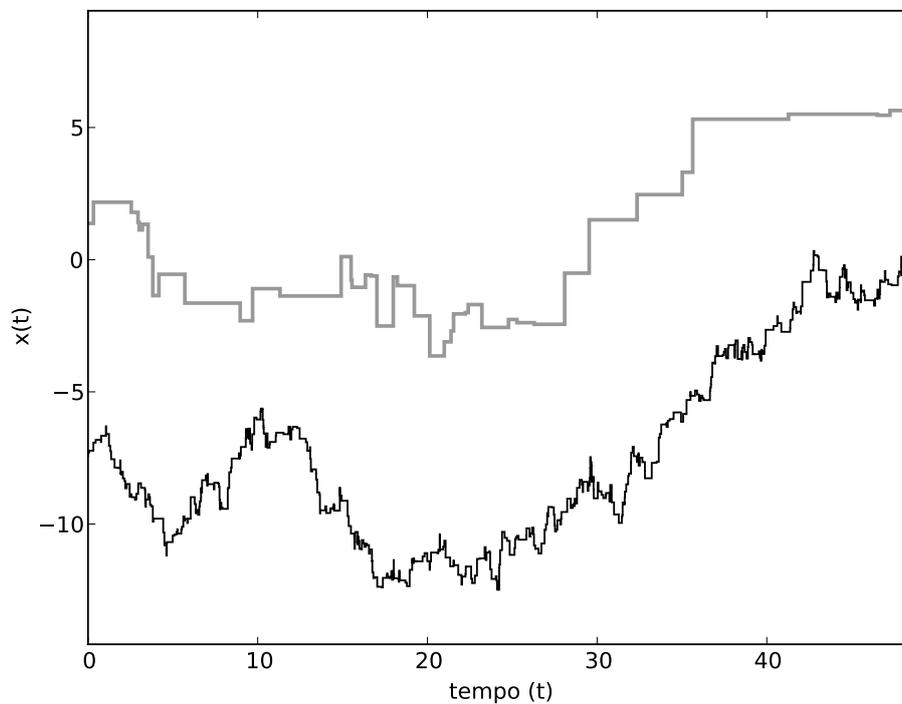


Figura 3.1: Processo de saltos em duas escalas de tempo distintas. A linha clara resulta de um processo com $\lambda = 1$ e saltos Gaussianos de variância 1 e a linha escura foi produzida de maneira semelhante com $\lambda = 10$ e variância 0.3. Em frequências altas e saltos pequenos as trajetórias se assemelham às trajetórias de um movimento Browniano.

A resposta é positiva: as formas canônicas de De Finetti definem processos que evoluem por saltos discretos, como mostrado na Fig. 3.1. Para ver isto,

expandimos a Eq. (3.5) em uma série de potências em torno de $\tau = 0$,

$$\phi_\tau(z) = e^{-\lambda\tau} \left[1 + \lambda\tau\psi(z) + \frac{(\lambda\tau)^2}{2!}\psi(z)^2 + \dots \right] \quad (3.6)$$

$$= \sum_{n=0}^{\infty} P(n|\tau) \psi(z)^n, \quad (3.7)$$

onde $P(n|\tau) = (\lambda\tau)^n / n!$. A equação para a probabilidade de transição é portanto

$$G_\tau(x) = \sum_{n=0}^{\infty} P(n|\tau)q(x|n), \quad (3.8)$$

onde $q(x|0) = \delta(x)$ e $q(x|1) = q(x)$ é a densidade de probabilidades associada à FC $\psi(z)$. As probabilidades $q(x|n)$ são definidas pelo produto de convolução

$$q(x|n) = \underbrace{q(x) * q(x) * \dots * q(x)}_{n \text{ vezes}}. \quad (3.9)$$

A Eq. (3.8) determina que, passado um tempo τ , existe uma probabilidade $P(n|\tau)$ de fazer uma transição segundo $q(x|n)$. É óbvio pela definição (3.9), que $q(x|n)$ é a probabilidade da soma de n variáveis distribuídas como $q(x)$. Em outras palavras, a evolução segundo $G_\tau(x)$ é decomposta em uma parte $P(n|\tau)$ que diz quantas transições foram feitas até o tempo τ e outra parte que diz o estado final após somar as n transições distribuídas segundo $q(x)$.

Uma maneira útil de conceber esse processo é utilizando a noção de tempo de sobrevivência. O termo $P(0|\tau) = e^{-\lambda\tau}$ determina a chance que nenhuma transição ocorra até o tempo τ . Queremos determinar a probabilidade $p(t) dt$ para que a primeira transição ocorra exatamente entre t e $t + dt$. É claro que $P(0|\tau)$ contém essencialmente esta informação já que corresponde à probabilidade de que o instante t em que ocorre a primeira transição seja maior que τ . Sendo assim, $1 - P(0|\tau)$ é a probabilidade cumulativa $\int_0^\tau dt p(t)$. Diferenciando esse

termo por τ obtemos a probabilidade desejada

$$p(\tau) = \frac{\partial}{\partial \tau} [1 - P(0|\tau)] = \lambda e^{-\lambda \tau}. \quad (3.10)$$

A simulação de um processo descrito pela Eq. (3.10) é então muito simples: (1) utilizando $p(\tau)$, sorteia-se a duração do intervalo de espera no qual o estado do sistema não se altera, (2) realiza-se uma transição de tamanho distribuído de acordo com $q(x)$, e (3) itera-se o processo indefinidamente.

3.2 Processos analíticos em τ

Os resultados obtidos na seção anterior dependem da condição de homogeneidade espacial que reduz a equação de Chapman-Kolmogorov a uma convolução. Nessa seção, relaxamos a restrição de homogeneidade espacial e obtemos a solução para qualquer processo autônomo. Esse é o principal resultado original deste capítulo.

Supomos que a probabilidade de transição $G_\tau(x|x')$ é analítica em τ . Deste modo, é razoável esperar que a equação de Chapman-Kolmogorov para $G_\tau(x|x')$ implique em relações tratáveis para os coeficientes da série de potências de $G_\tau(x|x')$. Antes de fazê-lo, convém escrever a probabilidade de transição como

$$G_\tau(x|x') = \sum_{n=0}^{\infty} P(n|\tau, x') q(x|x', n), \quad (3.11)$$

onde $q(x|x', 0) = \delta(x - x')$ e tanto $P(n|\tau, x')$ quanto $q(x|x', n)$ são probabilidades ainda indefinidas.

A probabilidade cumulativa $\int_{-\infty}^x dy G_\tau(y|x')$ possui necessariamente uma descontinuidade em $x = x'$ com uma amplitude que, em geral, depende tanto de τ quanto de x' . Essa descontinuidade se traduz em um termo proporcional a uma delta de Dirac na densidade $G_\tau(x|x')$. Obviamente, para $\tau = 0$ a des-

continuidade assume seu valor máximo de 1 já que $G_0(x|x') = \delta(x - x')$. A amplitude desta descontinuidade é a função de sobrevivência mencionada anteriormente, que aqui denotamos por $S_\tau(x')$. A suposição que $G_\tau(x|x')$ é analítica em τ implica que $S_\tau(x')$ também o é.

Considere a expansão em série de potências de

$$\frac{G_\tau(x|x')}{S_\tau(x')} = R_\tau(x, x') \equiv \sum_{n=0}^{\infty} c_n(x, x') \tau^n. \quad (3.12)$$

Cada coeficiente $c_n(x, x')$ deve ser normalizável para que $G_\tau(x|x')$ também o seja. Definimos $a_n(x') \equiv \int dx c_n(x, x')$ para introduzir os novos coeficientes $q(x|x', n) \equiv c_n(x, x')/a_n(x')$ e $P(n|\tau, x') \equiv S_\tau(x') a_n(x') \tau^n$. Estes são os mesmos coeficientes que aparecem na Eq. (3.11).

As probabilidades $q(x|x', n)$ são normalizadas por definição. Já a normalização de $P(n|\tau, x')$ é estabelecida facilmente ao integrar a Eq. (3.11) por x . Note que o fator $S_\tau(x')$ é comum a todos os termos $P(n|\tau, x')$ e portanto podemos reorganizar a condição de normalização como

$$S_\tau(x')^{-1} = \sum_{n=0}^{\infty} a_n(x') \tau^n. \quad (3.13)$$

Daqui segue que os coeficientes da expansão em série de potências de $S_\tau(x')^{-1}$ determinam completamente os coeficientes $a_n(x')$.

Dos vários processos descritos pela Eq. (3.11) (ex., Ref. [3]), a propriedade Markoviana surge apenas de famílias de probabilidades $P(n|\tau, x')$ e $q(x|x', n)$ muito específicas. Primeiramente, mostramos que a função de sobrevivência $S_\tau(x')$ de um processo Markoviano é necessariamente um decaimento exponencial. Posteriormente, mostramos que a seqüência de probabilidades $q(x|x', n)$ depende apenas da distribuição $q(x|x') \equiv q(x|x', 1)$.

3.2.1 Função de sobrevida

Toda distribuição de probabilidades pode ser fatorada de maneira unívoca em uma parte contínua e em outra parte descontínua (esta fatoração será tratada explicitamente na Sec. 4.2.3, vale observar que aqui nos referimos à continuidade da distribuição cumulativa de probabilidades). A parte descontínua consiste em uma mistura de funções delta de Dirac.

Consideramos distribuições que possuem pelo menos um ponto de descontinuidade em $x = x'$. Deste modo, é possível fatorar $G_\tau(x|x')$ como

$$G_\tau(x|x') = S_\tau(x')\delta(x - x') + H_\tau(x|x'), \quad (3.14)$$

onde $H_\tau(x|x')$ não possui nenhuma descontinuidade em $x = x'$, mas pode possuir uma descontinuidade em outros pontos. Substituindo (3.14) na equação de Chapman-Kolmogorov, obtemos

$$G_{\tau+\tau'}(x|x') = S_{\tau+\tau'}\delta(x - x') + H_{\tau+\tau'}(x|x'), \quad (3.15)$$

$$\begin{aligned} &= S_\tau(x')S_{\tau'}(x')\delta(x - x') + \int dy H_\tau(x|y)H_{\tau'}(y|x') \\ &+ S_\tau(x)H_{\tau'}(x|x') + S_{\tau'}(x')H_\tau(x|x'). \end{aligned} \quad (3.16)$$

Basta que a convolução $\int dy H_\tau(x|y)H_{\tau'}(y|x')$ não possua nenhuma descontinuidade em x' para que seja possível afirmar que

$$S_{\tau+\tau'}(x') = S_\tau(x')S_{\tau'}(x'). \quad (3.17)$$

Essa equação é satisfeita somente se

$$S_\tau(x') = e^{-\lambda(x')\tau}, \quad (3.18)$$

onde $\lambda(x')$ é uma função arbitrária de x' . Utilizando a relação (3.13), segue

imediatamente que

$$p(n|\tau, x') = \frac{(\lambda(x')\tau)^n}{n!} e^{-\lambda(x')\tau}. \quad (3.19)$$

Assim como no caso homogêneo, as probabilidades $p(n|\tau, x')$ definem um processo de Poisson. A diferença está essencialmente em que agora a taxa de transições $\lambda(x')$ pode depender do estado x' .

Existe uma lacuna nesta demonstração: é necessário definir em que condições o termo $\int dy H_\tau(x|y)H_{\tau'}(y|x')$ não resulta em uma descontinuidade em x' . Para isto, consideramos a maneira mais geral que um ponto de descontinuidade pode aparecer na probabilidade de transição $G_\tau(x|x')$, dada pela fatoração abaixo

$$G_\tau(x|x') = S_\tau(x')\delta(x - x') + \sum_{n=1}^N A_n(\tau, x')\delta(x - f_n(x')) + H_\tau^{(c)}(x|x'). \quad (3.20)$$

A função $H_\tau^{(c)}(x|x')$ representa a parte contínua da probabilidade de transição e as funções $f_n(x')$ e $A_n(\tau, x')$ são até certo ponto arbitrárias. Cada termo no somatório representa portanto uma transição de x' para $f_n(x')$ que ocorre com uma probabilidade $A_n(\tau, x')$. É possível mostrar que as funções $f_n(x')$ não podem depender de τ para que a propriedade Markoviana seja satisfeita.

A equação de Chapman-Kolmogorov implica em restrições fortes sobre a probabilidade de transição e, em especial, sobre as funções $f_n(x')$. Antes de proceder, é conveniente introduzir a notação $A_0(\tau, x') \equiv S_\tau(x')$ e $f_0(x') = x'$ e desconsiderar a parte contínua de $G_\tau(x|x')$ (ela pode ser facilmente recuperada, mas a desconsideramos por simplicidade do argumento). Substituindo (3.20) em (3.1), obtemos

$$G_{\tau+\tau'}(x|x') = \sum_{n=0}^N A_n(\tau + \tau', x')\delta(x - f_n(x')), \quad (3.21)$$

$$= \sum_{n=0}^N \sum_{m=0}^N A_n(\tau, f_m(x')) A_m(\tau', x')\delta(x - f_n(f_m(x'))). \quad (3.22)$$

Os pontos de descontinuidade de (3.21) obviamente devem ser os mesmos que (3.22). Fica claro, portanto, que o conjunto de funções $\{f_0, f_1, \dots, f_N\}$ deve ser fechado pela operação de composição ¹. Escrevendo explicitamente, temos

$$f_n \circ f_m = f_{M_{nm}}, \quad (3.23)$$

onde M_{nm} é a “tabela de multiplicação” do conjunto. Todos os termos nulos de M_{nm} identificam composições de funções $f_n \circ f_m = f_0$. Estes termos contribuirão para a forma da função de sobrevivência.

A suposição que fizemos no início da seção corresponde à imposição que o único termo nulo na tabela de multiplicação seja M_{00} (esse termo é sempre nulo já que $f_0 \circ f_0 = f_0$). Caso isto não se verifique, ao invés da Eq. (3.17), chegaremos a equações funcionais do tipo

$$S_{\tau+\tau'}(x') = S_\tau(x')S_{\tau'}(x') + F(\tau, \tau'), \quad (3.24)$$

onde $F(\tau, \tau')$ é a soma sobre todos os termos $A_i(\tau, f_j(x'))A_j(\tau', x')$ em que $f_i \circ f_j = f_0$, exceto $f_0 \circ f_0$ que fornece a contribuição $S_\tau(x')S_{\tau'}(x')$ já considerada em (3.24).

Ainda que seja possível definir consistentemente processos onde relações do tipo (3.24) aparecem, não iremos considerá-los na presente exposição. Em princípio, é possível utilizar um raciocínio análogo ao que apresentaremos para definir estes processos completamente. No entanto, ignoramos esta possibilidade deixando-a para um trabalho futuro. De agora em diante, assumimos que $S_\tau(x) = e^{-\lambda(x)\tau}$.

¹Usamos a notação corrente $(f \circ g)(x) = f(g(x))$.

3.2.2 Probabilidade de saltos

Uma vez que $P(n|\tau, x')$ for conhecida, as probabilidades $q(x|x', n)$ seguem de substituir a Eq. (3.11) na equação de Chapman-Kolmogorov (3.1), comparando ordem a ordem de $\tau^m \tau^n$. Em especial, obtemos uma relação de recorrência escolhendo os termos com $\tau \tau^n$ em ambos os lados da equação,

$$q(x|x', n+1) = \int dy K(x, y, x') q(y|x', n), \quad (3.25)$$

onde

$$K(x, y, x') = \left[1 - \frac{\lambda(y)}{\lambda(x')} \right] \delta(x - y) + \frac{\lambda(y)}{\lambda(x')} q(x|y). \quad (3.26)$$

Na relação anterior, definimos $q(x|x') = q(x|x', 1)$ para simplificar a notação. Se a taxa de transições *não* depender do estado, então a Eq. (3.25) assumiria a forma mais simples

$$q(x_n|x_0, n) = \int dx_1 \dots dx_{n-1} \prod_{i=1}^n q(x_i|x_{i-1}). \quad (3.27)$$

Processos com uma taxa de transição dependente do estado podem exibir difusão anômala [61, 30], sendo que alguns processos definidos em fractais também possuem esta propriedade [53]. No entanto, nos concentramos no caso em que a taxa de transição é constante, que resulta nas relações mais simples para as probabilidades $q(x|x', n)$ na Eq. (3.27).

Em uma série de casos, as probabilidades de transição de processos Markovianos autônomos descrevem processos de saltos e são dadas por

$$G_\tau(x|x') = \sum_{n=0}^{\infty} \frac{(\lambda(x')\tau)^n}{n!} e^{-\lambda(x')\tau} q(x|x', n). \quad (3.28)$$

Mostraremos por um argumento simples e genérico que mesmo os casos não contemplados na demonstração anterior podem ser escritos como um caso li-

mite de um processo de saltos [16]. No entanto, ao abordar a equação de Chapman-Kolmogorov explicitamente, como foi feito nesta seção, obtemos uma noção mais precisa de como cada termo em (3.28) deve ser interpretado e quais são as possibilidades contempladas em diferentes processos Markovianos.

3.3 Limite de processos de saltos

Considere operadores G_τ que agem em elementos $f(x)$ do espaço de funções integráveis como

$$G_\tau \cdot f(x) \equiv \int dy G_\tau(x|y) f(y). \quad (3.29)$$

Essa notação é especialmente conveniente pois a equação de Chapman-Kolmogorov vira uma definição de semigrupo

$$G_{\tau+\tau'} = G_\tau \cdot G_{\tau'}, \quad (3.30)$$

cujas solução geral é dada por

$$G_\tau = e^{\tau \bar{W}}, \quad (3.31)$$

onde \bar{W} é o gerador de evolução temporal. No entanto, talvez a Eq. (3.31) seja excessivamente geral: a positividade e normalização de $G_\tau(x|x')$ são preservadas apenas para geradores de evolução temporal muito específicos.

O operador de transição satisfaz a relação

$$G_\tau = (G_{\tau/m})^m, \quad (3.32)$$

para todo m finito. A equação (3.32) é exata, mas de uso muito limitado se ainda não se conhece G_τ . Para obter uma prescrição útil a partir da Eq. (3.32), consideramos uma aproximação $T_{\delta\tau} \simeq G_{\delta\tau}$ válida para $\delta\tau$ muito pequeno. O

operador aproximante $T_{\delta\tau}$ não precisa ser um processo Markoviano, mas deve ser semelhante o suficiente à $G_{\delta\tau}$ no sentido preciso que

$$\lim_{\delta\tau \rightarrow 0} \frac{T_{\delta\tau} - G_{\delta\tau}}{\delta\tau} = 0. \quad (3.33)$$

Uma escolha conveniente é $T_{\delta\tau} = 1 + \delta\tau \bar{W}$, que satisfaz a Eq. (3.33) — mas, naturalmente, existem várias outras possibilidades. De todo modo, se a Eq. (3.33) for satisfeita, podemos trocar $G_{\tau/n}$ por $T_{\tau/n}$ em (3.32). O erro é da ordem de $(T_{\delta\tau} - G_{\delta\tau})/\delta\tau$ e portanto desaparece no limite $\delta\tau \rightarrow 0$. De uma maneira bem formal, esse raciocínio corresponde ao formalismo de integrais de trajetória para processos Markovianos.

Reescrevemos \bar{W} , sem perda de generalidade, como

$$\bar{W} = \lambda(Q - 1), \quad (3.34)$$

onde λ é um parâmetro arbitrário com dimensão de $[\text{tempo}]^{-1}$ e Q é um operador. O aproximante $T_{\delta\tau}$ considerado anteriormente se torna

$$T_{\delta\tau} = (1 - \lambda\delta\tau) + (\lambda\delta\tau) Q, \quad (3.35)$$

que é simplesmente a definição infinitesimal de um processo de saltos. Assim, após substituir esse resultado em (3.32) e tomar o limite $m \rightarrow \infty$, ficamos com

$$G_\tau = \sum_{n=0}^{\infty} \frac{e^{-\lambda\tau}}{n!} (\lambda\tau)^n Q^n. \quad (3.36)$$

É óbvio que se Q representar uma probabilidade condicional válida, então $G_\tau(x|x')$ também o será. A Eq. (3.34) portanto fornece uma prescrição para definir um operador de evolução temporal \bar{W} que resulta em probabilidades de transição positivas. No entanto, não está claro se Eq. (3.36) é uma

representação geral de um processo Markoviano ou não.

É possível aproximar qualquer processo Markoviano autônomo por uma série convergente de processos de saltos. O limite mencionado é o seguinte:

$$G_\tau = \lim_{\lambda \rightarrow \infty} e^{\lambda\tau(G_{1/\lambda}-1)}, \quad (3.37)$$

$$= \lim_{\lambda \rightarrow \infty} \sum_{n=0}^{\infty} \frac{(\lambda\tau)^n e^{-\lambda\tau}}{n!} (G_{1/\lambda})^n. \quad (3.38)$$

A prova deste resultado se resume a mostrar que apenas os termos $n \simeq \lambda\tau$ contribuem para o somatório na Eq. (3.38). (Esse teorema é uma versão simplificada de um resultado encontrado em [16].) A média e o desvio padrão da distribuição de Poisson que pesa os fatores $(G_{1/\lambda})^n$ são respectivamente $\lambda\tau$ e $\sqrt{\lambda\tau}$. Assim, a massa de probabilidade é praticamente toda concentrada no intervalo $n_\pm = \lambda\tau \pm k\sqrt{\lambda\tau}$ para um número k suficientemente alto de desvios-padrão em torno da média². Deste modo, desprezamos os termos no somatório com $n \notin [n_-, n_+]$ pois estes contribuem com uma fração arbitrariamente pequena da massa de probabilidade total.

Para $\lambda \rightarrow \infty$, cada termo do somatório truncado converge para um termo proporcional à G_τ . Para verificar esse fato, observamos que

$$(G_{1/\lambda})^{n_\pm} = G_{n_\pm/\lambda} = G_{\tau \pm k\sqrt{\tau/\lambda}}, \quad (3.39)$$

que para $\lambda \rightarrow \infty$, claramente se reduz a G_τ . A série truncada converge para G_τ , fora um fator arbitrariamente próximo de 1 (a massa de probabilidade que a série carrega). Chamamos o termo em que $n = \lambda\tau$, de termo principal da série. (3.38). Esse argumento estabelece que, à parte de uma vizinhança arbitrariamente próxima do termo principal, todas as outras contribuições à

²Uma colocação mais precisa desta afirmação pode ser feita invocando a desigualdade de Chebyshev. Assim, estabelecemos que a massa de probabilidade incluída nesse intervalo é maior ou igual a $1 - 1/k^2$. Para um limite mais preciso, também seria possível calcular explicitamente a distribuição de quantis da probabilidade de Poisson às custas de uma forma matemática mais elegante.

série (3.38) desaparecem para $\lambda \rightarrow \infty$.

3.4 Conclusão

Em um processo Markoviano, deve-se garantir que a probabilidade de transição obedece à equação de Chapman-Kolmogorov. Existem algumas maneiras de definir modelos que satisfazem esta restrição, onde já mencionamos as integrais de trajetória e as equações mestras. Nos dois casos, a resposta é obtida respectivamente pela solução de uma integral funcional ou de uma equação integro-diferencial. Esse passo, de escrever a probabilidade de transição como a solução de um problema matemático específico não é estritamente necessário.

Resolvemos a equação de Chapman-Kolmogorov de processos autônomos supondo que a probabilidade de transição de $x(t)$ para $x(t + \tau)$ é analítica em τ . A solução obtida representa um processo de saltos, que é parametrizado pela probabilidade de cada salto e uma taxa de transições. Feller mostrou que todo processo Markoviano pode ser aproximado por um processo de saltos no limite em que a taxa de transições $\lambda \rightarrow \infty$. Para que isto seja verdade, a probabilidade de saltos deve ser definida como

$$q(x|x') = G_{1/\lambda}(x|x'). \quad (3.40)$$

Esse resultado é denominado de fórmula exponencial para os processos Markovianos. O resultado que obtivemos se superpõe com o teorema de Feller em alguns aspectos, mas as duas demonstrações não são inteiramente equivalentes.

Nossa preocupação foi obter explicitamente as soluções da equação de Chapman-Kolmogorov. A solução geral obtida permite-nos construir explicitamente todos processos Markovianos que satisfazem as hipóteses da nossa demonstração. O teorema de Feller, ou a fórmula exponencial do processo Markoviano, não faz algumas destas hipóteses (como a analiticidade em τ) e

nesse sentido é mais geral que o nosso. No entanto, não podemos encará-lo como a solução geral da equação de Chapman-Kolmogorov. A fórmula exponencial mostra como criar *aproximações de um dado processo Markoviano* como um processo de saltos.

A fórmula exponencial do processo Markoviano justifica a suposição de analiticidade que fizemos anteriormente. Ainda que nem todo processo Markoviano autônomo seja analítico, é sempre possível construir uma aproximação analítica tão semelhante quanto se queira do processo estocástico original.

Capítulo 4

Covariância, ergodicidade e difusão

As diferentes condições de mistura tentam capturar a noção de uma variável estocástica que “esquece” seu valor inicial. Já a ergodicidade determina se as médias temporais são iguais às médias de ensemble. Mostramos que duas condições de mistura (uma fraca e outra forte) implicam respectivamente duas condições de ergodicidade (uma fraca e outra forte). Nosso resultado está em desacordo com um resultado recente de Lee [42] e é consistente com um teorema que Khinchin demonstrou para sistemas clássicos [33].

Morgado, et. al, mostraram como caracterizar o expoente de difusão de uma partícula descrita pela equação de Langevin generalizada a partir da densidade espectral da força aleatória[49]. Reformulamos esse problema sem mencionar a força aleatória. Ainda que os resultados se assemelhem, o critério apresentado envolve apenas a densidade espectral do processo, enquanto que o de Morgado utiliza transformadas de Laplace.

Considere um copo de água pura. Em princípio, não há muito o que observar: a água possui uma temperatura, volume, densidade, bem definidos e com

umas poucas variáveis é possível especificar as suas características mais relevantes. Essa simplicidade obviamente desaparece na descrição microscópica: são cerca de 10^{23} moléculas fazendo movimentos complicados, trombando umas com as outras, etc. Fundamentalmente, esta interação complexa determina aqueles poucos graus de liberdade que nos interessam no sistema macroscópico. No entanto, transitar de um nível de descrição para outro é muitas vezes impraticável.

Para fazer a ponte com a dinâmica microscópica é conveniente adotar uma abordagem estocástica. Ao invés de considerar o estado de todas as partículas que formam o sistema, é possível obter várias informações úteis considerando apenas o estado de uma destas partículas e tratar o resto do sistema como uma influência aleatória sobre esta partícula. Para evitar uma discussão muito abstrata, denotamos a variável de interesse de $u(t)$ e assumimos que se trata da velocidade da partícula selecionada.

Muitas vezes, especificar o processo estocástico *exato* relacionado a $u(t)$ pode ser tão complicado quanto desenvolver a dinâmica microscópica explicitamente. Felizmente, várias informações de interesse podem ser obtidas sem que seja necessário especificar um processo estocástico completamente. Em alguns casos, apenas determinar os primeiros momentos estatísticos destas distribuições pode ser o suficiente. Este capítulo trata de processos definidos parcialmente apenas pela média e pela covariância da variável estocástica.

4.1 Matrizes de covariância

4.1.1 Definição

Considere a probabilidade a dois pontos $P(u_t, u_{t'}) \equiv f_{tt'}(u, u') du du'$. A probabilidade marginal $P(u_t) \equiv f_t(u) du$ é definida como sendo $f_t(u) = \int du' f_{tt'}(u, u')$

ou $f_{t'}(u') = \int du f_{tt'}(u, u')$. A média $\langle u(t) \rangle$ corresponde a

$$\mu(t) = \int du f_t(u) u; \quad \mu(t') = \int du' f_{t'}(u'). \quad (4.1)$$

Usando a probabilidade a 2-pontos, escrevemos a matriz de covariância do processo. O segundo momento da distribuição é

$$\langle u(t)u(t') \rangle = \int du du' f_{tt'}(u, u') u u' \quad (4.2)$$

e a covariância é dada explicitamente como

$$\Sigma(t, t') \equiv \langle u(t)u(t') \rangle - \mu(t)\mu(t'). \quad (4.3)$$

Na presente discussão, estamos interessados em processos estacionários, que são aqueles em que $f_t(u)$ não depende do tempo e todas as probabilidades de N -pontos são invariantes por deslocamento temporal (e.g., $f_{tt'}(u, u') = f_{t+\tau, t'+\tau}(u, u')$). A motivação física é que os resultados não sejam explicitamente dependentes da escolha do instante inicial $t = 0$. Espera-se que esta simetria seja violada apenas quando houver uma influência externa que dependa do tempo ou na descrição do regime transiente de um estado fora do equilíbrio.

A condição de estacionariedade implica que a média $\mu(t)$ não depende do tempo e a matriz de covariância $\Sigma(t, t')$ depende apenas da diferença entre os tempos. A partir destas duas condições não é possível deduzir a estacionariedade de um processo estocástico, então seria mais correto dizer que os sistemas aqui considerados são *fracamente estacionários*. Por fracamente estacionário, é lógico, entendemos que $\mu(t) = \text{cte}$ e $\Sigma(t, t') = \Sigma(t - t')$.

4.1.2 Propriedades matemáticas

É possível identificar dois efeitos independentes codificados na matriz de covariância: são a correlação e a flutuação. A correlação linear, definida como

$$\chi(t, t') = \frac{\Sigma(t, t')}{\sigma(t)\sigma(t')}; \quad \sigma(t) \equiv \sqrt{\Sigma(t, t)}, \quad (4.4)$$

fornece uma idéia da similaridade da variável $u(t)$ avaliada em tempos diferentes. O caso extremo em que correlação entre $u(t)$ e $u(t')$ é ± 1 , significa que existe uma relação linear entre $u(t)$ e $u(t')$. Já o desvio padrão $\sigma(t)$ fornece uma medida da flutuação da variável $u(t)$ no instante t . Tipicamente, $u(t)$ permanecerá dentro do intervalo $\mu \pm \sigma(t)$, portanto quanto maior $\sigma(t)$, maior será o intervalo de valores em que $u(t)$ costuma percorrer. (Podemos reformular esta afirmação de maneira mais precisa usando a desigualdade de Chebyshev: a probabilidade que $u(t)$ esteja até k desvios-padrão longe da média é maior ou igual à $1 - 1/k^2$ para qualquer distribuição.)

Nem toda função do tipo $\Sigma(t, t')$ é uma matriz de covariância válida. É óbvio, pela definição da covariância na Eq. (4.3), que $\Sigma(t, t')$ deve ser simétrica nos argumentos t e t' . Existe, no entanto, uma condição bem mais restritiva. A matriz de covariância deve ser positivo-definida. Essa propriedade será apresentada na seção que se segue.

Matrizes de covariância

Definimos a matriz $M_{ij} \equiv \Sigma(t_i, t_j)$, onde $\{t_1, \dots, t_n\}$ é uma malha temporal arbitrária. O processo estocástico para $u(t)$ prescreve a probabilidade conjunta $P(u_1, \dots, u_n)$ de se obter qualquer vetor $\mathbf{u} = (u_1, \dots, u_n)$ nesta malha temporal. Para efeito da presente discussão, consideramos momentaneamente que esse processo seja conhecido e que seja possível sortear amostras \mathbf{u}_i a partir de $P(\mathbf{u}) \equiv P(u_1, \dots, u_n)$.

Supomos por simplicidade que a média sobre esse *ensemble* é nula. A covariância de um *ensemble* com N amostras fica então dada por

$$M_{ij} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N (\mathbf{u}_k)_i (\mathbf{u}_k)_j, \quad (4.5)$$

onde $(\mathbf{u}_k)_i$ denota a i -ésima componente do k -ésimo vetor. Podemos reescrever a relação anterior de maneira mais elegante definindo o vetor $\mathbf{w}_k = \frac{1}{\sqrt{N}} (u_{k1}, u_{k2}, \dots, u_{kn})$. Assim, $\frac{1}{N} \sum_{k=1}^N u_{ik} u_{jk} = \langle \mathbf{w}_i, \mathbf{w}_j \rangle$, onde $\langle \cdot, \cdot \rangle$ é o produto escalar definido usualmente. Generalizando a idéia anterior, vemos que M_{ij} é uma matriz de covariância se e somente se puder ser escrita como

$$M_{ij} = \langle \mathbf{w}_i, \mathbf{w}_j \rangle, \quad (4.6)$$

onde $\{\mathbf{w}_i\}$ é um conjunto de vetores com dimensões arbitrárias.

A propriedade anterior é uma das maneiras como os matemáticos definem uma *matriz positivo-definida*. Na realidade, existe uma série de propriedades equivalentes que cumprem o mesmo efeito. Deste modo, a matriz M será positivo-definida¹ se e somente se:

1. Vale a desigualdade $\mathbf{v}^\dagger M \mathbf{v} \geq 0$ para todo \mathbf{v} (estritamente positiva se $\mathbf{v}^\dagger M \mathbf{v} = 0$ apenas para $\mathbf{v} = 0$).
2. Uma matriz M é (estritamente)-positivo-definida se for Hermitiana e todos seus autovalores forem (estritamente) positivos.
3. M pode ser decomposta como $M_{ij} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle$, onde $\{\mathbf{v}_i\}$ é um conjunto de vetores e $\langle \cdot, \cdot \rangle$ denota um produto interno. A matriz resultante será *estritamente-positivo-definida* se o conjunto de vetores for linearmente independente.

¹Note que alguns autores utilizam uma convenção diferente: usam “positivo-semi-definido” e “positivo-definido” quando nós utilizamos “positivo-definido” e “estritamente positivo-definido”.

4. Existe uma noção semelhante à de raiz quadrada de uma matriz positivo definida. Ela é dada pela decomposição de Cholesky, e significa que existe uma matriz triangular superior R tal que

$$M = R^\dagger R. \quad (4.7)$$

(Equivalentemente, existe uma matriz triangular *inferior* L tal que $M = LL^\dagger$). Os elementos diagonais de R e L são maiores que zero no caso estritamente positivo, além da decomposição de Cholesky ser unívoca.

5. Todos menores principais de M possuem determinante positivo ou nulo. Os menores principais são formados removendo um número arbitrário de linhas e as correspondentes colunas da matriz M (i.e.: se removermos a segunda linha é necessário remover a segunda coluna). Determinantes estritamente positivos definem uma matriz estritamente-positivo-definida.

No caso estritamente positivo, a forma bilinear $\langle \mathbf{v}, \mathbf{u} \rangle \equiv \mathbf{v}^\dagger M \mathbf{u}$ define um produto interno.

Operadores positivo-definidos

A relação entre operadores positivo-definidos e matrizes positivo-definidas é bastante direta. Definimos a matriz de covariância $M_{ij} = \Sigma(t_i, t_j)$ sobre uma malha $\{t_1, \dots, t_n\}$. Dizemos que o operador $\Sigma(t, t')$ é positivo definido se todas matrizes de covariância M_{ij} definidas pela prescrição anterior sobre qualquer malha temporal forem positivo-definidas[57].

Podemos inferir algumas propriedades de $\Sigma(t, t')$ diretamente desta definição. Escolhendo uma malha com apenas um tempo, vemos que $\Sigma(t, t) \geq 0$ (podemos utilizar a propriedade (2), por exemplo, para estabelecer isto). Escolhendo $\{t, t'\}$, junto com a propriedade (5), mostramos imediatamente que

$\Sigma(t, t) \geq |\Sigma(t, t')|$. Observe que M é Hermitiana e real, o que implica em $\Sigma(t, t') = \Sigma(t', t)$. Estas propriedades são úteis, mas obviamente não são o suficiente para assegurar a positividade de um operador.

Similarmente ao caso anterior, existem diferentes maneiras de caracterizar um operador positivo-definido que espelham as propriedades das matrizes positivo-definidas. Assim, considere o operador Σ que atua em um elemento $g(t)$ do espaço de funções,

$$(\Sigma \cdot g)(t) \equiv \int d\mu(t') \Sigma(t, t') g(t'). \quad (4.8)$$

O operador de covariância é positivo-definido se e somente se $g \cdot \Sigma \cdot g \geq 0$ para qualquer função g . Similarmente, um operador é *estritamente positivo-definido* se $g \cdot \Sigma \cdot g = 0$ apenas para $g = 0^2$.

Funções positivo-definidas: teorema de Bochner

Toda função $f(\tau)$ define um núcleo pela prescrição $K(t, t') = f(t - t')$. Dizemos que uma função $f(\tau)$ é positivo-definida se $K(t, t')$ representar o núcleo de um operador positivo definido. Segundo um teorema devido à Bochner [45], uma função é positivo-definida se e somente se puder ser escrita como

$$f(\tau) = \int d\nu(\omega) e^{-i\omega\tau}, \quad (4.9)$$

onde $\nu(\omega)$ é uma medida.

Nos casos em que $\mu(\omega)$ possui uma densidade, $d\nu(\omega) = \tilde{\Sigma}(\omega)d\omega$, podemos escrever

$$\Sigma(\tau) = \int d\omega \tilde{\Sigma}(\omega) \frac{e^{-i\omega\tau}}{2\pi}; \quad \tilde{\Sigma}(\omega) = \int d\tau \Sigma(\tau) e^{i\omega\tau}, \quad (4.10)$$

onde identificamos $\Sigma(\tau) = f(\tau)$. Esse resultado é conhecido como o teorema

²A Eq. (4.8) é uma integral de Stieltjes sobre a medida $\mu(t)$. Caso $\mu(t)$ possua uma densidade, então $d\mu(t) = \frac{d\mu}{dt} dt$, e a integral se reduz à integral de Riemann usual.

de Wiener-Khinchin no contexto de processos estocásticos, mas simplesmente reproduz o teorema de Bochner na teoria geral de operadores lineares. Para obter um processo estocástico bem definido, é necessário que a variância seja finita, ou seja,

$$\Sigma(0) = \frac{1}{2\pi} \int d\omega \tilde{\Sigma}(\omega) < \infty. \quad (4.11)$$

Veremos adiante que esta propriedade determina se as trajetórias $u(t)$ são contínuas em quase todos os pontos ou não.

No contexto mais específico de matrizes de covariância, adiantamos as seguintes conclusões. Para que $\Sigma(t - t')$ seja uma matriz de covariância positivo definida, basta que sua transformada de Fourier seja positiva, ou seja $\tilde{\Sigma}(\omega) \geq 0$ (o que também garante que a matriz de covariância é Hermitiana). Além disso, para que a matriz de covariância seja real, $\tilde{\Sigma}(\omega)$ deve ser uma função par. Enfim, $\tilde{\Sigma}(\omega)$ é uma distribuição simétrica (e geralmente normalizável) das frequências ω .

Análise de autovalores

A equação de autovalores para um núcleo positivo-definido $\Sigma(t, t')$ com respeito a uma medida $\nu(t)$ se escreve como

$$\int d\nu(t') \Sigma(t, t') \phi_i(t') = \lambda_i \phi_i(t). \quad (4.12)$$

Assumimos que os autovalores são ordenados ($\lambda_1 \geq \lambda_2 \geq \dots \geq 0$) e que as auto-funções são normalizadas ($\int d\nu(t) \phi_i(t) \phi_j(t) = \delta_{ij}$). O teorema espectral (ou teorema de Mercer), diz que o operador $\Sigma(t, t')$ pode ser escrito como

$$\Sigma(t, t') = \sum_{i=1}^{\infty} \lambda_i \phi_i(t) \phi_i^*(t') \quad (4.13)$$

e que os autovalores são absolutamente somáveis, ou seja, $\sum_{i=1}^{\infty} |\lambda_i|$ não diverge.

Um fato interessante a respeito dos resultados acima é que tanto a decomposição espectral quanto os próprios autovalores dependem da medida $\nu(t)$. O espectro de autovalores torna-se contínuo se a decomposição é feita sobre uma medida infinita ($\int d\nu(t) = \infty$). É possível obter trivialmente o espectro de autovalores de um núcleo estacionário a partir do teorema de Bochner. Escrevemos

$$\Sigma(t - t') = \int d\mu(\omega) \frac{e^{-i\omega(t-t')}}{2\pi} = \int d\mu(\omega) \frac{e^{i\omega t}}{\sqrt{2\pi}} \left(\frac{e^{i\omega t'}}{\sqrt{2\pi}} \right)^*, \quad (4.14)$$

que, comparando com a Eq. (4.13), demonstra que as autofunções são do tipo ondas planas e o espectro de autovalores é dado por $d\mu(\omega)$. Se $\mu(\omega)$ possui uma densidade, $d\mu(\omega) = \tilde{\Sigma}(\omega)d\omega$, então convém utilizar a *densidade espectral* $\tilde{\Sigma}(\omega)$ como uma representação do núcleo estacionário.

Funções características e covariância

É conveniente explorar a relação entre funções características e as matrizes de covariância de processos estacionários. Matematicamente, as condições sobre ambas são muito semelhantes, onde a maior dificuldade consiste em garantir o caráter positivo-definido. Primeiramente, existe a diferença trivial que a função característica deve ser igual à 1 na origem para impor a normalização das probabilidades. As matrizes de covariância podem assumir qualquer valor positivo, que identificamos com $\Sigma(0) = \frac{k_B T}{m}$. Mais importante é o fato que matrizes de covariância devem ser Hermitianas e reais, enquanto as funções características não exigem tais propriedades.

A parte real de uma função característica é sempre par e sua parte imaginária é ímpar. Para que uma função característica seja real, é necessário que ela defina uma lei de probabilidades simétrica. Deste modo, as matrizes de covariância são sempre proporcionais às funções características de variáveis

aleatórias simétricas.

Escrevemos explicitamente a matriz de covariância como

$$\Sigma(\tau) = \Sigma(0)\phi(\tau), \quad (4.15)$$

onde $\phi(\tau)$ é a função característica de uma distribuição simétrica. De maneira análoga, temos a densidade espectral

$$\tilde{\Sigma}(\omega) = \Sigma(0)f(\omega), \quad (4.16)$$

onde $f(x)$ é a densidade de probabilidades da distribuição definida por $\phi(\tau)$. Esse resultado simples é de enorme conveniência: ele nos permite utilizar a vasta literatura sobre funções características no estudo de matrizes de covariância e vice-versa.

Mostramos alguns exemplos de distribuições simétricas cuja função característica é conhecida em forma fechada. Note que em todos os casos $s > 0$ é um parâmetro de escala e $\nu > 0$ é um parâmetro de forma. (Ver Apêndice 8 para a definição das funções especiais).

Também podemos utilizar distribuições definidas apenas para argumentos positivos e fazer uma extensão par das mesmas. A função característica desta nova variável é simplesmente a parte real da função característica original. Aqui mostramos alguns exemplos deste tipo (sendo que a extensão par não foi realizada).

Núcleos integrais e transformações lineares

A combinação linear de variáveis estocásticas define outra variável estocástica. Considere um processo $x(t)$ com média $\mu(t)$ e covariância $\Sigma(t, t')$. Definimos um segundo processo $y(t)$ de acordo com a relação

Distribuição	Densidade $f(x)$	Função Característica $\phi(z)$
Gaussiana	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2s^2}}$	$e^{-\frac{1}{2}s^2z^2}$
Student-t/Matérn	$\frac{\Gamma(\nu+\frac{1}{2})}{s\sqrt{2\pi\nu}\Gamma(\nu)} \left(1 + \frac{x^2}{2\nu s^2}\right)^{-\nu-\frac{1}{2}}$	$\frac{K_\nu(\sqrt{2\nu} sz)(\sqrt{2\nu} sz)^{\frac{\nu}{2}}}{\Gamma(\nu)2^{\nu-1}}$
Cauchy/Lorentz	$\frac{s}{\pi[x^2+s^2]}$	$e^{-s z }$
Laplace/Exponencial	$\frac{1}{s} e^{- x /s}$	$\frac{1}{1+(sz)^2}$
Lévy estáveis ³	(sem forma fechada)	$e^{- sz ^\alpha}$
Retangular ⁴	$\frac{1}{s} \text{rect}\left(\frac{x}{s}\right)$	$\frac{2}{sz} \sin\left(\frac{1}{2}sz\right)$
Triangular ⁵	$\frac{1}{s} \text{tri}\left(\frac{x}{s}\right)$	$\frac{2}{(sz)^2} (1 - \cos(sz))$
Logística	$\frac{\pi}{4s} \text{sech}^2\left(\frac{\pi x}{2s}\right)$	$\frac{sz}{\sinh(sz)}$
Secante Hiperbólica	$\frac{1}{2s} \text{sech}\left(\frac{\pi x}{2s}\right)$	$\text{sech}(sz)$
Semi-círculo	$\frac{2}{\pi s^2} \sqrt{s^2 - x^2}$	$2 \frac{J_1(sz)}{sz}$
Cosseno truncado ⁶	$\frac{1}{2s} \left[1 + \cos\left(\pi \frac{x}{s}\right)\right]$	$\frac{\pi^2 \sin(sz)}{st(\pi^2 - (sz)^2)}$
Log-truncado ⁷	$-\frac{1}{s} \log\left(\frac{ x }{s}\right)$	$\frac{\text{Si}(kz)}{kz}$

Tabela 4.2: Funções características de distribuições simétricas

Distribuição	Densidade $f(x)$	Função Característica $\phi(z)$
Gama	$\frac{x^{\nu-1}}{s^\nu \Gamma(\nu)} e^{-x/s}$	$(1 - isz)^{-\nu}$
Gama inversa	$\frac{s^\nu}{x^{\nu+1} \Gamma(\nu)} e^{-s/x}$	$\frac{2(-isz)^{\nu/2}}{\Gamma(\nu)} K_\nu(2\sqrt{-isz})$
Lévy	$\sqrt{\frac{s}{2\pi}} \frac{e^{-\frac{s}{x}}}{x^{3/2}}$	$e^{-\sqrt{-2i}sz}$
Wald ⁸	$\sqrt{\frac{\mu^2}{\pi s x^3}} e^{-\frac{(x-\mu)^2}{s x}}$	$\exp\left[\frac{\lambda}{\mu} (1 - \sqrt{1 - isz})\right]$
Weibull	$\frac{\nu x^{\nu-1}}{s^\nu} e^{-(x/s)^\nu}$	$1 + \sum_{n=0}^{\infty} \Gamma\left(\frac{n+1}{\nu}\right) \frac{(isz)^{n+1}}{\nu n!}$
Pareto	$\Theta(x-s) \frac{\nu s^\nu}{x^{\nu+1}}$	$\nu (-isz)^\nu \Gamma(-\nu, -isz)$
Rayleigh	$\frac{x e^{-\frac{x^2}{2s^2}}}{s^2}$	$1 + i\sqrt{\frac{\pi}{2}} sz e^{-\frac{(sz)^2}{2}} \left[\text{erf}\left(\frac{sz}{\sqrt{2}}\right) + 1\right]$
χ^2 não-central	$\frac{e^{-\frac{x+\mu}{2s}}}{2s} \left(\frac{x}{\mu}\right)^{\frac{\nu-1}{2}} I_{\nu-1}\left(\frac{\sqrt{\mu x}}{s}\right)$	$(1 - 2isz)^{-\nu} e^{\frac{i\mu z}{1-2sz}}$ para $z < \frac{1}{2s}$
Beta	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{x^{\alpha-1}(s-x)^{\beta-1}}{s^{\alpha+\beta-1}}$	${}_1F_1(\alpha; \alpha + \beta; isz)$

Tabela 4.4: Funções características de distribuições definidas na reta real

$$y(t) = \int dt' g(t, t')x(t') + a(t). \quad (4.17)$$

É simples mostrar que a média e covariância da variável transformada se escrevem explicitamente como

$$\bar{\mu}(t) = \int dt' g(t, t')x(t') + a(t), \quad (4.18)$$

$$\bar{\Sigma}(t, t') = \int ds ds' g(t, s)g(t', s')\Sigma(s, s'). \quad (4.19)$$

Além destas propriedades, também temos a correlação cruzada

$$\langle\langle x(t)y(t') \rangle\rangle = \int ds' g(t', s')\Sigma(t, s'), \quad (4.20)$$

que é útil em alguns contextos.

É interessante considerar o caso em que tanto o processo $x(t)$ quanto os núcleos integrais dependem apenas da diferença $\tau = t - t'$. Considere as transformadas de Fourier definidas como $\tilde{g}(\omega) = \int dt g(t)e^{i\omega t}$ e $\tilde{\Sigma}(\omega) = \int dt \Sigma(t)e^{i\omega t}$ ⁹. Um cálculo simples a partir de (4.19) resulta em

$$\bar{\Sigma}(\tau) = \int d\omega |\tilde{g}(\omega)|^2 \tilde{\Sigma}(\omega) \frac{e^{-i\omega\tau}}{2\pi}. \quad (4.21)$$

Essa expressão é particularmente útil porque a estacionariedade é uma propriedade esperada em vários processos físicos. De maneira similar, a covariância cruzada fica

$$\int d\tau \langle\langle x(t)y(t+\tau) \rangle\rangle e^{i\omega\tau} = \tilde{g}(\omega)\tilde{\Sigma}(\omega). \quad (4.22)$$

⁹No caso comum em que $g(t)$ e $x(t)$ não são de quadrado integrável, é necessário truncar o domínio de integração em um intervalo $t \in [-T, T]$ e trabalhar com os resultados no limite de T arbitrariamente grande. Para alguns processos estocásticos, esse limite pode não ser rigorosamente bem definido.

Continuidade e diferenciabilidade

Existem diferentes critérios para avaliar se um processo estocástico é contínuo ou não. A motivação é inferir a continuidade das trajetórias aleatórias, o que infelizmente é um problema matemático muito difícil. Apresentamos alguns critérios simples que, em certos casos (como processos Gaussianos) permitem acessar a continuidade das trajetórias decorrentes do processo estocástico.

Dizemos que um processo estocástico é contínuo no sentido quadrático médio se o limite abaixo existe

$$\lim_{t' \rightarrow t} \langle [x(t') - x(t)]^2 \rangle = 0. \quad (4.23)$$

Essa é uma condição de continuidade relativamente forte, ainda que não garanta totalmente a continuidade das amostras de $x(t)$ ¹⁰. De todo modo, a continuidade quadrática média é a única forma conveniente para nossos propósitos: como a probabilidade a N -pontos não foi fornecida, a continuidade quadrática média é o máximo que se pode dizer sobre a continuidade ou descontinuidade das trajetórias.

Resolvendo explicitamente a Eq. (4.23), ficamos com

$$\begin{aligned} \lim_{t \rightarrow t'} \langle [x(t') - x(t)]^2 \rangle &= \lim_{t \rightarrow t'} [\langle x(t)^2 \rangle + \langle x(t')^2 \rangle - 2 \langle x(t)x(t') \rangle], \quad (4.24) \\ &= \lim_{t \rightarrow t'} [\Sigma(t, t) + \mu(t)^2 + \Sigma(t', t') + \mu(t')^2 - 2\Sigma(t, t') - 2\mu(t)\mu(t')]. \quad (4.25) \end{aligned}$$

Onde vemos que a relação é satisfeita somente se média for uma função contí-

¹⁰É possível mostrar que (4.23) implica que as funções $x(t)$ são “contínuas em quase todos os pontos”. No entanto, um processo contínuo no sentido quadrático médio ainda pode ser descontínuo na acepção comum: basta, para isto, que a trajetória $x(t)$ possua uma quantidade enumerável de pontos de descontinuidade — ou melhor, que os pontos de descontinuidade formem um conjunto de medida nula em t [59]. A violação da Eq. (4.23) não garante que o processo estocástico seja descontínuo em quase todos os pontos, apesar de isto ser verdade em alguns casos, como em processos Gaussianos.

nua em t e se $\lim_{t \rightarrow t'} \Sigma(t, t')$ convergir para um valor finito (se isto for verdade, então é possível mostrar que $\Sigma(t, t')$ é contínua em todos os pontos).

Partindo da condição de continuidade mencionada anteriormente, é possível determinar a diferenciabilidade de um processo estocástico $x(t)$. Dizemos que um processo $x(t)$ é diferenciável se o processo estocástico $y(t) = \frac{dx}{dt}$ for contínuo. Deste modo, é necessário que a média e a covariância de $y(t)$,

$$\bar{\mu}(t) = \frac{d\mu}{dt}; \quad \bar{\Sigma}(t, t') = \frac{d^2}{dt dt'} \Sigma(t, t'), \quad (4.26)$$

sejam funções contínuas em todos os pontos t e t' . De maneira similar, a diferenciabilidade até n -ésima ordem decorre da continuidade de $\frac{d^n x}{dt^n}$, o que depende de $\frac{d^n \mu}{dt^n}$ e $\frac{d^{2n}}{dt^n dt'^n} \Sigma(t, t')$ serem funções contínuas dos seus argumentos.

4.2 Ergodicidade e condição de mistura

Várias definições não-equivalentes de ergodicidade e mistura existem na literatura [6]. Definimos uma condição de mistura forte e outra condição fraca e mostramos que cada uma delas é suficiente para estabelecer a ergodicidade (fraca e forte) de um processo estacionário.

Em um trabalho recente, Lee [42] apontou que a condição de mistura linear *não* é uma condição suficiente para um sistema físico ser ergódico. Esse resultado contradiz um teorema devido a Khinchin[33], ainda que o teorema de Khinchin não se aplique ao caso considerado por Lee. Obtemos uma condição de ergodicidade consistente com o resultado de Khinchin. Analisando sistemas descritos pela equação de Langevin generalizada, Lapas, *et. al.* [40] também chegaram à mesma conclusão por nós obtida.

O teorema de Khinchin foi formulado originalmente para sistemas Hamiltonianos clássicos. Ele mostra que se a correlação de uma variável dinâmica

tender à zero,

$$\lim_{\tau \rightarrow \infty} \langle\langle f(t)f(t+\tau) \rangle\rangle = 0, \quad (4.27)$$

é possível concluir que $f(t)$ é ergódico. A média poderia então ser calculada sobre o ensemble microcanônico¹¹. Sendo um resultado puramente clássico, em princípio não há porque esse teorema deva ser satisfeito em um processo estocástico arbitrário. No entanto, mostramos que as mesmas conclusões do teorema de Khinchin se aplicam a qualquer processo estocástico estacionário, o que contradiz a conclusão de Lee.

4.2.1 Ergodicidade fraca

Um sistema é ergódico se as médias de *ensemble* se igualam às médias temporais. É razoável que um sistema que viole a condição de mistura não seja ergódico já que as condições iniciais afetariam a estatística após um tempo muito longo. Mostramos, nesta seção, que esta intuição está correta.

Definimos a condição de mistura fraca

$$\lim_{\tau \rightarrow \infty} \Sigma(\tau) = 0, \quad (4.28)$$

onde $\Sigma(\tau)$ é a função de covariância de um processo estacionário. De maneira similar, definimos o critério de ergodicidade fraco como sendo

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt u(t) = \int du f(u) u, \quad (4.29)$$

onde o lado esquerdo é uma média temporal e o lado direito representa a média sobre a distribuição $f(u)$. O objetivo desta seção é verificar que é possível inferir a Eq. (4.29) a partir da Eq. (4.28).

A igualdade (4.29) deve ser interpretada no sentido "quase certo" da esta-

¹¹O ensemble microcanônico pode estar sujeitos a vínculos adicionais fora o da energia.

tística: o lado esquerdo é uma função aleatória de $u(t)$, mas o lado direito é determinístico. Mais explicitamente, a Eq. (4.29) diz que a probabilidade da variável $F_T \equiv \frac{1}{T} \int_0^T dt u(t)$ assumir um valor diferente daquele ao lado direito converge para zero. A estratégia na nossa demonstração para determinar a ergodicidade do processo estocástico consiste em mostrar sob quais condições a média de F_T converge para um valor fixo e o desvio padrão de F_T converge para zero no limite $T \rightarrow \infty$.

Calculamos a média $\langle F_T \rangle$ de um processo estacionário $u(t)$ como ¹²

$$\langle F_T \rangle = \frac{1}{T} \int_0^T dt \langle u(t) \rangle. \quad (4.30)$$

$$= \frac{1}{T} \int_0^T dt \mu = \mu, \quad (4.31)$$

onde μ é a média de $u(t)$. Para o processo ser ergódico, o desvio padrão de F_T deve ser nulo no limite $T \rightarrow \infty$, garantindo a convergência em probabilidade da Eq. (4.29). Calculamos

$$\langle\langle F_T \rangle\rangle = \frac{1}{T^2} \int_0^T dt \int_0^T dt' \langle (u(t) - \mu)(u(t') - \mu) \rangle, \quad (4.32)$$

$$= \frac{1}{T^2} \int_0^T dt \int_0^T dt' \Sigma(t - t'), \quad (4.33)$$

$$= \frac{2}{T^2} \int_0^T dt \int_0^t dt' \Sigma(t - t'). \quad (4.34)$$

Na passagem de (4.32) para (4.33) definimos $\Sigma(t - t')$ como a matriz de covariância de $u(t)$ e na passagem de (4.33) para (4.34) utilizamos o fato que $\Sigma(\tau) = \Sigma(-\tau)$ para mudar os limites de integração.

É possível avaliar o limite $\lim_{T \rightarrow \infty} \langle\langle F_T \rangle\rangle$ facilmente utilizando a regra de

¹²A rigor, a demonstração que fazemos exige apenas que a média e a covariância sejam estacionárias.

L'Hôpital. Aplicando-a uma vez em (4.34), obtemos

$$\lim_{T \rightarrow \infty} \langle \langle F_T \rangle \rangle = \frac{1}{T} \int_0^T dt' \Sigma(T - t') = \frac{1}{T} \int_0^T dt' \Sigma(t'). \quad (4.35)$$

Aplicando novamente a regra de L'Hôpital em (4.35) o resultado é

$$\lim_{T \rightarrow \infty} \langle \langle F_T \rangle \rangle = \lim_{T \rightarrow \infty} \Sigma(T), \quad (4.36)$$

que simplesmente reproduz a condição de mistura fraca na Eq. (4.28). Mostramos portanto que a condição de mistura fraca é equivalente ao processo estocástico estacionário ser fracamente ergódico.

4.2.2 Ergodicidade forte

Dizemos que um processo é fortemente ergódico se para qualquer função $g(u)$ vale a igualdade

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt g(u(t)) = \int du f(u) g(u) \quad (4.37)$$

para alguma distribuição de equilíbrio $f(u)$. Podemos entender a condição de ergodicidade fraca (4.29) como se “médias temporais são iguais às médias de ensemble”. A Eq. (4.37) é uma condição mais forte, onde está em questão se qualquer estatística calculada sobre qualquer trajetória suficientemente longa possa ser calculada a partir de uma distribuição de ensemble $f(u)$. A condição de ergodicidade forte é necessária se quisermos garantir que cada trajetória $u(t)$ funcione como um gerador de números aleatórios da distribuição $f(u)$.

Definimos a condição de mistura forte como sendo

$$\lim_{\tau \rightarrow \infty} G_\tau(u|u') = f(u). \quad (4.38)$$

onde $G_\tau(u|u')$ é a probabilidade de transição de u' para u após um tempo τ e $f(x)$ é a probabilidade marginal de um processo estacionário. A equação anterior postula que toda informação sobre a condição inicial se perderia para tempos muito longos. A condição de mistura fraca (4.28) diz respeito apenas às correlações lineares. De um modo geral, as Eqs. (4.28) e (4.38) não se equivalem. A Eq. (4.28) é uma condição necessária para que valha a Eq. (4.38) em processos com desvio padrão finito. Em alguns processos estocásticos (ex.: processos Gaussianos) as condições de mistura fraca e forte se equivalem. Vale notar que existem várias outras condições de mistura disponíveis na literatura [6].

A rigor, a Eq. (4.37) deve valer para qualquer função $g(u)$ escolhida arbitrariamente. No entanto, basta provarmos que a equação vale para $g(u) = e^{izu}$: nesse caso o lado direito vira a função característica, que define completamente a distribuição de equilíbrio $f(u)$. A análise que se segue é muito semelhante àquela realizada na seção anterior, portanto apenas apontamos os resultados principais.

Para começar, definimos a variável aleatória

$$F_T(z) = \frac{1}{T} \int_0^T dt e^{iz u(t)}. \quad (4.39)$$

A média $\langle F_T(z) \rangle$ é facilmente calculada se a distribuição marginal de $u(t)$ é estacionária,

$$\langle F_T(z) \rangle = \frac{1}{T} \int_0^T dt \langle e^{iz u(t)} \rangle. \quad (4.40)$$

$$= \frac{1}{T} \int_0^T dt \phi(z) = \phi(z), \quad (4.41)$$

onde $\phi(z)$ é a função característica da distribuição marginal. Aqui assumimos a condição de estacionariedade explicitamente para as distribuições a 1-ponto

e a 2-pontos. Para o processo ser ergódico, o desvio padrão de $F_T(z)$ deve ser nulo. Deste modo,

$$\left\langle \lim_{T \rightarrow \infty} F_T(z)^2 \right\rangle - \phi(z)^2 = 0, \quad (4.42)$$

que pode ser reescrita explicitamente como

$$\lim_{T \rightarrow \infty} \frac{1}{T^2} \int_0^T dt dt' \left\langle e^{iz u(t) + iz' u(t')} \right\rangle \Big|_{z=z'} = \phi(z)^2. \quad (4.43)$$

O termo $\langle \dots \rangle$ corresponde à função característica da distribuição a dois pontos.

O limite $T \rightarrow \infty$ pode ser avaliado da mesma maneira que na seção anterior já que $\phi_{t-t'}(z, z')$ também é uma função par de $t - t'$. Apresentamos somente o resultado final,

$$\lim_{\tau \rightarrow \infty} \phi_\tau(z, z) = \phi(z)^2. \quad (4.44)$$

Queremos relacionar a Eq. (4.44) com a condição de mistura forte dada em (4.38). Para comparar as duas equações, multiplicamos a Eq. (4.38) por $f(u')$ para obter

$$\lim_{\tau \rightarrow \infty} f_\tau(u, u') = f(u)f(u'), \quad (4.45)$$

onde $f_\tau(u, u')$ representa a probabilidade a 2-pontos. Reescrevemos a equação anterior utilizando a função característica, de modo que

$$\lim_{\tau \rightarrow \infty} \phi_\tau(z, z') = \phi(z)\phi(z'), \quad (4.46)$$

onde z diz respeito à variável u e z' à variável u' . É lógico que se fizermos $z' = z$, a Eq. (4.46) reproduz exatamente a condição de ergodicidade na Eq. (4.44). Assim, a condição de mistura forte é suficiente para haver ergodicidade. No entanto, é possível que existam processos que não satisfazem a condição de mistura forte e ainda assim sejam ergódicos.

4.2.3 Representação da densidade espectral

Existem basicamente duas maneiras como a condição de mistura fraca pode ser violada. Na primeira, temos $\lim_{\tau \rightarrow \infty} \chi(\tau) = \text{cte}$, para uma constante não-nula qualquer. A outra possibilidade é que $\chi(\tau)$ se comporte como uma função periódica e o limite considerado anteriormente não exista. Em qualquer um destes casos, o comportamento assintótico se reflete na densidade espectral $\tilde{\Sigma}(\omega)$. Para traduzir a condição de mistura para a linguagem de densidade espectral, usamos um teorema importante que diz respeito à representação de distribuições. Lembramos também das Eqs. (4.15-4.16) que relacionam funções características de variáveis simétricas com as matrizes de covariância.

Seja $F(x) = \int_{-\infty}^x dx' f(x')$ a probabilidade cumulativa de uma distribuição (pode ser uma distribuição de probabilidades ou também a densidade espectral). Toda distribuição cumulativa $F(x)$ possui uma decomposição unívoca em três componentes [45]

$$F(x) = a F_c(x) + b F_d(x) + c F_s(x). \quad (4.47)$$

As distribuições cumulativas $F_c(x)$, $F_d(x)$ e $F_s(x)$ são respectivamente as partes contínua, descontínua e singular.

Uma distribuição contínua é diferenciável em todos os pontos e possui uma densidade $f(x)$ dada por $dF_c(x)/dx$. Já a distribuição descontínua representa uma função de saltos, e cresce apenas em um número enumerável de pontos. Rigorosamente, esta distribuição não é diferencial nos pontos de descontinuidade, mas normalmente representamos a derivada como uma delta de Dirac. A densidade da parte descontínua portanto corresponde a uma mistura de funções delta de Dirac.

Finalmente, temos a componente singular $F_s(x)$. Esse é um objeto curioso: trata-se de uma função contínua, crescente, mas que a derivada é nula em

quase todos os pontos. Devido a estas propriedades, várias distribuições singulares são conhecidas pelo nome genérico de “escadaria do diabo” [66]. Existem efeitos físicos modelados por distribuições singulares [50], mas via de regra sua aplicabilidade é bastante limitada.

É interessante verificar como a função característica de cada parte da decomposição se comporta. Na linguagem de matrizes de covariância, a função característica representa a covariância enquanto que dF/dx estaria relacionado com a densidade espectral. A decomposição em (4.47) automaticamente se reflete em

$$\phi(z) = a\phi_c(z) + b\phi_d(z) + c\phi_s(z). \quad (4.48)$$

No entanto, as funções características de cada parte não são imediatamente reconhecíveis. É possível, até um certo ponto, inferir as propriedades de continuidade da distribuição a partir do comportamento da função característica em $z \rightarrow \infty$. Segundo o lema de Riemann-Lebesgue, a transformada de Fourier de qualquer função mensurável se anula quando o argumento tende ao infinito. Esse é exatamente o caso da parte contínua da distribuição, e portanto

$$\lim_{z \rightarrow \infty} \phi_c(z) = 0. \quad (4.49)$$

A distribuição da parte descontínua é uma mistura de deltas de Dirac. Tomando a transformada de Fourier, $\phi_d(z)$ seria uma mistura de ondas planas que descrevem uma função quasi-periódica. Deste modo, o limite para $z \rightarrow \infty$ não existe, mas podemos tomar o limite superior que resulta em

$$\limsup_{z \rightarrow \infty} \phi_d(z) = 1. \quad (4.50)$$

Finalmente, a parte singular possui um comportamento intermediário,

$$\limsup_{z \rightarrow \infty} \phi_s(z) = L, \quad (4.51)$$

onde L pode ser um número no intervalo $0 \leq L \leq 1$.

Agora utilizamos a decomposição mencionada anteriormente explicitamente no contexto de funções de covariância. Na discussão que se segue, assumimos que a parte singular da distribuição é nula, que é uma condição necessária para que seja possível escrever uma densidade espectral. Deste modo, fatoramos a densidade espectral univocamente como

$$\tilde{\Sigma}(\omega) = \tilde{\Sigma}_c(\omega) + \sum_{n=0}^N \frac{a_n}{2} [\delta(\omega - \omega_n) + \delta(\omega + \omega_n)], \quad (4.52)$$

onde utilizamos o fato de se tratar de uma distribuição simétrica (lembre-se que a Eq. (4.52) representa a derivada da Eq. (4.47). Transformando esse resultado de volta para τ , temos

$$\Sigma(\tau) = \Sigma_c(\tau) + \sum_{n=0}^N a_n \cos(\omega_n \tau). \quad (4.53)$$

Segundo a Eq. (4.47), a parte contínua da covariância se anula quando $\tau \rightarrow \infty$. Deste modo, a condição de mistura fraca é violada somente se a parte descontínua for não-nula.

4.3 Difusão

Difusão é o fenômeno em que uma partícula se move devido a colisões aleatórias com as moléculas de um fluido. Um conjunto de partículas que difunde em um recipiente uniforme tende a se espalhar até atingir uma distribuição uniforme. A maneira como esse processo se dá é importante para compreender vários

fenômenos em física, química e biologia.

A difusão é um dos mecanismos dominantes na determinação da velocidade de diversas reações químicas [37]. Ela também é responsável pelo transporte de proteínas, RNA e alguns nutrientes no meio intracelular [20]. Em física da matéria condensada, a difusão participa do transporte de condutores [48], entre várias outras aplicações. Sistemas não-moleculares também podem apresentar uma fenomenologia semelhante à difusão. Alguns exemplos são a transmissão de doenças ou genes em uma população [4], ativos financeiros [62, 56], entre muitos outros.

Os exemplos anteriormente citados são apenas uma amostragem dispersa. De todos sistemas onde ocorre difusão, o movimento Browniano ainda é o caso emblemático. O primeiro modelo matemático bem sucedido para explicar o fenômeno foi proposto por Einstein [14]. Einstein caracterizou a difusão pelo desvio padrão da distribuição de posições. Considerando o movimento aleatório de uma partícula esférica de raio a em um fluido de viscosidade η , temos

$$\langle x(\tau)^2 \rangle \simeq 2D\tau, \quad (4.54)$$

onde o coeficiente de difusão D é igual a

$$D = \frac{k_B T}{6\pi a \eta}. \quad (4.55)$$

Esse resultado, conhecido como a fórmula de Einstein, foi importante historicamente porque permitia estimar o número de Avogadro de uma maneira relativamente direta. Como é bem sabido, o número de Avogadro está escondido na Eq. (4.55) dentro da constante de Boltzmann, $k_B = R/N_A$. A constante dos gases R e os outros fatores da fórmula são facilmente mensuráveis¹³.

¹³Existem diversas maneiras mais convenientes e precisas de determinar o número de Avogadro experimentalmente. O próprio Einstein propôs algumas alternativas antes mesmo da publicação de seu celebrado artigo sobre o movimento Browniano [54].

A difusão ainda é um tópico de pesquisa atual. Do ponto de vista teórico, a atenção, agora, se voltou para os sistemas com a chamada difusão anômala nos quais a expressão de Einstein falha qualitativamente. Em sistemas anômalos, a variância da posição não cresce mais linearmente com o tempo.

É possível chegar qualitativamente na Eq. (4.54) a partir de um raciocínio estatístico muito simples, daí o desconforto causado por sistemas com difusão anômala. Considere o movimento Browniano em uma dimensão. Um modelo razoável postula que a cada pequeno intervalo δt , a partícula em observação sofre um deslocamento aleatório. Seja a variância da distribuição de saltos dada por v_0 . Em $t_n = n \delta t$, a posição da partícula é a soma dos n passos aleatórios executados. É claro que, se estes passos são estatisticamente idênticos e independentes, então a variância seria $v_0 n = v_0 t_n / \delta t$, que cresce linearmente com o tempo.

A difusão anômala é observada em várias situações. Citamos o transporte de portadores em sistemas com armadilhas [48], difusão em sistemas fractais [53] e ainda o transporte de portadores em uma proteína [36], só para nos ater a uma lista muito breve — a Ref. [47] apresenta vários exemplos de processos deste tipo. É comum generalizar a fórmula de Einstein assumindo um expoente diferente de 1 para τ . Assim, a relação de difusão fica

$$\langle x(\tau)^2 \rangle \simeq 2D_\alpha \tau^\alpha. \quad (4.56)$$

A difusão normal, obviamente, corresponde à $\alpha = 1$ e denomina-se subdifusão e superdifusão os casos em que α é respectivamente menor ou maior que 1. Note que existem sistemas que não se enquadram na formulação anterior. As leis de Lévy, por exemplo, possuem variância infinita e devem ser analisadas de outras maneiras.

4.3.1 Expoentes anômalos

O objetivo desta seção é determinar o expoente de difusão assintótico a partir da matriz de covariância de um processo que descreve a velocidade $u(t)$ da partícula Browniana. Assumimos que o processo para a velocidade é estacionário e que a lei de equipartição é respeitada, de modo que $\Sigma(0) = \frac{k_B T}{m}$.

Determinar o expoente de difusão se resume a avaliar como a variância de $x(\tau) \equiv \int_0^\tau ds u(s)$ se comportaria. Um cálculo direto determina a variância de $x(\tau)$,

$$\langle x(\tau)^2 \rangle = \int_0^\tau ds ds' \Sigma(s - s') \quad (4.57)$$

$$= \int_{-\infty}^{\infty} d\omega \tilde{\Sigma}(\omega) \int_0^\tau ds ds' \frac{e^{-i\omega(s-s')}}{2\pi} \quad (4.58)$$

$$= \int_{-\infty}^{\infty} d\omega \tilde{\Sigma}(\omega) \left[\frac{2 \sin^2(\omega\tau/2)}{\tau \pi \omega^2} \right] \times \tau, \quad (4.59)$$

onde a passagem da segunda para a terceira linha é feita resolvendo as integrais em s e s' após trocar a ordem de integração. O termo em colchetes corresponde a uma representação da delta de Dirac como distribuição, como mostrado abaixo

$$\lim_{a \rightarrow \infty} \delta_a(\omega) = \delta(\omega); \quad \delta_a(\omega) \equiv \frac{a \sin^2(\omega/a)}{\pi \omega^2}. \quad (4.60)$$

Deste modo, podemos computar imediatamente o limite de $\tau \rightarrow \infty$ da Eq. (4.59), (que equivale a $a \rightarrow 0$, uma vez que identificamos $a = 2/\tau$)

$$\lim_{\tau \rightarrow \infty} \frac{\langle x(\tau) \rangle^2}{\tau} = \tilde{\Sigma}(0). \quad (4.61)$$

Devido à simetria da matriz de covariância e à relação entre pares de transformadas de Fourier, é possível escrever o lado direito de (4.61) como

$$\tilde{\Sigma}(0) = \int d\tau \Sigma(\tau) e^{i(0)\tau} = 2 \int_0^\infty d\tau \Sigma(\tau) = 2D \quad (4.62)$$

A Eq. (4.62) é conhecida como a relação de Green-Kubo para o coeficiente de difusão. Juntas, as Eqs. (4.62) e (4.54) caracterizam o processo de difusão normal, onde assintoticamente a variância cresce linearmente com o tempo, e a constante de proporcionalidade é definida como duas vezes o coeficiente de difusão.

A difusão anômala ocorre quando a integral em (4.62) se anula ou diverge. Vamos então reescrever a Eq. (4.57) de uma maneira conveniente para extrair o expoente de difusão. Tirando o logaritmo de ambos os lados da Eq. (4.57), e diferenciando com relação à τ , obtemos

$$\frac{\partial}{\partial \ln \tau} \ln \langle x(\tau)^2 \rangle = \frac{\partial}{\partial \ln \tau} [\ln(2D) + \alpha \ln \tau] = \alpha. \quad (4.63)$$

Note que o crescimento da variância como lei de potência no lado direito da Eq. (4.56) só é esperado assintoticamente. Após substituir a Eq. (4.59) em (4.63), tomamos o limite $\tau \rightarrow \infty$. Um cálculo direto leva a

$$\alpha = 1 - \lim_{\tau \rightarrow \infty} \frac{\int d\omega \delta_a(\omega) \omega \frac{\partial \tilde{\Sigma}(\omega)}{\partial \omega}}{\int d\omega \delta_a(\omega) \tilde{\Sigma}(\omega)}, \quad (4.64)$$

Na passagem anterior, foi necessário assumir que $\tilde{\Sigma}(\pm\infty) = 0$ ao realizar a integração por partes, o que só é verdade em distribuições contínuas. As distribuições descontínuas podem ser tratadas a partir de um raciocínio bem mais simples, que será feito separadamente na Sec. 4.3.3.

A Eq. (4.64) deve ser avaliada com um pouco de cautela pois à exceção do caso que apresenta difusão normal, onde $\tilde{\Sigma}(0)$ é um valor finito não-nulo, trata-se de um limite singular. A função $\delta_a(\omega)$ efetivamente realiza um corte no domínio de integração da Eq. (4.64) já que para $|\omega| \gg a^{-1}$, $\delta_a(\omega)$ atinge um valor suficientemente próximo de zero para que o restante da integral possa ser desprezado. Avaliamos como o limite (4.64) se comporta quando a densidade

espectral pode ser escrita como uma lei de potências próximo à origem:

$$\tilde{\Sigma}(\omega) \sim \omega^{-\beta}. \quad (4.65)$$

Para um valor de a o suficientemente pequeno, seria possível trocar os integrandos em (4.64) pelo valor em (4.65): $\delta_a(\omega)$ garantiria, então, que o integrando se tornaria desprezível no intervalo que a densidade espectral começasse a divergir consideravelmente de (4.65).

Substituindo a Eq. (4.65) em (4.64) ficamos com

$$\alpha = 1 - \lim_{\tau \rightarrow \infty} \frac{\int d\omega \delta_a(\omega) \omega (-\beta \omega^{-\beta-1})}{\int d\omega \delta_a(\omega) \omega^{-\beta}} = 1 + \beta. \quad (4.66)$$

Observe, no entanto, que esse resultado só vale se $-1 < \beta < 1$. Para $\beta > 1$, a densidade espectral carregaria uma divergência não-normalizável na origem. Isto implica que $\Sigma(0) = \infty$, o que é fisicamente absurdo. Por outro lado, se $\beta < -1$, as integrais em (4.66) divergem e ficamos novamente com um limite singular (que pode ter qualquer valor). De certo modo, já era possível antever algum problema nesse limite uma vez que o expoente de difusão α seria negativo, o que não é plausível.

O caso $\beta < -1$ está associado à ausência de difusão, ou seja, $\langle x(\tau)^2 \rangle = \text{cte}$. A prova é muito simples, uma vez que se percebe que uma eventual divergência da integral (4.59) para $\tau \rightarrow \infty$ ocorreria na origem (o fato da integral não divergir implica que $\alpha = 0$). Reorganizamos a Eq. (4.59) como

$$\langle x(\tau)^2 \rangle = \frac{4}{\pi} \int_0^\infty d\omega \frac{\tilde{\Sigma}(\omega)}{\omega^2} \sin(\omega\tau/2)^2, \quad (4.67)$$

$$\leq \frac{4}{\pi} \int_0^\infty d\omega \frac{\tilde{\Sigma}(\omega)}{\omega^2}. \quad (4.68)$$

A segunda integral converge sempre que $\beta < -1$ e portanto $\omega^{-\beta}/\omega \rightarrow 0$ para $\omega = 0$. Isto garante a convergência da primeira integral e portanto um processo

deste tipo não apresenta difusão, como queríamos demonstrar.

Resumindo, se a densidade espectral $\tilde{\Sigma}(\omega)$ se comportar como lei de potência para ω muito pequeno, a difusão ocorreria segundo

$$\tilde{\Sigma}(\omega) \sim \omega^{-\beta}; \quad \langle x(\tau)^2 \rangle \sim \begin{cases} \tau^{1+\beta} & -1 \leq \beta < 1 \\ \tau^0 & \beta \leq -1 \end{cases}. \quad (4.69)$$

Deste modo, os processos superdifusivos possuem $\beta > 0$ e a densidade espectral *diverge* na origem. Enquanto isso, processos subdifusivos são caracterizados por $\beta < 0$ e a densidade espectral se *anula* na origem. De um modo geral, a escala no espaço de frequências em que a aproximação em (4.65) começa a valer define inversamente a escala de tempo necessária para a difusão de fato se comporte como lei de potência. O resultado anterior é uma caracterização assintótica da difusão. É possível que em escalas de tempo mais curtas, a difusão ocorra de maneira consideravelmente mais complicada. Há situações em que efeitos transientes são dominantes na escala de tempo do laboratório.

É interessante que exista uma relação simples entre o expoente de difusão e a densidade espectral do processo para as velocidades. Um resultado semelhante à Eq. (4.69) foi obtido por Morgado, et. al. [49] no contexto da equação de Langevin generalizada. Morgado caracteriza a difusão a partir do comportamento na origem da transformada de Laplace da covariância da *força aleatória*. Os dois resultados não são diretamente comparáveis. No entanto, utilizamos a Eq. (4.69) na Sec. 6.3 para obter uma relação semelhante à de Morgado para a equação de Langevin generalizada.

4.3.2 Exemplos

Exemplo I: Decaimento em leis de potência

Considere correlações que decaem assintoticamente como leis de potência

$$\Sigma(\tau) \propto |\tau|^{-\gamma}. \quad (4.70)$$

Vale observar que no regime subdifusivo ($\gamma > 1$), a integral de $\Sigma(\tau)$ se anula e para isto acontecer a função de covariância deve oscilar. Nestes casos, interpretamos (4.70) como o decaimento da amplitude de $\Sigma(\tau)$.

De todo modo, a transformada de Fourier de $\Sigma(\tau)$ existe para γ positivo e se comporta na origem como $\tilde{\Sigma}(\omega) \propto \omega^{\gamma-1}$. Comparando com (4.69), reconhecemos $\beta = 1 - \gamma$ e portanto

$$\langle x(\tau)^2 \rangle \propto \tau^{2-\gamma}. \quad (4.71)$$

Vale observar que de acordo com a Eq. (4.69), se $\gamma > 2$, o expoente de difusão seria nulo.

Uma correlação positiva indica a propensão de se manter o sinal da velocidade inicial ao longo do movimento. O caso limite em que $\Sigma(t) = \text{cte} = \Sigma(0)$ e portanto $\gamma = 0$ representa a situação onde a correlação é 1 e assim representa o extremo em que a velocidade evolui de maneira reversível. Esse é um exemplo emblemático de difusão balística. Note que a transformada de Fourier da matriz de covariância não decai como lei de potência, sendo dada por

$$\tilde{\Sigma}(\omega) = \Sigma(0)\delta(\omega). \quad (4.72)$$

O resultado anterior é de caráter geral. Os núcleos associados à transformação balística não decaem à zero quando $\tau \rightarrow \infty$. Chamamos Eq. (4.72) de

núcleo balístico puro, que descreve o caso em que as partículas simplesmente se espalham sem realizar colisões (e.g., efusão de um gás ideal).

Exemplo II: Função gama

Por suas propriedades matemáticas, a função gama é um exemplo especialmente interessante de matriz de covariância (ver tabela 4.4 na página 81). O comportamento na origem é regulado por um índice que permite cobrir toda a faixa de expoentes anômalos, exceto a difusão balística.

A Eq. (4.69), que relaciona o expoente de difusão com o comportamento da densidade espectral na origem, se aplica na faixa $0 < \nu < 2$. É óbvio que para a função gama, $\tilde{\Sigma}(\omega) \sim \omega^{\nu-1}$, o que resulta em $\langle x(\tau)^2 \rangle \sim \tau^{2-\nu}$. Na faixa $\nu \geq 2$, a análise é um pouco diferente. Para $\nu = 2$, a covariância é

$$\Sigma(\tau) = \Sigma(0) \Re [(1 - is\tau)^{-2}] = \Sigma(0) \frac{1 - s^2\tau^2}{(1 + s^2\tau^2)^2}, \quad (4.73)$$

onde s é uma constante arbitrária com dimensão de frequência. Substituindo a expressão anterior na integral (4.57), segue, que

$$\langle x(\tau) \rangle = \frac{k_B T}{m s^2} \ln(1 + s^2\tau^2). \quad (4.74)$$

A difusão, nesse caso, não é descrita por uma lei de potências, mas podemos entendê-la como um caso intermediário: o logaritmo cresce mais lentamente que qualquer potência de τ . Para $\nu > 2$, a integral (4.57) converge, e isto caracteriza um expoente de difusão nulo.

O motivo pode ser entendido heurísticamente. A covariância decai com ordem $\tau^{-\nu} = (t - t')^{-\nu}$. Como temos uma integral em t e outra em t' , é necessário que $\nu > 2$ para que as integrais convirjam no limite $t, t' \rightarrow \infty$. Para $\nu < 2$, a integral diverge como lei de potência, caracterizando um expoente anômalo, mas para $\nu = 2$ a divergência é logarítmica, resultando na Eq. (4.74).

Finalmente, em $\nu > 2$, a integral converge e não há difusão.

4.3.3 Difusão em distribuições descontínuas

A análise que segue da Eq. (4.59) não se aplica a distribuições descontínuas. No entanto, é possível estudar a difusão facilmente nesse caso. A covariância associada a uma densidade espectral puramente descontínua é uma mistura de cossenos, como abaixo

$$\Sigma(\tau) = \sum_{n=0}^N a_n \cos(\omega_n \tau), \quad (4.75)$$

onde assumimos que $\omega_0 < \omega_1 < \dots < \omega_N$. Deste modo, basta calcular a integral (4.57) para a distribuição acima. O resultado é

$$\langle x(\tau)^2 \rangle = \sum_{n=0}^N 2 a_n \frac{1 - \cos(\omega_n \tau)}{\omega_n^2}, \quad (4.76)$$

onde não há qualquer restrição quanto a N ser finito.

Aparentemente, a Eq. (4.76) implica que um sistema com densidade espectral descontínua não difunde. No entanto, se ω_0 for nulo ou muito pequeno, a difusão em uma escala de tempo $\tau \ll \omega_0^{-1}$ será balística, como se percebe pelo limite

$$\lim_{\omega_0 \rightarrow 0} \frac{1 - \cos(\omega_0 \tau)}{\omega_0^2} = \frac{1}{2} \tau^2. \quad (4.77)$$

Segundo a equação anterior, a ausência de difusão se dará somente no limite em que $\tau \gg \omega_0^{-1}$. Se alguns ω_n forem muito baixos, esse limite poderia não ser alcançado em escalas de tempo do laboratório e portanto a difusão se apresentaria como uma função potencialmente complicada do tempo. Por outro lado, se a menor frequência for $\omega_0 \ll \tau^{-1}$, o comportamento será aproximadamente balístico.

A expressão completa para a densidade espectral também inclui a parte contínua (e a parte singular que estamos desprezando). Nesse caso, a difusão

é dominada pelo termo que difunde mais rapidamente.

Vale notar que, a rigor, os sistemas físicos tendem a um estado de equilíbrio. Nesse caso, partícula Browniana se encontraria em qualquer posição dentro do recipiente em que está confinada com igual probabilidade. Obviamente, nesse caso $\langle x(t)^2 \rangle$ tenderia a uma constante e obteríamos um expoente de difusão nulo. Esse comportamento aparece naturalmente em uma distribuição descontínua se ω_0 não for rigorosamente nulo. Para que o mesmo ocorra na parte contínua da distribuição, seria necessário que a densidade espectral fosse nula em um intervalo $\omega \in [0, \omega_0]$, para ω_0 muito pequeno. Assim, a integral (4.68) convergiria.

4.4 Conclusão

4.4.1 Mistura e ergodicidade

Ao tratar a ergodicidade de um sistema estocástico, obtemos um resultado que contradiz Lee [42], mas é compatível com o trabalho de Lapas, *et. al.* [40]. Possivelmente, a contradição surge porque Lee utiliza considera uma teoria quântica. É importante destacar que a noção de processo estocástico não se aplica a sistemas quânticos. Para que um processo estocástico seja definido, é necessário construir todas as probabilidades a N -pontos $P(u_1, \dots, u_N)$. No caso de uma variável quântica, isto não é possível porque um operador de Heisenberg $\hat{A}(t)$ geralmente não comuta com $\hat{A}(t')$ ($t \neq t'$). Não existe uma prescrição bem definida para atribuir uma probabilidade conjunta a operadores que não comutam. Mais ainda, é possível mostrar que nenhuma probabilidade conjunta pode ser definida para determinados conjuntos de operadores.

Ainda que não exista uma definição precisa de processo estocástico quântico, se define funções de covariância quântica e outros conceitos estatísticos. Estas definições são inerentemente ambíguas e nada garante que se comportem

como os objetos da teoria de probabilidades. O teorema de ergodicidade fraca apresentado na Sec. 4.2.1 pode ser formalmente generalizado para o regime quântico definindo-se a covariância entre dois operadores como

$$Cov(A, B) = \frac{1}{2} \langle (A - \langle A \rangle) (B - \langle B \rangle) \rangle, \quad (4.78)$$

onde $\langle \dots \rangle$ denota a média no espaço de Hilbert. Lee utiliza uma definição diferente de covariância quântica. Acreditamos que a discrepância entre seu resultado e o resultado para processos estocásticos se deva fundamentalmente à esta escolha.

4.4.2 Difusão

Estabelecemos um critério para obter o expoente de difusão a partir da densidade espectral do processo estocástico para as velocidades. Esse resultado lembra o de Morgado, et. al. [49], mas ainda não pode ser diretamente comparável à ele. Faremos uma comparação na Sec. 6.3, após apresentar a equação de Langevin generalizada. O método aqui apresentado não faz uso desta equação. A única suposição é que o processo estocástico para as velocidades seja estacionário.

Outro ponto importante é que mostramos que a difusão balística é uma propriedade dos sistemas que violam a condição de mistura fraca e, de acordo com o resultado anterior, também *não* são ergódicos.

Capítulo 5

Movimento Browniano e equação de Langevin

A equação de Langevin foi formulada para descrever o movimento de partículas mesoscópicas em um fluido. Estudamos sua extrapolação para o regime microscópico, que talvez seja indevida [17, 65]. O estudo apresentado nesse capítulo está baseado em um modelo cinético simples, onde as partículas trocam momentum por colisões elásticas. A equação de Langevin não descreve bem esse modelo, e representa uma aproximação razoável apenas em alguns regimes. Ainda assim, mostramos que a melhor aproximação do tipo equação de difusão para descrever um gás de esferas rígidas utilizaria um coeficiente de difusão que depende da velocidade da partícula Browniana.

Pouco após Einstein publicar seu celebrado artigo sobre o movimento Browniano [14], Paul Langevin propôs um modelo que obteve os mesmos resultados utilizando um raciocínio matematicamente mais simples [43]. No modelo de Langevin, a partícula Browniana evolui de acordo com as leis de Newton, e

assim, a sua velocidade $u(t)$ é governada por

$$m \frac{du}{dt} = -\gamma m u + \varphi(t), \quad (5.1)$$

onde $-\gamma m u$ corresponde a uma força de atrito viscoso e $\varphi(t)$ é uma força aleatória. A força aleatória $\varphi(t)$ modela as flutuações na velocidade média das partículas do fluido que seriam responsáveis pelo movimento Browniano.

Apresentamos o argumento de Langevin, que visa calcular a dispersão das posições $\langle x^2(t) \rangle$. Identificamos $u = dx/dt$ com a velocidade das partículas. Substituindo esta expressão em (5.1) e após multiplicar (5.1) por x e reorganizar alguns termos, o resultado é

$$\frac{m}{2} \frac{d^2 x^2}{dt^2} - m \dot{x}^2 = -\frac{\gamma}{2} \frac{dx^2}{dt} + x \varphi. \quad (5.2)$$

Ao tirar a média desta equação, Langevin supôs que $\langle \dot{x}^2 \rangle = \frac{k_B T}{m}$ de acordo com a lei de equipartição. A segunda suposição, que $\langle x \varphi \rangle = \langle x \rangle \langle \varphi \rangle = 0$, determina que as forças aleatórias seriam independentes da posição da partícula, possuindo média nula. O resultado destas considerações é uma equação diferencial de segunda ordem para $\langle x(t)^2 \rangle$. É fácil mostrar que a solução geral é

$$\langle x(\tau)^2 \rangle = 2 \frac{k_B T}{m \gamma} \tau + A + B e^{-2\gamma \tau}, \quad (5.3)$$

onde apenas o primeiro termo com um crescimento linear em τ é relevante assintoticamente. Obtemos a fórmula de Einstein ao desprezar os termos com as constantes de integração A e B para tempos muito longos ($\tau \gg \gamma^{-1}$).

O modelo de Langevin é simples e aparentemente robusto. Posteriormente, Ornstein e Uhlenbeck forneceram um tratamento mais detalhado da Eq. (5.1), postulando certas propriedades para o ruído aleatório $\varphi(t)$ [64]. Essa é uma descrição mais detalhada e que realiza suposições mais fortes sobre a força

aleatória. Hoje em dia, os dois modelos se confundem e o que a literatura de física geralmente chama de modelo de Langevin é na verdade aquele formulado por Ornstein e Uhlenbeck.

As hipóteses introduzidas por Ornstein e Uhlenbeck permitem calcular o processo estocástico completo associado à velocidade e posição da partícula Browniana. Os próprios autores comentam que “a justificativa precisa, ou eventualmente uma crítica, a estas suposições deve vir de uma teoria cinética mais precisa”. Apresentamos algumas críticas nesse trabalho.

Antes de proceder, é necessário distinguir entre duas aplicações comuns da equação de Langevin: a primeira, proposta originalmente por Langevin, descreve uma partícula macroscópica suspensa em um fluido. Nesse regime valem as leis da hidrodinâmica e a força de arrasto é a força de Stokes. O movimento errático da partícula Browniana é atribuído às flutuações espontâneas do campo de velocidades do fluido com relação ao estado de equilíbrio.

Outro modelo caracteriza a dinâmica de uma partícula microscópica sujeita a colisões com as partículas de um fluido. Aqui, a hidrodinâmica não se aplica já que o fluido não é tratado como um objeto contínuo. Os movimentos aleatórios da partícula Browniana decorrem das colisões com as partículas do fluido. Para distinguir explicitamente entre as duas situações, chamamos a primeira de partícula Stokes (em referência à força viscosa) e a segunda de partícula de Rayleigh e para nos referir a ambas, usamos o termo partícula Browniana indistintamente.

5.1 Partícula de Stokes

5.1.1 Força viscosa

A força de Stokes para uma partícula esférica com velocidade $u(t)$, raio a imersa em um fluido de viscosidade η é dada pela expressão abaixo

$$F_v(t) = -c\eta a u(t), \quad (5.4)$$

onde c é uma constante que depende das condições de contorno utilizadas. Normalmente, $c = 6\pi$ que corresponde ao caso em que o fluido adere à superfície da esfera. Caso o fluido escorra sem esta restrição, teríamos $c = 4\pi$ [65]. Ajustando o valor desta constante, podemos também descrever outros formatos diferentes que uma esfera.

A força de Stokes é uma entre inúmeras forças de arrasto hidrodinâmico. Algumas destas forças dependem de características específicas do objeto imerso, como por exemplo, a presença de rugosidades na superfície ou da sustentação dinâmica (como na asa de um avião). De todo modo, força viscosa em (5.4) e a chamada força de arrasto inercial possuem um papel predominante. Essa última decorre do objeto movente precisar “abrir espaço” pelas partículas do fluido, transferindo momentum para as mesmas nesse processo. Para uma esfera, temos a força de arrasto inercial dada por

$$F_I(t) = \pi\rho a^2 u(t)^2, \quad (5.5)$$

onde ρ é a densidade do fluido.

Todo objeto que se move em um fluido estará sujeito, no mínimo, a estas duas forças. Na maior parte das aplicações, no entanto, apenas um destes dois efeitos é relevante. Definimos o número de Reynolds, que é essencialmente a

razão entre as duas forças, como

$$N_R = \frac{2a\rho u(t)}{\eta}. \quad (5.6)$$

Um número de Reynolds alto $N_R \gg 1$ indica a predominância da força de arrasto inercial, enquanto um número baixo significa que a força viscosa é mais importante e a força de arrasto inercial pode ser desprezada.

No caso do movimento Browniano, podemos substituir $u(t)$ em (5.6) pela média térmica. Desprezando os fatores numéricos e definindo β como a razão da densidade do fluido e da partícula Browniana, isto resulta em

$$N_R \sim \frac{1}{\eta} \sqrt{\rho k_B T} \frac{1}{\sqrt{a\beta}}. \quad (5.7)$$

Ao considerar partículas Brownianas de densidade fixa, o número de Reynolds dependeria de $a^{-1/2}$, que diverge para $a \rightarrow 0$. Como o movimento Browniano só é visível em objetos muito pequenos, parece sensato investigar a possibilidade que a força de arrasto inercial cumpra um papel mais importante que a força viscosa em alguns regimes.

Substituindo alguns valores numéricos na Eq. (5.7) vemos que em diversas aplicações, os números de Reynolds são consideravelmente baixos, validando a utilização da lei de Stokes. Como exemplo, considere o movimento Browniano das organelas no meio celular (a célula geralmente é muito pesada para que o efeito seja visível). Fazendo uma substituição simples de valores, o número de Reynolds para uma mitocôndria se situaria na faixa de $\sim 10^{-3}$, que é mais que o suficiente para justificar o uso da lei de Stokes.

De um modo geral, é isto que acontece. Avaliando o termo $\frac{1}{\eta} \sqrt{\rho k_B T}$ para a água na situação extrema em que $T = 100^\circ C$ e a viscosidade atinge um valor mínimo de $0,28cp$, teríamos $N_R \sim \sqrt{10^{-12}/\beta a}$. Mesmo para materiais consideravelmente mais densos que a água, o número de Reynolds seria maior

que 1 somente se as partículas Brownianas possuíssem dimensões moleculares. É lógico que a hidrodinâmica falha nestes casos e já não se espera que a lei de Stokes seja correta.

No caso microscópico, um argumento comumente repetido para atribuir uma força linear com a velocidade diz que ao se deslocar dentro de um fluido, a partícula de Rayleigh sofreria mais colisões na sua parte frontal que nas outras direções. Isto contribuiria com uma troca de momentum maior na direção contrária ao movimento relativo ao fluido, resultando em uma força proporcional a $-u(t)$.

É tentador pensar que, na medida que consideramos partículas de dimensões maiores, esta força resultaria na força de Stokes. No entanto, são efeitos diferentes: a origem microscópica da viscosidade não está na observação trivial que as colisões ocorrem com mais frequência na direção contrária ao movimento. Se isto fosse verdade, fluidos mais densos sempre seriam mais viscosos que fluidos mais leves. Veremos adiante que esse efeito só é relevante em partículas razoavelmente pequenas. A força linear com a velocidade aparece na partícula de Rayleigh por razões totalmente diferentes daquelas da partícula de Stokes. Por simetria, podemos imaginar que a força aleatória também terá origens diferentes.

5.1.2 Força aleatória

Vale observar que o movimento Browniano, como descrito por Einstein e Langevin e observado por Brown e Perrin se situa no regime hidrodinâmico: trata-se da agitação de uma partícula macroscópica imersa em um fluido. Nesse caso, a lei de Stokes se aplica, e ignorando a presença de flutuações, a velocidade da partícula Browniana seria descrita por

$$u_{t+\tau} = u_t e^{-\gamma\tau}, \quad (5.8)$$

onde definimos $\gamma = c\eta a/m$ de acordo com a Eq. (5.4).

Contrariamente a uma concepção comum, a “dança” da partícula Browniana não é ocasionada por choques individuais com as partículas do fluido. Essa imagem talvez possa ser inferida (incorretamente) do modelo de Langevin, mas já se sabia desde o fim do século *XIX* que tal explicação é inviável. Só para efeito de comparação, as partículas observadas por Brown possuem uma massa da ordem de 10^{10} moléculas de água. É fácil deduzir que, nestas situações, o momento médio transferido em cada colisão seria da ordem de 10^{-5} vezes menor que o momentum que a partícula Browniana possui devido à agitação térmica: não há flutuação estatística que contemple esta enorme discrepância.

É possível obter o processo de Ornstein e Uhlenbeck a partir de algumas considerações simples, que discutiremos em maiores detalhes a seguir. Apresentamos uma demonstração que evita falar sobre ruído branco nesse sentido difere do argumento original. O motivo para isto é evitar definir o ruído branco como uma hipótese fundamental do modelo. O ruído branco é uma idealização que possui várias características abstrusas e não é necessário para determinar o processo estocástico para $u(t)$.

As hipóteses que consideramos são:

1. $u(t)$ é um processo estacionário em um estado de equilíbrio termodinâmico.
2. Vale a lei de van't Hoff que soluções muito diluídas se comportam gases ideais e portanto $\langle u(t)^2 \rangle = \frac{k_B T}{m}$.
3. O movimento de $u(t)$ é Markoviano.

Os pontos (1) e (2) são facilmente justificáveis dentro da física estatística de equilíbrio. Mostraremos que o ponto (3) equivale à imposição do ruído branco para a força aleatória na equação de Langevin.

É possível integrar a equação de Langevin explicitamente, escrevendo-a como

$$u(t + \tau) = u(t)e^{-\gamma\tau} + \delta u_t(\tau). \quad (5.9)$$

O primeiro termo do lado direito representa o decaimento devido à força de Stokes e o segundo termo corresponde a uma contribuição aditiva decorrente das flutuações. Interpretamos a variável estocástica $\delta u_t(\tau)$ como o momentum total transferido pelas flutuações do fluido em no intervalo entre t e $t + \tau$. O processo estocástico para $u(t)$ segue do teorema de Doob generalizado demonstrado no Apêndice III.

Theorem 5.1 (Doob generalizado). *A probabilidade de transição de qualquer processo estocástico aditivo e autônomo é Markoviana se e somente se a média e a matriz de covariância da variável estocástica $u(t)$ são da forma*

$$\langle u(t) \rangle = \begin{cases} a + (b - a)e^{-\lambda t} & \lambda \neq 0 \\ a + ct & \lambda = 0 \end{cases}, \quad (5.10)$$

$$\langle\langle u(t)u(t') \rangle\rangle = \begin{cases} [A + (B - A)e^{-2\lambda \min(t,t')}] e^{-\lambda|t-t'|} & \lambda \neq 0 \\ 2D \min(t, t') & \lambda = 0 \end{cases}, \quad (5.11)$$

onde $\langle\langle \dots \rangle\rangle$ denota os momentos centrais (e.g., $\langle\langle u^n \rangle\rangle = \langle\langle (u - \langle u \rangle)^n \rangle\rangle$).

Dizemos que um processo é aditivo se puder ser escrito na forma (5.9). Ou seja, os incrementos de um processo aditivo são variáveis aleatórias independentes da variável de interesse (que no nosso caso é $u(t)$). Uma outra consequência deste teorema, que está demonstrada no apêndice, é que se $\gamma \neq 0$, então o processo é Gaussiano.

Ao utilizar a forma integral da equação de Langevin (5.9) e impor a propriedade Markoviana, satisfazemos as hipóteses deste teorema. Para que o

processo estocástico resultante seja estacionário, é necessário que $a = b$, $c = d$ e $\lambda \neq 0$. Identificamos o parâmetro $\lambda = \gamma$ comparando a Eq. (5.9) com a Eq. (8.27) no Apêndice III. As outras constantes seguem da lei de equipartição, $a = b = 0$ e $A = B = \frac{k_B T}{m}$. Deste modo, a covariância do processo estocástico para $u(t)$ se escreveria como

$$\langle\langle u(t)u(t') \rangle\rangle = \frac{k_B T}{m} e^{-\gamma|t-t'|}. \quad (5.12)$$

Mostraremos que esse resultado é equivalente ao processo de Ornstein-Uhlenbeck.

5.1.3 Conseqüências matemáticas

Difusão e mistura

A principal motivação de Langevin ao apresentar seu modelo para descrever o movimento Browniano foi obter o coeficiente de difusão e resolver uma inconsistência entre o resultado obtido por Einstein e outro obtido por Smoluchowski. Calculamos a variância de $x(t)$ aplicando a transformação

$$x(t) = x_0 + \int_0^t dt' u(t'), \quad (5.13)$$

à Eq. (5.12). O resultado é

$$\langle\langle x(t)^2 \rangle\rangle = \int_0^t ds ds' \langle\langle u(s)u(s') \rangle\rangle, \quad (5.14)$$

$$= 2 \frac{k_B T}{m\gamma^2} [e^{-\gamma t} + \gamma t], \quad (5.15)$$

que para tempos muito longos, $t \gg \gamma^{-1}$, se simplifica como

$$\langle x(\tau)^2 \rangle \simeq 2 \frac{k_B T}{m\gamma} \tau \equiv 2D\tau; \quad D = \frac{k_B T}{m\gamma}. \quad (5.16)$$

Esse resultado é a fórmula de Einstein para o coeficiente de difusão que também foi obtida em (5.3).

Processo condicional

Usando a Eq. (5.10), podemos calcular a média e a covariância do processo condicionado à $u(t = 0) = u_0$. A média e a covariância condicionais são respectivamente

$$\langle u(t) \rangle = u_0 e^{-\gamma t}, \quad (5.17)$$

$$\langle \langle u(t)u(t') \rangle \rangle = \frac{k_B T}{m} \left[1 - e^{-2\gamma \min(t,t')} \right] e^{-\gamma |t-t'|}, \quad (5.18)$$

que é obtido escolhendo $a = 0$, $b = u_0$, $A = \frac{k_B T}{m}$ e $B = 0$ na Eq. (5.10). Ao preparar uma partícula em u_0 , dizemos que ela caminha para o equilíbrio em uma escala de tempo da ordem de γ^{-1} : quando $t, t' \gg \gamma^{-1}$, as probabilidades a N -pontos com a condição inicial em $u(0) = u_0$ convergem para as probabilidades do processo estacionário. Nesse contexto, "fora do equilíbrio" significa apenas uma lembrança sobre a condição inicial e não diz nada sobre o fluido onde a partícula Browniana está imersa.

Força aleatória: ruído branco

A equação de Langevin formula a evolução estocástica como se valessem as leis de Newton¹. Deste modo, é interessante identificar as propriedades estatísticas da força aleatória $\varphi(t)$. Identificamos

$$F_T = m \frac{d}{dt} u(t) = -\gamma m u + \varphi(t), \quad (5.19)$$

¹O fato das trajetórias não possuírem derivada causa certa apreensão pois a partícula Browniana não possuiria momentum definido.

onde separamos um termo devido à força viscosa para analisar a parte puramente flutuativa $\varphi(t)$. Usando (5.17) e (5.18), mostra-se que a média condicional para tempos muito curtos é

$$\langle u_{t+\delta t} \rangle = u_t e^{-\gamma\delta t} \simeq u_t (1 - \gamma\delta t) \quad (5.20)$$

e a variância é

$$\langle [u_{t+\delta t} - \langle u_{t+\delta t} \rangle]^2 \rangle = \frac{k_B T}{m} (1 - e^{-2\gamma\delta t}) \simeq \frac{2\gamma k_B T}{m} \delta t. \quad (5.21)$$

Deste modo, o deslocamento infinitesimal $du_t = u_{t+\delta t} - u_t$ possui uma média $-\gamma u_t$ e variância $2\gamma k_B T/m \delta t$, o que pode ser representado por uma equação diferencial estocástica

$$du = -\gamma u dt + b dW_t \quad (5.22)$$

onde $b \equiv \sqrt{2\gamma \frac{k_B T}{m}}$ e dW_t é uma variável Gaussiana com média nula e variância dt .

Ao dividir a Eq. (5.22) por dt , teríamos a equação diferencial

$$m \frac{du}{dt} = -\gamma m u + \varphi(t), \quad (5.23)$$

onde identificamos $\varphi(t) = b dW_t/dt$. Observe que $u(t)$ não é diferenciável já que a variância de $\varphi(t)$ é infinita pois $dW_t/dt \sim dt^{-1/2}$. A Eq. (5.23) possui a qualidade indesejável de escrever a derivada de uma função que não possui derivada em nenhum ponto. É necessário conceber a Eq. (5.23) como uma definição formal para o processo $u(t)$ ².

Obtemos o processo estocástico associado à $\varphi(t)$ relacionando-o com $u(t)$

²Mesmo interpretando a equação de Langevin de uma maneira completamente formal, é possível encontrar problemas se o termo que multiplica a força aleatória depender do estado $u(t)$. Nesse caso, existe uma ambigüidade relacionada à escolha do cálculo de Itô vs. Stratonovich [31].

pela transformação $\varphi(t) = \left(\frac{d}{dt} + \gamma\right) u(t)$. Deste modo, a matriz de covariância da nova variável seria dada por

$$\langle\langle \varphi(t)\varphi(t') \rangle\rangle = m^2 \left(\frac{d}{dt} + \gamma\right) \left(\frac{d}{dt'} + \gamma\right) \frac{k_B T}{m} e^{-\gamma|t-t'|} = 2\gamma m k_B T \delta(t-t'). \quad (5.24)$$

A interpretação usual deste resultado é que $\varphi(t)$ não é auto-correlacionado: a força $\varphi(t)$ é independente de $\varphi(t')$ para quaisquer dois instantes $t \neq t'$. Esse é o postulado que Ornstein e Uhlenbeck utilizaram ao formular a equação de Langevin.

Observe que o desvio padrão de $\varphi(t)$ é infinito. As trajetórias resultantes são descontínuas e podem variar em uma faixa de valores arbitrariamente grande. O processo definido como (5.24) representa uma série contínua de impulsos, onde a força diverge em quase todos os pontos e onde cada impulso é estatisticamente independente dos outros.

No entanto, os impulsos gerados pela força aleatória transferem uma quantidade infinitesimal de momentum para a partícula Browniana. O momentum médio transmitido em cada colisão é dado aproximadamente por

$$\Delta p = \int_t^{t+\delta t} |\varphi(t)| dt \simeq 2m \sqrt{\gamma \frac{k_B T}{\pi m}} \delta t, \quad (5.25)$$

onde usamos o limite da delta de Dirac como distribuição $\delta(t-t') \simeq 1/\delta t$. Assim, interpretamos o processo de Langevin-Ornstein-Uhlenbeck como uma série de impulsos que modificam o momentum (ou a velocidade) da partícula em saltos infinitesimais de tamanho médio (5.25). Estes impulsos não representam colisões individuais. Imagina-se que a intensidade destas colisões não estaria relacionada com a viscosidade do fluido, mas sim com as propriedades (massas e seção de choque) das partículas que constituem o fluido.

5.1.4 Relação de flutuação e dissipação

Substituindo a relação de Stokes $\gamma = 6\pi a\eta/m$ em (5.24) vemos o fato curioso que a força flutuativa depende da viscosidade do fluido. Esse resultado segue de impormos a lei de van't Hoff na Eq. (5.12), que determina que a variância de $u(t)$ é $\frac{k_B T}{m}$.

Se concebêssemos as flutuações como resultado direto das colisões com as moléculas do fluido, esse resultado pareceria bastante inesperado: a frequência e intensidade das colisões depende de fatores como a velocidade das partículas do fluido, a seção de choque de colisão, a concentração de partículas, etc. O aparecimento de um termo proporcional à viscosidade só pode ser entendido quando concebemos $\varphi(t)$ como o resultado de flutuações no campo de velocidades do fluido. O mesmo mecanismo que gera o arrasto, a lei de Stokes, também seria responsável por transferir momentum das flutuações do fluido para a partícula Browniana. Lembramos ainda do paradoxo de D'Alembert, que mostra que um fluido não viscoso seria incapaz de transmitir momentum para um objeto submerso.

De um modo geral, é necessário que exista uma determinada relação entre os termos de flutuação e dissipação de energia para que o sistema físico atinga o equilíbrio termodinâmico. A tentativa de encontrar explicitamente a relação entre estes termos é o conteúdo dos chamados teoremas de flutuação e dissipação.

Mesmo na equação de Langevin usual, dada em (5.1), existe uma ambigüidade na identificação dos termos de flutuação e dissipação. Uma vez especificada a matriz de covariância $\langle\langle u(t)u(t') \rangle\rangle = \frac{k_B T}{m} e^{-\lambda|t-t'|}$ existe uma arbitrariedade em escolher o parâmetro γ na força de fricção (não é necessário

que $\gamma = \lambda$). Para ver isto, calculamos a covariância do ruído

$$\langle\langle\varphi(t)\varphi(t')\rangle\rangle = m^2 \left(\frac{d}{dt} + \gamma\right) \left(\frac{d}{dt'} + \gamma\right) \langle\langle u(t)u(t')\rangle\rangle, \quad (5.26)$$

$$= mk_B T \left[\lambda \delta(t-t') + (\gamma^2 - \lambda^2) e^{-\lambda|t-t'|} \right]. \quad (5.27)$$

Ao escolher a covariância do ruído como (5.27), é possível controlar o parâmetro γ de maneira arbitrária (basta que $\gamma > \lambda$) e ainda assim obter o mesmo processo estocástico para $u(t)$.

Para desacoplar o efeito de flutuação e de dissipação é necessário apelar para causas externas. No modelo de Langevin, postulamos que o decaimento da função velocidade média segue a lei de Stokes e portanto $\gamma = \lambda$ (esta condição é as vezes denominada de hipótese de Onsager). Ainda que pareça fisicamente razoável, esta é uma condição adicional à equação de Langevin que não é necessária pela consistência matemática.

5.2 Gás de esferas rígidas

Os argumentos apresentados anteriormente não são válidos caso a partícula Browniana seja microscópica. Em que pese, a hidrodinâmica não se aplica e portanto não é possível invocar a lei de Stokes para descrever a força de fricção. Apresentamos um modelo elementar que acompanha a mudança na velocidade da partícula Browniana colisão a colisão. O modelo descreve um gás diluído onde a interação com a partícula Browniana se dá por um potencial de esferas rígidas.

Seja uma partícula Browniana de massa M imersa em um fluido de partículas com massa m . A dinâmica deste sistema não é contínua: cada colisão implica na troca de uma quantidade discreta de momentum linear δp , mas entre duas colisões sucessivas, nenhuma mudança ocorre. Uma dinâmica apro-

ximadamente contínua surge se a partícula Browniana for muito mais massiva que as partículas do fluido: assim, as trocas de momentum correspondem a mínimos acréscimos de velocidade que, ao final, se configuram em uma força de fricção efetiva. É possível recuperar o modelo de Langevin nesse limite. Se as partículas forem idênticas, a dinâmica resultante é consideravelmente diferente do processo de Ornstein-Uhlenbeck.

Denotamos por u' a velocidade inicial da partícula Browniana e u a velocidade final após colidir com uma partícula do fluido de velocidade v . Usando as leis de conservação de momentum e energia em uma dimensão, obtemos

$$u = \frac{M - m}{M + m}u' + \frac{2m}{M + m}v. \quad (5.28)$$

Se o fluido estiver em equilíbrio, v possuirá uma distribuição de Maxwell com variância $k_B T/m$. Nesse caso, a probabilidade da partícula Browniana realizar uma transição de u' para u após colidir com a partícula de velocidade v é uma Gaussiana com média e variância dadas respectivamente por

$$\mu = \frac{M - m}{M + m}u' \equiv \zeta u'; \quad \sigma^2 = \frac{4m^2}{(M + m)^2} \frac{k_B T}{m}. \quad (5.29)$$

A frequência média de colisões, pode ser obtida por considerações cinéticas como sendo

$$\lambda \sim \frac{\rho}{m} \bar{\sigma} \sqrt{\frac{8 k_B T}{\pi m}} \quad (5.30)$$

onde $\bar{\sigma}$ é seção de choque de colisão e ρ é a densidade do fluido³. De acordo com as considerações feitas na Sec. 3.2, o processo estocástico para $u(t)$ seria descrito por uma probabilidade de saltos Gaussiana que define a equação

³A rigor, deveríamos utilizar a velocidade relativa ao invés de $\sqrt{\frac{k_B T}{M}}$, que introduziria uma dependência explícita em u . No entanto, se $M \gg m$, a velocidade térmica da partícula Browniana seria muito menor que a das partículas do fluido, tornando esse efeito irrelevante.

mestra abaixo

$$\partial_\tau G_\tau(u|u') = \int dw \lambda [q(u|w) - \delta(u-w)] G_\tau(w|u'). \quad (5.31)$$

Onde convêm explicitar a distribuição de um salto

$$q(u|u') = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(u-\zeta u')^2}. \quad (5.32)$$

Ainda que a solução da equação mestra (5.31) não seja trivial, é possível obter uma série de resultados importantes de forma fechada.

5.2.1 Probabilidades $q(u|u', n)$

As probabilidades $q(u|u', n)$ que descrevem a variação de velocidade após n colisões aleatórias podem ser determinadas diretamente realizando as várias integrais na Eq. (3.27). No entanto, é possível calculá-las por um procedimento muito mais simples que utiliza a relação (5.28). Definimos $g \equiv 2m/(M+m)$ e reescrevemos esta relação como

$$u_1 = \zeta u_0 + g v_0, \quad (5.33)$$

onde o subscrito em u_1 enfatiza que esse é o resultado da primeira colisão. Interpretando a equação anterior como uma relação de recorrência ($u_{i+1} = \zeta u_i + g v_i$), é simples mostrar que

$$u_n = \zeta^n u_0 + g [\zeta^{n-1} v_0 + \zeta^{n-2} v_1 + \cdots + \zeta^0 v_{n-1}], \quad (5.34)$$

onde v_i é a velocidade da i -ésima partícula do fluido que colidiu com a partícula Browniana.

A relação anterior define o que chamamos de *processos de saltos aditivos*.

É importante ressaltar que estes processos, em princípio, não são do mesmo tipo que os considerados na generalização do teorema de Doob no Apêndice III. Aqui, postulamos apenas que as transições se somam, enquanto a definição (8.27) é uma imposição mais forte sobre a variável aleatória u_t .

Na soma de variáveis independentes, tanto as médias quanto as variâncias se somam. Assim, é simples mostrar a partir de (5.34) que $q(u|u', n)$ resultante das n colisões possui média $\zeta^n u'$ e variância $\sum_{k=0}^n \sigma^2 \zeta^k = \sigma^2 \frac{1-\zeta^{2n}}{1-\zeta^2}$. A soma de variáveis Gaussianas independentes resulta em outra variável Gaussiana. Esse resultado caracteriza completamente a distribuição, dada abaixo

$$q(u|u', n) = \left[\frac{1 - \zeta^2}{2\pi\sigma^2(1 - \zeta^{2n})} \right]^{1/2} e^{-\frac{(u - \zeta^n u')^2}{2\sigma^2} \frac{1 - \zeta^2}{1 - \zeta^{2n}}}. \quad (5.35)$$

Usando a Eq. (5.35) podemos escrever diretamente o processo de saltos para a variável $u(t)$,

$$G_\tau(u|u') = \sum_{n=0}^{\infty} \frac{(\lambda\tau)^n e^{-\lambda\tau}}{n!} q(u|u', n). \quad (5.36)$$

A expressão anterior possui a inconveniência de se tratar de uma série infinita de termos. É possível truncar (5.36) e estimar o erro de fazê-lo de maneira muito simples. Basta observar que os fatores de Poisson $\frac{(\lambda\tau)^n}{n!} e^{-\lambda\tau}$ regulam a massa de probabilidade que cada termo $q(u|u', n)$ carrega. Além disto, os termos $q(u|u', n)$ convergem e a partir de um certo valor de n a diferença entre $q(u|u', n)$ e $q(u|u', \infty)$ se torna desprezível.

Um caso de interesse onde é possível avaliar (5.36) de forma fechada é quando $M = m$, representando um processo de auto-difusão. Aqui, $\zeta = 0$ e portanto a série em (5.36) se simplifica pois $q(u|u', n) = q(u|u')$ para $n \geq 1$. O resultado é

$$G_\tau(u|u') = e^{-\lambda\tau} \delta(u - u') + (1 - e^{-\lambda\tau}) \sqrt{\frac{m}{2\pi k_B T}} e^{-\frac{mu^2}{2k_B T}}, \quad (5.37)$$

e descreve uma escolha binária: existe uma probabilidade de $e^{-\lambda\tau}$ de permanecer no estado inicial e uma probabilidade $1 - e^{-\lambda\tau}$ de escolher um novo estado de acordo com $u \sim N(0, k_B T/m)$. É importante ressaltar que a solução dada pela Eq. (5.37) é uma particularidade do modelo unidimensional. Na colisão elástica de partículas idênticas em uma dimensão, elas simplesmente trocam de velocidades. Em duas ou mais dimensões, é necessário considerar os ângulos de espalhamento e o problema se complica enormemente.

5.2.2 Cumulantes

Apresentamos um método baseado na equação mestra que fornece equações diferenciais simples para os cumulantes da distribuição $G_\tau(u|u')$. Vale enfatizar que o cálculo pode ser facilmente adaptado para qualquer processo de saltos aditivo.

Comece multiplicando a Eq. (5.31) por e^{izu} e calculando a integral sobre u . O resultado é

$$\partial_\tau \psi_\tau(z|u') = \lambda \left[e^{-\frac{1}{2}\sigma^2 z^2} \psi_\tau(\zeta z|u') - \psi_\tau(z|u') \right], \quad (5.38)$$

onde definimos $\psi_\tau(z|u')$ como a função característica de $G_\tau(u|u')$. Os cumulantes estão associados com a expansão em série do logaritmo da função característica. Deste modo, definimos

$$\psi_\tau(z|u') = e^{W_\tau(z|u')} \quad (5.39)$$

onde a equação dinâmica para W é

$$\partial_\tau W_\tau(z|u') = \lambda \left[e^{-\frac{1}{2}\sigma^2 z^2 + W_\tau(\zeta z) - W_\tau(z)} - 1 \right]. \quad (5.40)$$

Agora, expandimos $W_\tau(z)$ em série

$$W_\tau(z|u') = \sum_{n=1}^{\infty} \frac{c_n(\tau; u')}{n!} (iz)^n,$$

onde $c_n(\tau; u')$ são os cumulantes da distribuição $G_\tau(u|u')$.

A equação para cada cumulante é obtida igualando ordem a ordem de z^n na Eq. (5.40). Lembramos que a condição inicial $G_0(u|u') = \delta(u - u')$ reflete nos cumulantes, implicando em $c_n(0; u') = 0$ exceto para a média que é $c_1(0; u') = u'$. Listamos as primeiras equações diferenciais para os cumulantes, onde definiu-se $\bar{c}_n = (1 - \zeta^n) c_n$,

$$\partial_\tau \bar{c}_1 = -\lambda (1 - \zeta^1) \bar{c}_1 \quad (5.41)$$

$$\partial_\tau \bar{c}_2 = -\lambda (1 - \zeta^2) [\bar{c}_2 - \sigma^2 - \bar{c}_1^2] \quad (5.42)$$

$$\partial_\tau \bar{c}_3 = -\lambda (1 - \zeta^3) [\bar{c}_3 - 3\bar{c}_1 (\bar{c}_2 - \sigma^2 - \bar{c}_1^2) + 4\bar{c}_1^3] \quad (5.43)$$

$$\partial_\tau \bar{c}_4 = -\lambda (1 - \zeta^4) [\bar{c}_4 - 3(\bar{c}_2 - \sigma^2 - \bar{c}_1^2)^2 - 4\bar{c}_1 \bar{c}_3 + 2\bar{c}_1^4] \quad (5.44)$$

É óbvio a partir do sistema de Eqs. (5.41-5.44) que todos os cumulantes relaxam exponencialmente para a condição de equilíbrio. No entanto, a velocidade com que esse processo ocorre depende do fator $(1 - \zeta^n)$ que faz com que cada cumulante relaxe com uma velocidade diferente dos demais. Uma vez que o estado de equilíbrio só é atingido para $|\zeta| < 1$, vemos que os cumulantes de mais alta ordem tendem a atingir o valor de equilíbrio mais rapidamente que os cumulantes de ordem baixa. Isso indica um processo de Gaussianização, já que a Gaussiana possui apenas os dois primeiros cumulantes não-nulos.

Definimos $g = 1 + \zeta = \frac{2M}{M+m}$ e $\bar{\gamma} = (1 - \zeta) \lambda = \frac{2m}{M+m} \lambda$ e escrevemos

explicitamente a solução para a média e o desvio padrão respectivamente como

$$c_1(\tau; u') = u' e^{-\bar{\gamma}\tau} \quad (5.45)$$

$$c_2(\tau; u') = \frac{k_B T}{M} [1 - e^{-g\bar{\gamma}\tau}] + (g u')^2 [e^{-g\bar{\gamma}\tau} - e^{-2\bar{\gamma}\tau}]. \quad (5.46)$$

É interessante comparar esse resultado com a solução da equação de Langevin (5.17-5.18). Para que os dois formalismos coincidam, é necessário que $g \rightarrow 2$ e $\bar{\gamma} \rightarrow \gamma$ (note que $\alpha = 2$). Esse resultado é possível somente no limite $M \gg m$.

A distribuição de equilíbrio pode ser avaliada fazendo o lado esquerdo das Eqs. (5.41-5.44) ir para zero. Nesse caso, fica claro que o único cumulante não nulo é a variância, que assume o valor $(1 - \zeta^2) \sigma^2 = \frac{k_B T}{M}$. É de certa forma impressionante que a lei de equipartição segue como uma consequência direta do modelo. Vale observar que, na equação de Langevin, é necessário impor a equipartição *a priori*.

5.2.3 Probabilidade de equilíbrio

Fisicamente, esperamos que o estado de equilíbrio $f(u) = G_\infty(u|u')$ seja descrito por $q(u|u', \infty)$, que representa a probabilidade de realizar uma transição com infinitas colisões. É possível justificar esta expectativa a partir da Eq. (5.36) pelo fato que a média e o desvio padrão da probabilidade de Poisson são respectivamente $\langle n \rangle = \lambda\tau$ e $\sigma_n = \sqrt{\lambda\tau}$. Deste modo, a massa de probabilidade na série (5.36) tende a se concentrar em torno dos fatores com $n \simeq \lambda\tau$ que para $\tau \rightarrow \infty$ correspondem a $q(u|u', n \rightarrow \infty)$. Assim, identificamos $f(u) = q(u|u', n \rightarrow \infty) = G_{\tau \rightarrow \infty}(u|u')$.

No modelo aqui considerado, a distribuição de equilíbrio corresponde a uma probabilidade Gaussiana com média nula e variância $\frac{k_B T}{m}$. Uma questão interessante surge se a distribuição de probabilidades para v na Eq. (5.28) não for uma Gaussiana. Nesse caso, para onde a distribuição de equilíbrio de $u(t)$

convergiria?

Para considerar esta situação, fazemos um raciocínio semelhante ao da seção anterior, mas utilizando uma distribuição arbitrária para v . O efeito disto é trocar o fator Gaussiano na Eq. (5.38) pela função característica da distribuição de velocidades, que denotamos por $e^{\Theta(z)}$. Usando a mesma notação anterior, é fácil mostrar que o estado de equilíbrio ($\partial_\tau \psi_{eq} = 0$) implica na relação

$$W(z) - W(\zeta z) = \Theta(z), \quad (5.47)$$

onde $W(z) \equiv \lim_{\tau \rightarrow \infty} W_\tau(z|u')$. Como o lado direito de (5.47) não depende de u' , $W(z)$ também não deve fazê-lo.

Usando o fato que a função característica é igual a 1 na origem e assim $W(0) = 0$, vemos imediatamente que se $\zeta = 0$ então $W(z) = \Theta(z)$. Deste modo, a distribuição de equilíbrio se iguala à distribuição das moléculas do fluido.

Para valores arbitrários de ζ , é difícil antever o comportamento assintótico. Ainda assim, a Eq. (5.47) permite determinar facilmente os cumulantes da distribuição de equilíbrio a partir da distribuição das partículas do fluido. Vemos por comparação direta que

$$c_n = \frac{k_n}{1 - \zeta^n}, \quad (5.48)$$

onde c_n são os cumulantes da probabilidade de equilíbrio e k_n são os cumulantes da distribuição do fluido.

Mostramos agora que no limite em que $M \gg m$ ou $m \gg M$, onde $|\zeta| \rightarrow 1$ a distribuição de equilíbrio da variável reduzida converge para uma Gaussiana. Esse é o mesmo tipo de convergência apresentada no Teorema do Limite Central.

Para ver isto, considere a variância $c_2 = k_2 / (1 - \zeta^2)$, que utilizamos para

definir a variável reduzida como

$$u^* = \frac{u - c_1}{\sqrt{c_2}}. \quad (5.49)$$

É óbvio pela definição, que a média de \bar{u} é nula e seu desvio padrão é igual a 1. Os outros cumulantes reduzidos são dados pela fórmula

$$c_n^* = \frac{c_n}{c_2^{n/2}} = \frac{k_n}{k_2^{n/2}} (1 - \zeta^2)^{n/2}. \quad (5.50)$$

Todos cumulantes reduzidos de ordem maior que 2 se anulam no limite $\zeta^2 \rightarrow \pm 1$, caracterizando portanto uma distribuição Gaussiana. O argumento mostra que se as massas M e m forem muito díspares, a Gaussianização ocorre mesmo se o fluido onde a partícula Browniana está imersa estiver em um estado estacionário não-Gaussiano.

5.2.4 Covariância e difusão

Os processos de saltos aditivos possuem uma função de covariância exponencial, como será demonstrado a seguir.

O primeiro passo consiste em obter a relação

$$\int du q(u|u', n) u = \zeta^n u' + g \langle v \rangle \sum_{k=0}^n \zeta^k, \quad (5.51)$$

que segue da Eq. (5.34). A média da distribuição de equilíbrio é

$$\mu_e = g \langle v \rangle \sum_{k=0}^{\infty} \zeta^k. \quad (5.52)$$

Comparando (5.51) e (5.52) com (5.36), obtemos facilmente

$$\int du G_\tau(u|u') (u - \mu_e) = \sum_{n=0}^n e^{-\lambda\tau} \frac{(\lambda\tau\zeta)^n}{n!} (u' - g\langle v \rangle) = e^{-\bar{\gamma}\tau} (u' - g\langle v \rangle), \quad (5.53)$$

onde usamos a definição $\bar{\gamma} \equiv (1 - \zeta) \lambda$.

Utilizando o resultado anterior, a covariância se escreve como

$$\int du f(u) (u - \mu_e)^2 e^{-\bar{\gamma}\tau} = (1 - \zeta^2) g^2 \langle \langle v^2 \rangle \rangle e^{-\bar{\gamma}\tau}. \quad (5.54)$$

Nesse ponto, é conveniente identificar a variância da distribuição de velocidades das moléculas do fluido com $\frac{k_B T}{m}$. Se isto for correto, então a covariância pode ser escrita como

$$\langle \langle u(t)u(t') \rangle \rangle = \frac{k_B T}{M} e^{-\bar{\gamma}|t-t'|}, \quad (5.55)$$

que caracteriza um comportamento comum a diversos processos Markovianos.

A difusão pode ser calculada utilizando as mesmas considerações feitas para obter a Eq. (5.16), de onde resulta

$$\langle x(\tau)^2 \rangle = 2 \frac{k_B T}{M \bar{\gamma}} \tau. \quad (5.56)$$

5.2.5 Equação de Kramers-Moyal

Obtemos os coeficientes da expansão de Kramers-Moyal (2.39) relacionando-os com os momentos da distribuição de saltos $q(u|u')$. É possível computá-los explicitamente a partir da função característica $\phi(z|u)$ usando a fórmula

$$D_n(u) = \frac{\lambda}{n!} (-i)^n \left. \frac{d^n}{dz^n} \phi(z|u) \right|_{z=0} = \frac{\lambda}{n!} (-i)^n \left. \frac{d^n}{dz^n} e^{i(\zeta-1)uz - \frac{1}{2}\sigma^2 z^2} \right|_{z=0}. \quad (5.57)$$

Ainda que o cálculo acima seja razoavelmente simples, convém considerar

uma aproximação válida para $r \equiv m/M \ll 1$. Reescrevemos os termos

$$1 - \zeta = \frac{2r}{1+r}; \quad \sigma^2 = \frac{4r^2}{(1+r)^2} \frac{k_B T}{m}. \quad (5.58)$$

Uma vez que o segundo termo é de $O(r^2)$, é possível aproximar a expressão (5.57) expandindo o termo com σ^2 no expoente até a primeira ordem em r . O resultado final pode ser avaliado explicitamente como

$$D_n(u) \simeq \frac{\bar{\gamma}}{2n!} (-2u)^n r^{n-1} - \frac{(-2u)^{n-2}}{(n-2)!} \frac{\bar{\gamma} k_B T}{M} r^{n-2}, \quad (5.59)$$

que é exato na ordem principal de r . Vale ressaltar que $\bar{\gamma}$ depende de r pela expressão $\bar{\gamma} = 2r/(1+r)\lambda$ e assim o limite $r \rightarrow 0$ implica que todos os coeficientes de Kramers-Moyal se anulam.

A solução completa da equação de de Kramers-Moyal é relevante apenas para descrever o regime transiente. Para tempos longos, $G_\tau(u|u') \rightarrow q(u|u', \infty)$, que é a distribuição de Maxwell. O problema original que motivou Kramers a estudar as expansões com derivadas de várias ordens era justamente avaliar quais equações diferenciais possuem a Gaussiana como solução estacionária [37]. É possível mostrar, no caso aqui considerado, que qualquer truncagem em ordem maior ou igual a 2 possui a mesma solução estacionária que a equação completa.

Toda truncagem da equação de Kramers-Moyal recupera *precisamente* os momentos até a ordem de truncagem. Por exemplo, ao truncar até $n = 4$, a média, a variância, o fator de assimetria e a kurtosis da solução aproximada seriam os mesmos da solução completa.

O problema de reconstruir uma distribuição a partir de um conjunto finito de momentos é amplamente estudado em estatística. Em geral, é mais conveniente utilizar os cumulantes, mas estes podem ser facilmente calculados a

partir dos momentos. Um método comum é o da expansão de Edgeworth [8]

$$f(x) dx = \frac{e^{-x^2}}{\sqrt{2\pi}} \left[1 + \frac{\bar{\kappa}_3 H_3(x)}{6\sqrt{n}} + \frac{\bar{\kappa}_4 H_4(x)}{24n} + \frac{\bar{\kappa}_3^2 H_6(x)}{72n} + \dots \right] dx, \quad (5.60)$$

onde $\bar{\kappa}_n$ são os cumulantes reduzidos e $x = \frac{u - \langle u(t) \rangle}{\sqrt{\langle (u(t)^2) \rangle}}$. Vale notar que a série de Edgeworth não gera necessariamente densidades positivas. Existem outros métodos mais sofisticados que contornam esse problema, mas esse assunto foge do escopo da tese.

5.2.6 Equação de Langevin

Estritamente falando, não é possível antecipar a convergência da expansão de Kramers-Moyal em (5.57), para a equação de Fokker-Planck, mesmo quando $M \gg m$. O resultado depende da relação entre M e λ , mas estas são duas grandezas físicas independentes *a priori*.

Usando a Eq. (5.57) junto com a definição $\bar{\gamma} = (1 - \zeta) \lambda$, obtemos a força de fricção

$$D_1(u) = -M \frac{2r \lambda}{1+r} u \simeq -\frac{2\rho\bar{\sigma}}{(1+r)} \sqrt{\frac{8 k_B T}{\pi m}} u, \quad (5.61)$$

vale enfatizar que esse resultado vale para qualquer razão entre massas $r = m/M$. Diferentemente da lei de Stokes, que cresce linearmente com as dimensões da partícula Browniana, o resultado que obtemos é linear com a seção de choque e usualmente dependeria do raio da partícula Browniana ao quadrado. Isto ilustra que a força de fricção devido ao choque com as partículas do fluido é gerada por um mecanismo diferente da lei de Stokes.

Um fato interessante a respeito do modelo apresentado é que a truncagem da expansão de Kramers-Moyal até a segunda ordem não recupera o modelo de Langevin. Isto porque o coeficiente de difusão⁴ calculado a partir de (5.57)

⁴Não confundir com o coeficiente de difusão relacionado às posições dado pela fórmula de Einstein.

seria dado por

$$D_2(u) = \frac{\bar{\gamma}}{1+r} \left(\frac{k_B T}{M} + r u^2 \right), \quad (5.62)$$

que possui um termo adicional proporcional a $r u^2$. Esse termo viola uma suposição comum que a força aleatória deveria ser independente da velocidade. No entanto, ao considerar a Eq. (5.62), vemos que sua origem física é muito simples: o impulso transmitido por cada colisão depende da velocidade relativa entre as partículas do fluido e a partícula Browniana. No referencial do laboratório, a velocidade relativa antes do choque se escreve em função tanto de u quanto de v . Essa é, fundamentalmente, a razão para o aparecimento de um termo de fricção linear em u e também de um termo de difusão proporcional a u^2 na equação de Fokker-Planck associada ao processo.

É lógico que poderíamos desprezar esse termo adicional no coeficiente de difusão se $r \ll 1$. Esse é o limite onde o modelo de Langevin se aplica como aproximação. Vale observar que na situação oposta, onde as partículas Brownianas são muito leves ($r \geq 1$), é necessário fazer uma correção adicional à $D_1(u)$ e $D_2(u)$: a rigor, a taxa de transições λ depende da velocidade da partícula Browniana e, portanto, nem mesmo o termo de drift do modelo de Langevin tradicional seria recuperado.

5.3 Conclusão

O modelo de Langevin para o movimento Browniano se baseia em uma idéia aparentemente poderosa e de caráter geral. Para construir um modelo estocástico a partir de um modelo determinístico descrito por uma equação diferencial, bastaria acrescentar um termo de ruído aleatório. Mais concretamente, de posse do modelo

$$\frac{dx}{dt} = Y(x, t), \quad (5.63)$$

a prescrição de Langevin induziria a escrever o modelo estocástico correspondente como

$$\frac{dx}{dt} = Y(x, t) + \varphi(t), \quad (5.64)$$

onde $\varphi(t)$ é uma influência aleatória.

No entanto, existem vários problemas com esta abordagem. Primeiramente, não existe uma forma universal para o ruído aleatório $\varphi(t)$: a “escolha natural” do ruído branco é motivada apenas pela conveniência matemática e é muitas vezes difícil de se justificar⁵. A influência aleatória não pode ser decidida em um nível puramente matemático: é necessário investigar a origem física do ruído e assim modelá-lo adequadamente.

No entanto, o problema com (5.64) é ainda mais grave que isto: nem mesmo o termo de drift pode ser determinado univocamente. É necessário definir como a lei determinística (5.63) se relaciona com a distribuição de probabilidades vinda de (5.64). Esse alerta é dado enfaticamente por van Kampen, que recomenda interpretar a equação diferencial estocástica como um objeto puramente formal. Uma equação como (5.64) pode ajudar a visualizar certas propriedades das funções aleatórias $x(t)$, mas não representa necessariamente uma definição legítima de um processo estocástico. Nesse contexto, vale mencionar o imbróglio Itô vs. Stratonovich: a mesma equação de Langevin resulta em diferentes equações de Fokker-Planck caso se utilize uma definição de integral estocástica ou a outra [31].

Uma maneira razoável de produzir uma generalização estocástica da Eq. (5.63), seria definir uma probabilidade de transição onde a média $\langle x(t) \rangle$ seria dada pela solução desta equação. Essa prescrição não equivale à equação diferencial estocástica (5.64), a não ser que Y seja uma função linear de x . Para vermos

⁵O argumento de que o ruído deve ser decorrelacionado, apenas garante a Markovianidade de $x(t)$, mas existe uma infinidade de possibilidades consistentes com esta imposição.

isto, bastaria avaliar a média de ambos os lados de (5.64),

$$\frac{d\langle x \rangle}{dt} = \langle Y(x, t) \rangle \neq Y(\langle x \rangle, t). \quad (5.65)$$

Van Kampen mostra como montar uma equação diferencial estocástica onde o lado esquerdo de (5.64), ao menos aproximadamente, seria igual ao lado direito [32]. Não vamos nos aprofundar nesta discussão, que é longa e cheia de sutilezas matemáticas. O ponto que queríamos fazer é o seguinte: equações como a de Langevin são construtos formais: para definir um processo estocástico legítimo é necessário especificar as probabilidades a N -pontos ou equações que permitam determinar estas probabilidades.

Isto nos leva à discussão sobre o gás de esferas rígidas. Muitas vezes, esse sistema é descrito utilizando o formalismo de Langevin, identificando a equação de Langevin com a lei de Newton. No entanto, esta identificação não é satisfatória: existe uma série de questões a serem respondidas antes de interpretar o objeto formal da equação de Langevin como a definição de um processo estocástico. A começar: $\varphi(t)$ depende de $x(t)$? quais as propriedades estatísticas de $\varphi(t)$? as regras adotadas supõe o cálculo de Itô ou Stratonovich?

No modelo para o gás de esferas rígidas discutido no Cap. 5, evitamos todas estas questões pois não se escreveu uma equação diferencial estocástica para descrever a dinâmica da partícula Browniana. A aleatoriedade reflete o fato que não se pode antecipar a velocidade da próxima partícula do fluido que colidirá com a partícula Browniana. Sua origem física é transparente.

O modelo do gás de esferas rígidas também ilustra um ponto interessante: muitas vezes o processo responsável por introduzir a aleatoriedade na relação determinística (5.63) é do tipo descontínuo. Isto é verdade no gás de esferas rígidas, onde a partícula Browniana realiza trajetórias retilíneas até que uma colisão muda repentinamente sua velocidade para uma outra direção. O

mesmo pode ser dito de outros processos: um ativo financeiro mantém seu valor de mercado até que ocorra uma transação, uma população se mantém constante até que ocorra a morte ou o nascimento de um indivíduo. Em várias situações, o processo estocástico é, no fundo, um processo de saltos. As equações diferenciais estocásticas, como a Eq. (5.64) seriam encaradas apenas como uma aproximação para o processo real (obviamente é necessário verificar se tais aproximações seriam muito grosseiras ou não).

O ponto geral que queríamos destacar é esse: processos de saltos geralmente oferecem uma linguagem mais transparente que as equações diferenciais estocásticas. O efeito da aleatoriedade é separado de uma eventual evolução determinística quando especificamos a distribuição de saltos e a taxa de ocorrência dos mesmos. A equação diferencial estocástica é um objeto mais problemático: existem certas dificuldades matemáticas que não são óbvias em princípio (e.g., cálculo de Itô vs. Stratonovich). Em várias circunstâncias estes problemas impedem, ou ao menos dificultam, a interpretação de (5.64) como uma definição legítima de um processo estocástico.

No caso específico da equação de Langevin, é necessário ainda discriminar o regime hidrodinâmico, onde a partícula Browniana possui dimensões macroscópicas, do regime molecular, onde o caráter descontínuo do fluido deve ser contabilizado. Em cada caso, a origem do termo de arrasto e do termo de flutuação é diferente, ainda que os dois modelos se assemelhem do ponto de vista formal.

Capítulo 6

Equação de Langevin Generalizada

Os sistemas de resposta linear são aqueles que uma influência externa não altera a matriz de covariância de equilíbrio. Kubo mostrou que é possível formular tais sistemas a partir de uma equação de Langevin generalizada que possui com uma força de fricção retardada, um ruído não-branco e satisfaz a relação de Nyquist generalizada (ou teorema de flutuação e dissipação).

No entanto, existem sistemas que não satisfazem a relação Nyquist e ainda assim são bem descritos pela teoria de resposta linear. Construimos um destes sistemas explicitamente e encontramos as condições para que a relação de Nyquist seja satisfeita. A equação proposta por Kubo não permite formular estes sistemas. Consideramos algumas generalizações diferentes da equação de Langevin e argumentamos que aquela proposta por Kubo é a menos plausível.

No entanto, a generalização da equação de Langevin não é unívoca. Argumentamos que a escolha de Kubo é particularmente infeliz: se interpretada como uma lei de Newton, ela assume propriedades inaceitáveis para as forças aleatórias. Se interpretada como uma teoria de resposta linear ela praticamente impõe o teorema de dissipação e flutuação (hipótese de Onsager) que relaciona a admitância e covariância de um sistema linear. Ainda assim, Kubo oferece

uma demonstração que parece particularmente atraente pois se baseia em hipóteses mínimas. Existem outras generalizações da equação de Langevin que não sofrem das mesmas críticas. No entanto, a hipótese de Onsager não resulta facilmente do formalismo e permanece, como o nome diz, como uma mera hipótese. Em vista desta discussão, obtemos a covariância e a resposta linear de um sistema de osciladores clássicos e mostramos que a hipótese de Onsager pode ser violada de uma maneira trivial.

Uma questão importante na física de sistemas macroscópicos é inferir a resposta de uma variável termodinâmica, que aqui denotamos por $u(t)$ a uma influência externa, ou “força termodinâmica” $K(t)$. Um sistema de resposta linear admite que estas duas grandezas se relacionam por

$$\langle u(t) \rangle = \int_{-\infty}^t dt Y(t-s) K(s). \quad (6.1)$$

Assumimos que $u(t)$ é uma grandeza sujeita a flutuações. A matriz de covariância de $u(t)$ é então definida como

$$\Sigma(\tau) = \langle \langle u(t) u(t+\tau) \rangle \rangle, \quad (6.2)$$

onde se impõe a simetria por translação temporal. Geralmente os dois núcleos $Y(\tau)$ e $\Sigma(\tau)$ são relacionados. A relação de Nyquist generalizada, ou teorema de flutuação e dissipação, estabelece que

$$\Sigma(\tau) = mk_B T Y(|\tau|), \quad (6.3)$$

onde m é a massa da partícula e T a temperatura do banho térmico. Essa relação foi demonstrada por Nyquist no contexto de circuitos elétricos [52] e por Callen e Welton para sistemas Hamiltonianos [7].

Posteriormente, Kubo mostrou que um sistema de resposta linear que obedece à relação de flutuação e dissipação pode ser reescrito de maneira mais sugestiva como uma equação de Langevin generalizada[38],

$$\frac{du}{dt} = - \int_0^t ds g(t-s)u(s) + \frac{\varphi(t)}{m} + \frac{K(t)}{m}, \quad (6.4)$$

onde $\varphi(t)$ representa o ruído que induz as flutuações de $u(t)$ e $g(t-s)$ é interpretado como o núcleo de uma força de fricção retardada. A equação de Kubo lembra a lei de Newton, e permite pensar nas forças termodinâmicas como se fossem foras usuais. Mas aqui vale uma ressalva: as três forças $\frac{\varphi(t)}{m}$, $K(t)$ e $-\int_{-\infty}^t ds g(t-s)u(s)$ são entendidas em um contexto abstrato, e nem sempre representam forças legítimas. Por exemplo, $K(t)$ pode representar o gradiente de temperatura, que é “força” que direciona o fluxo de calor ($u(t)$).

Uma dificuldade com a expressão proposta por Kubo é que mesmo no caso em que $K(t) = 0$, os processos $u(t)$ e $\varphi(t)$ não podem ser simultaneamente estacionários [21, 15]. Essa situação pode ser contornada facilmente reformulando o termo de dissipação em (6.4), tornando-o explicitamente invariante por translação temporal:

$$\frac{du}{dt} = - \int_{-\infty}^t ds g(t-s)u(s) + \frac{\varphi(t)}{m} + \frac{K(t)}{m}. \quad (6.5)$$

É óbvio que (6.4) e (6.5) não correspondem à mesma equação, mas é possível mostrar que sob certas condições, as variáveis relevantes na teoria de resposta linear (admitância e covariância de $u(t)$) são as mesmas.

A relação de Nyquist generalizada para a equação de Kubo e para a Eq. (6.5) pode ser escrita como

$$\langle \varphi(t)\varphi(0) \rangle = m k_B T g(t); \quad \langle \varphi(t) \rangle = 0. \quad (6.6)$$

Essa é a versão de Kubo para o teorema de flutuação e dissipação. O objetivo deste capítulo é mostrar que, mesmo em sistemas microscópicos simples, a relação de flutuação e dissipação pode ser violada. Usamos o sistema de osciladores clássicos como exemplo e definimos $u(t)$ como um processo estocástico para a velocidade de um dos osciladores. A relação de flutuação e dissipação é respeitada somente quando a força externa interagir *apenas* com o oscilador estudado. Na hipótese mais realística que a interação influencia todos os osciladores, a relação (6.6) será violada.

Entendemos a equação de Langevin generalizada como a definição de um processo estocástico. Em contraste, Lee e Mori [41] mostraram que é possível obter uma equação como (6.4) a partir da lei de Heisenberg. Deste modo, é possível utilizar a equação de Langevin generalizada como uma formulação alternativa para a mecânica quântica, onde a variável $u(t)$ em (6.4) seria um operador. Vale notar que a existência de processos estocásticos em sistemas quânticos é um tema controverso. A mecânica quântica fornece uma prescrição para definir a probabilidade conjunta de operadores que comutam. Como um operador de Heisenberg $\hat{A}(t_1)$ geralmente não comuta com $\hat{A}(t_2)$, não existe uma maneira bem definida de escrever a probabilidade a dois pontos $P(A_1, A_2)$, além das probabilidades para outras malhas temporais. Isto nos limita a sistemas clássicos e exclui o formalismo de Mori da discussão.

6.1 Soluções da equação de Langevin

6.1.1 Equação estacionária

Dado um núcleo de fricção $g(t - t')$ e a covariância do ruído $\Gamma(t - t')$, qual é o processo para $u(t)$? Como obtemos sua densidade espectral e a média?

Tratamos, primeiramente, da forma estacionária da equação de Langevin generalizada (ou Eq. [6.5]). É conveniente reescrever esta equação no espaço de

Fourier já a mesma que envolve derivadas e convoluções. Assim, definimos os termos $\tilde{u}(\omega)$, $\tilde{g}(\omega)$, $\tilde{\varphi}(\omega)$ e $\tilde{K}(\omega)$ como as transformadas de Fourier de $u(t)$, $g(t)$, $\varphi(t)$ e $K(t)$ ¹ de tal maneira que a equação de Langevin vira

$$-i\omega\tilde{u}(\omega) = -\tilde{g}(\omega)\tilde{u}(\omega) + \frac{1}{m} \left(\tilde{\varphi}(\omega) + \tilde{K}(\omega) \right). \quad (6.7)$$

Na passagem de (6.5) para (6.7), definimos $g(t) = 0$ quando $t < 0$, o que permite utilizar o teorema da convolução para escrever o primeiro termo no lado direito de (6.7).

Deste modo, podemos organizar $\tilde{u}(\omega)$ como uma transformação linear sobre as forças aleatórias. Ou seja,

$$\tilde{u}(\omega) = \frac{1}{m\tilde{g}(\omega) - i\omega} \left[\tilde{\varphi}(\omega) + \tilde{K}(\omega) \right] \equiv \tilde{Y}(\omega) \left[\tilde{\varphi}(\omega) + \tilde{K}(\omega) \right]. \quad (6.8)$$

Montando o mesmo argumento para o valor médio de $\langle u(t) \rangle$, obtemos

$$\tilde{\mu}(\omega) = \frac{1}{m\tilde{g}(\omega) - i\omega} \tilde{K}(\omega) \equiv \tilde{Y}(\omega) \tilde{K}(\omega), \quad (6.9)$$

onde $\mu(t)$ é a média de $u(t)$ e $\tilde{\mu}(\omega)$ é sua transformada de Fourier. O operador $\tilde{Y}(\omega)$ que aparece em (6.8) e (6.9) é a admitância da variável $u(t)$ com relação à força $K(t)$. Revertendo a Eq. (6.9) para t , obtemos

$$\langle u(t) \rangle = \int_{-\infty}^{\infty} dt' Y(t-t') K(t'), \quad (6.10)$$

onde definiu-se $Y(\tau) = \int d\omega \tilde{Y}(\omega) \frac{e^{-i\omega\tau}}{2\pi}$.

A causalidade é preservada se a admitância for nula para argumentos negativos. É possível mostrar que isto acontece se e somente se a mesma propriedade ocorre em $g(t-t')$, sendo que ambos estão relacionados diretamente pela Eq. (6.8). Essa propriedade foi imposta na Eq. (6.5) ao escrever a força

¹Aqui adotamos a convenção que $\tilde{x}(\omega) = \int dt x(t) e^{it\omega}$.

de fricção como uma integral de $-\infty$ até t ao invés de se estendê-la sobre toda a reta real.

O operador $Y(t-t')$ determina tanto a proporcionalidade de $u(t)$ com $\varphi(t)$ no caso onde não há perturbação externa, quanto entre $\langle u(t) \rangle$ e $K(t)$ quando há perturbação. De um certo modo, o sistema responde a uma flutuação interna da mesma maneira que responde macroscopicamente a uma força externa. Essa observação pode ser entendida como a definição do termo flutuativo na Eq. (6.5).

O cálculo da matriz de covariância de $u(t)$ é feito substituindo a Eq. (6.8) em Eq. (4.21),

$$\tilde{\Sigma}(\omega) = \frac{1}{m^2} \frac{\tilde{\Gamma}(\omega)}{|\tilde{g}(\omega) - i\omega|^2} = \frac{\Re[\tilde{Y}(\omega)]}{\Re[\tilde{g}(\omega)]} \tilde{\Gamma}(\omega). \quad (6.11)$$

Segundo o teorema de Bochner, basta que $\tilde{\Sigma}(\omega) \geq 0$ para que isto aconteça e a Eq.(6.11) garante a positividade de $\tilde{\Sigma}(\omega)$. Do ponto de vista da consistência matemática, os núcleos $\tilde{g}(\omega)$ e $\tilde{\Gamma}(\omega)$ são independentes e qualquer escolha razoável gera uma densidade espectral $\tilde{\Sigma}(\omega)$ válida.

No contexto de uma teoria de resposta linear, o processo estocástico para $u(t)$ é definido pela matriz de covariância (ou sua densidade espectral $\tilde{\Sigma}(\omega)$) e pelo operador de admitância $\tilde{Y}(\omega)$ que determina a média a partir da perturbação externa $K(t)$. Tanto a força flutuativa quanto a fricção retardada são, de certo modo, apenas forças fictícias que nos permitem interpretar a relação de resposta linear como a segunda lei de Newton. É óbvio pela Eq. (6.8) que o núcleo da força de fricção pode ser escrito biunivocamente como função da admitância,

$$\tilde{Y}(\omega) = \frac{1}{m} \frac{1}{\tilde{g}(\omega) - i\omega}; \quad \tilde{g}(\omega) = \frac{1}{m\tilde{Y}(\omega)} + i\omega. \quad (6.12)$$

Uma vez dado $\tilde{g}(\omega)$ e $\tilde{\Sigma}(\omega)$, processo estocástico para $\varphi(t)$ fica então definido

por

$$\tilde{\Gamma}(\omega) = m^2 |\tilde{g}(\omega) - i\omega|^2 \tilde{\Sigma}(\omega). \quad (6.13)$$

A admitância é uma relação entre a força externa e $\langle u(t) \rangle$. Na ausência de forças externas, a admitância seria indeterminada, assim como $\tilde{g}(\omega)$. Conseqüentemente, mesmo se conhecêssemos a o processo estocástico para $u(t)$, não seria possível definir univocamente a matriz de covariância do ruído $\Gamma(t - t')$ utilizando a Eq. (6.13). É necessário considerar uma força externa para separar a influência do núcleo de fricção e do núcleo de dissipação. Essa ambigüidade já aparecia na equação de Langevin tradicional e foi discutida na Sec. 5.1.4.

Caso não se conheça a admitância, diferentes equações de Langevin poderiam reproduzir o mesmo processo estocástico $u(t)$. É possível, por exemplo, escolher arbitrariamente o termo de dissipação e ainda assim produzir uma dada densidade espectral $\tilde{\Sigma}(\omega)$. Para isto, basta redefinir o ruído como $\tilde{\Gamma}' = \frac{|\tilde{g}' - i\omega|^2}{|\tilde{g} - i\omega|^2} \tilde{\Gamma}$ onde tanto as variáveis com linha como as sem linha definem a mesma densidade espectral.

Nos casos em que a relação de Nyquist é válida, as relações obtidas anteriormente se simplificam. A Eq. (6.3) pode ser reescrita no espaço de Fourier como

$$\tilde{\Sigma}(\omega) = m k_B T 2\Re \left[\tilde{Y}(\omega) \right], \quad (6.14)$$

onde usamos o fato que a transformada de Fourier de $Y(|\tau|)$ é $2\Re \left[\tilde{Y}(\omega) \right]$. Ao substituir esse resultado em (6.13), resulta que

$$\tilde{\Gamma}(\omega) = m k_B T 2\Re \left[\tilde{g}(\omega) \right], \quad (6.15)$$

que é a versão no espaço de Fourier da relação de flutuação e dissipação de Kubo (6.6).

6.1.2 Equação não-estacionária: Kubo

Na teoria de resposta linear, o estado de equilíbrio é atingido quando $K(t) = 0$. Ao tratar a equação de Kubo nesse regime, surge um certo dilema: não é possível que os processos estocásticos para $\varphi(t)$ e $u(t)$ sejam simultaneamente estacionários [21, 15]. Ao interpretar a Eq. (6.4) como uma lei de Newton, talvez pareça fisicamente mais intuitivo assumir que $\varphi(t)$ seja estacionário, pois descreveria o mecanismo que faz o sistema em equilíbrio produzir as flutuações. A não-estacionariedade de $u(t)$ seria atribuída ao preparo de $u(t)$ em uma condição inicial específica. No entanto, Kubo faz justamente o oposto e postula que $u(t)$ corresponde ao processo estacionário.

Assim como no caso anterior, a covariância $\Sigma(t - t')$ pode ser escrita como função dos núcleos $g(t - t')$ e $\Gamma(t) = \langle \varphi(t)\varphi(t') \rangle|^{t'=0}$ (é necessário escolher um instante específico $t' = 0$, pois $\varphi(t)$ não é um processo estacionário). Multiplicando a Eq. (6.4) por $\varphi(0) = m \left. \frac{du}{dt} \right|^{t=0}$ e tomando a média, obtemos

$$-\frac{d^2\Sigma(t)}{dt^2} = \int_0^t ds g(t-s) \frac{d\Sigma(s)}{ds} + \frac{1}{m} \langle \varphi(t)\varphi(0) \rangle. \quad (6.16)$$

A solução pode ser obtida pela transformada de Fourier-Laplace definida como $\tilde{\Sigma}_L(\omega) = \int_0^\infty d\tau \Sigma(\tau)e^{i\omega\tau}$. Como a informação sobre as funções $\Sigma(\tau)$, $\Gamma(\tau)$ e $g(\tau)$ está essencialmente contida apenas em metade da reta real, as transformadas de Fourier-Laplace aqui consideradas de algum modo recuperam as transformadas de Fourier definidas anteriormente.

Multiplicando a Eq. (6.16) por $e^{i\omega t}$ e integrando em t , obtemos

$$-\omega^2 \tilde{\Sigma}_L(\omega) = (-i\omega) \tilde{g}_L(\omega) \tilde{\Sigma}_L + \frac{1}{m} \tilde{\Gamma}_L(\omega). \quad (6.17)$$

Uma vez que $g(\tau) = 0$ para $\tau < 0$, então temos $\tilde{g}(\omega) = \tilde{g}_L(\omega)$, onde $\tilde{g}(\omega)$ foi definido na seção anterior. Já para as matrizes de covariância temos

$$\tilde{\Sigma}(\omega) = 2\Re \left[\tilde{\Sigma}_L(\omega) \right] \text{ e } \tilde{\Gamma}(\omega) = 2\Re \left[\tilde{\Gamma}_L(\omega) \right].$$

A densidade espectral é um pouco mais complicada que a obtida na seção anterior, mas ainda assim depende apenas da covariância do ruído e do núcleo de fricção

$$\tilde{\Sigma}(\omega) = \frac{(\tilde{g}_I(\omega) + \omega) \tilde{\Gamma}(\omega) - 2\tilde{g}_R(\omega) \Im \left[\tilde{\Gamma}_L(\omega) \right]}{\omega m^2 |\tilde{g}(\omega) - i\omega|^2}. \quad (6.18)$$

Na passagem anterior, separamos $\tilde{g}(\omega)$ nas partes real e imaginária, $\tilde{g}(\omega) = \tilde{g}_R(\omega) + i\tilde{g}_I(\omega)$. Vemos claramente que (6.18) não equivale a (6.11). Mais ainda, é possível que (6.18) não represente uma matriz de covariância válida já que depende do numerador ser maior que zero para ω 's positivos.

É um fato curioso que (6.18) e (6.11) tornam-se idênticos se vale a relação abaixo

$$\tilde{\Gamma}(\omega) = 2m k_B T \tilde{g}_R(\omega) \quad (6.19)$$

(e indiretamente $\Im \left[\tilde{\Gamma}_L(\omega) \right] = m k_B T \tilde{g}_I(\omega)$). Vários termos se cancelam e a covariância (6.18) seria a mesma obtida para a equação de Langevin estacionária. É fácil mostrar que o núcleo de admitância é idêntico a (6.12), independentemente de valer a relação de flutuação e dissipação, ou não.

6.1.3 Equação não-estacionária: Volterra

Essa é dedicada à solução da equação de Kubo (6.4) assumindo que $\varphi(t)$, e não $u(t)$, é um processo estacionário. A solução espelha o argumento utilizado para tratar a equação de Langevin tradicional, mas utiliza transformadas de Laplace e o formalismo de Volterra para equações integro-diferenciais. Fora isto, o tratamento é semelhante àquele dado à equação estacionária. Definimos $u_L(z), g_L(z), \varphi_L(z)$ e $K_L(z)$ como as transformadas de Laplace de $u(t), g(t), \varphi(t)$ e $K(t)$. Usando (6.4), calculamos

$$u_L(z) = \frac{1}{m} \frac{1}{g_L(z) + z} [m u_0 + \varphi_L(z) + K_L(z)], \quad (6.20)$$

onde a condição inicial $u_0 = u(t = 0)$ entra automaticamente na solução. Identificamos a admitância como

$$Y(t) = \mathcal{L}^{-1} \left[\frac{1}{m} \frac{1}{g_L(z) + z} \right], \quad (6.21)$$

que formalmente é o mesmo valor que (6.12) (comparamos os dois por continuação analítica, fazendo $z \mapsto -i\omega$ e $g_L(i\omega) = \tilde{g}(-\omega)$).

Mesmo na ausência de forças externas, a média de $\langle u(t) \rangle$ não é nula

$$\langle u(t) \rangle = u_0 m Y(t) + \int_0^t ds Y(t-s) K(s), \quad (6.22)$$

onde na expressão anterior usamos a relação de causalidade $Y(\tau) = 0$ para $\tau < 0$ para redefinir os limites da integral. A matriz de covariância deste processo é mais complicada pois depende do tempo inicial. Invertendo a transformada de Laplace na Eq. (6.21) e subtraindo a Eq. (6.22), obtemos

$$u(t) - \langle u(t) \rangle = \int_0^t ds Y(t-s) \varphi(s). \quad (6.23)$$

Multiplicamos a expressão anterior avaliada em dois tempos distintos e tiramos a média para assim obter

$$\Sigma(t, t') = \int_0^t ds \int_0^{t'} ds' Y(t-s) Y(t'-s) \Gamma(s-s'). \quad (6.24)$$

O processo estocástico para $u(t)$ fica totalmente definido pelos núcleos de fricção e a matriz de covariância para as flutuações. Fica claro que quando $t = t' = 0$, a matriz de covariância se anula e portanto é possível interpretar o resultado como um processo condicionado a $u(0) = u_0$.

Assim como no caso estacionário, não existe nenhum compromisso entre os núcleos $g(t-t')$ e $\Gamma(t-t')$ para garantir a consistência matemática de (6.24). A

Eq. (6.24) apenas descreve a regra de transformação de matrizes de covariância para variáveis relacionadas por uma transformação linear e portanto também define uma matriz de covariância válida.

6.2 Sistema de osciladores clássicos

É possível obter a covariância e a admitância de maneira exata se conseguirmos integrar as equações de movimento microscópicas. Ainda que geralmente isso não seja possível, vale a pena desenvolver esse raciocínio explicitamente em um sistema de muitas partículas onde as equações de movimento são integráveis. Temos em mente um sistema de osciladores sujeito à forças externas, cuja Hamiltoniana é

$$H(q, v) = \sum_{i=0}^n \frac{p_i^2}{2m_i} + \sum_{i=0}^n \sum_{j=0}^n \frac{K_{ij}q_iq_j}{2} + \sum_{i=0}^n q_i F_i(t). \quad (6.25)$$

Nesse caso (e possivelmente apenas esse) é possível resolver as equações microscópicas para um número n arbitrário de partículas. A demonstração que se segue se assemelha à Ref. [9], mas consideramos o termo da força para obter explicitamente a admitância enquanto estes autores tentam identificar diretamente os termos de flutuação e dissipação.

Supomos que m_0 , $p_0(t)$ e $q_0(t)$ são as coordenadas da partícula Browniana, e assim identificamos $u(t) = p_0(t)/m_0$ para manter a notação consistente com a seção anterior. O sistema descrito pela Eq. (6.25) consiste portanto em um conjunto de n osciladores harmônicos em contato térmico com uma partícula, a qual monitoramos o estado. Assumimos que as forças externas podem ser escritas como $F_i(t) = g_i K(t)$, onde $g_0 = 1$. Fisicamente, esta suposição descreve a aplicação de um campo externo dependente do tempo, onde cada partícula sofre uma força proporcional ao valor de sua carga g_i e ao próprio campo $K(t)$.

Calculamos a resposta linear do sistema ao campo externo $K(t)$ ligado em

$t = 0$. Para $t < 0$, o sistema se encontra em um estado de equilíbrio descrito pela distribuição canônica

$$P(p, q) = \frac{1}{Z} e^{-\beta H(p, q)} dq_0 \dots dq_n dp_0 \dots dp_n, \quad (6.26)$$

É fácil mostrar, utilizando a Hamiltoniana (6.25), que para $t < 0$ os valores médios se anulam, $\langle v_i(0) \rangle = \langle q_i(0) \rangle = 0$.

A integração formal das equações de movimento para as coordenadas $v_i(t)$ é expressa como

$$p_i(t) = f_i(q(0), p(0); t). \quad (6.27)$$

Em um sistema de osciladores, esta é uma função linear nas condições iniciais, como será demonstrado adiante. Por esse motivo, a resposta do oscilador harmônico é exatamente linear na força externa, mas fora esse caso, a linearidade é apenas uma aproximação da resposta real. O processo estocástico exato para $u(t)$ é definido a partir das condições iniciais aleatórias $p_i(0)$ e $q_i(0)$ descritas pela distribuição na Eq. (6.26) e pela relação (6.27). Desta forma temos todas as probabilidades a N -pontos escritas como

$$P(u_t) = \int dq_0 \dots dq_n dp_0 \dots dp_n \delta \left(u_t - \frac{f_0(q, v; t)}{m} \right) F(q, p), \quad (6.28)$$

$$P(u_t, u_{t'}) = \int dq_0 \dots dq_n dp_0 \dots dp_n \delta \left(u_t - \frac{f_0(q, v; t)}{m} \right) \delta \left(u_{t'} - \frac{f_0(q, v; t)}{m} \right) F(q, p), \quad (6.29)$$

e assim por diante.

No sistema de osciladores harmônicos, $f_0(\dots)$ é uma função linear e $F(q, p)$ é uma distribuição Gaussiana. Assim, todas as probabilidades a N -pontos também são Gaussianas. Quando o tamanho da malha temporal t_1, \dots, t_N se igualar ao número de partículas do banho, obteremos funções do tipo delta de Dirac, já que existe um número maior de vínculos no espaço de fase que graus

de liberdade.

É simples mostrar que as equações de movimento para as coordenadas q_i que surgem da Hamiltoniana (6.25) consistem em

$$\frac{d^2}{dt^2}q_i(t) = -\sum_{j=0}^N \frac{K_{ij}}{m_i}q_j(t) + \frac{1}{m_i}g_iK(t). \quad (6.30)$$

Esse sistema de equações pode ser resolvido pelo método da diagonalização de $M_{ij} = K_{ij}/m_i$. Definimos as variáveis $y_i = \sum_{j=0}^n U_{ij}q_j$, de tal maneira que a matriz de componentes U_{ij} diagonaliza M , de forma que $(U^{-1}MU)_{ij} = \lambda_i^2\delta_{ij}$. As equações de movimento para y_i portanto se desacoplam, e a solução geral é dada abaixo

$$y_i(t) = A_i \cos(\lambda_i t) + B_i \sin(\lambda_i t) + \frac{1}{\lambda_i} \int_0^t dt' \sin[\lambda_i(t-t')] \alpha_i(t'), \quad (6.31)$$

onde definimos $\alpha_i(t) = g_i K(t)/m_i$. Podemos relacionar as constantes A_i e B_i com as condições iniciais $q_i(0)$ e $p_i(0)$. Para isto, avaliamos $y_i(0)$ e $\dot{y}_i(0)$, que resultam em

$$A_i = \sum_{j=0}^N U_{ij}q_j(0); \quad B_i = \frac{1}{\lambda_i} \sum_{j=0}^N U_{ij}p_j(0)/m_j. \quad (6.32)$$

Finalmente, escrevemos a solução geral para a velocidade das partículas $p_i(t)/m_i$,

$$\begin{aligned} \frac{p_i(t)}{m_i} &= \frac{1}{m_i} \sum_{j=0}^N \sum_{k=0}^N U_{ij}^{-1} U_{jk} [v_k(0) \cos(\lambda_j t) + q_k(0) \lambda_j \sin(\lambda_j t)] \\ &\quad + \frac{1}{m_i} \sum_{j=0}^N \sum_{k=0}^N U_{ij}^{-1} U_{jk} \int_0^t dt' \cos[\lambda_j(t-t')] g_k K(t). \end{aligned} \quad (6.33)$$

Para definir uma teoria para a resposta linear de $u(t)$ ao campo $K(t)$, basta calcularmos a média de $\langle u(t) \rangle$, que fornece a admitância e a covariância. No estado de equilíbrio, as médias $\langle v_i(0) \rangle$ e $\langle q_i(0) \rangle$ se anulam. Usando esse fato

em (6.33), obtemos

$$\langle u(t) \rangle = \int_0^t dt' \left\{ \frac{1}{m_0} \sum_{j=0}^N \sum_{k=0}^N U_{0j}^{-1} U_{jk} \cos[\lambda_j(t-t')] g_k \right\} K(t'), \quad (6.34)$$

que caracteriza um sistema de resposta linear para $\langle u(t) \rangle$. Identificamos a admitância prontamente como o termo em colchetes. É conveniente trabalhar com a Eq. (6.34) no espaço de Fourier. Antes disto reescrevemos as integrais que aparecem na expressão anterior como

$$\int_0^t dt' \cos[\lambda_j(t-t')] K(t') = \int_{-\infty}^{\infty} dt' \Theta(t-t') \cos[\lambda_j(t-t')] K(t'). \quad (6.35)$$

O domínio de integração pode ser estendido para $t' = -\infty$ porque $K(t')$ se anula para $t' < 0$ e a função de Heaviside permite fazer o mesmo para $t' = \infty$. As transformadas de Fourier de cada termo $\Theta(t) \cos(\lambda_j t)$ são dadas por

$$\frac{1}{2} [\tilde{\Theta}(\omega - \lambda_j) + \tilde{\Theta}(\omega + \lambda_j)] = \frac{\pi}{2} \left[\delta(\omega - \lambda_j) + \delta(\omega + \lambda_j) + \frac{i}{\pi} \left(\frac{1}{\omega - \lambda_j} + \frac{1}{\omega + \lambda_j} \right) \right]. \quad (6.36)$$

Substituindo esse resultado na Eq. (6.25), após multiplicar ambos lados por $\int dt e^{i\omega t} (\cdot)$, obtemos

$$\langle \tilde{u}(\omega) \rangle = \tilde{Y}(\omega) \tilde{K}(\omega). \quad (6.37)$$

onde a admitância é dada por

$$\begin{aligned} \tilde{Y}(\omega) = \frac{1}{m_0} \sum_{j=0}^N \sum_{k=0}^N U_{0j}^{-1} U_{jk} \frac{\pi}{2} [\delta(\lambda_j - \omega) + \delta(\lambda_j + \omega)] g_k \\ + \frac{i}{m_0} \sum_{j=0}^N \sum_{k=0}^N U_{0j}^{-1} U_{jk} \frac{\omega}{\omega^2 - \lambda_j^2} g_k. \end{aligned} \quad (6.38)$$

O próximo passo é obter a covariância do sistema e a sua densidade es-

pectral. Para isto, multiplicamos a Eq. (6.33) avaliada em $i = 0$ por $u(0) = p_0(0)/m_0$. Tirando a média, o resultado é

$$\Sigma(t - t') = \frac{k_B T}{m_0^2} \sum_{j=0}^N U_{0j}^{-1} U_{j0} \cos[\lambda_j(t - t')]. \quad (6.39)$$

A densidade espectral se escreve como

$$\tilde{\Sigma}(\omega) = \frac{k_B T}{m_0^2} \sum_{j=0}^N U_{0j}^{-1} U_{j0} \pi [\delta(\lambda_j - \omega) + \delta(\lambda_j + \omega)], \quad (6.40)$$

onde vale notar que a covariância não é afetada pelo campo externo $K(t)$, o que é uma suposição básica da teoria de resposta linear.

A relação entre flutuação e dissipação é obtida utilizando as fórmulas (6.12) e (??) para identificar $\tilde{g}(\omega)$ e $\tilde{\Gamma}(\omega)$. Desse modo, é possível escrever

$$\tilde{\Gamma}(\omega) = \frac{\tilde{\Sigma}(\omega)}{\Re[\tilde{Y}(\omega)]} \Re[\tilde{g}(\omega)], \quad (6.41)$$

que reproduziria a Eq. (??) se a densidade espectral for proporcional à $\Re[\tilde{Y}(\omega)]$. Comparando (6.38) com (6.40), vemos que isto acontece somente se as partículas do fluido não sofrerem influência do campo externo $K(t)$ que afeta a partícula Browniana, exigindo, portanto, que $g_i = 0$ (exceto $g_0 = 1$). Comparando com a Eq. (6.41), recuperamos a relação de flutuação e dissipação de Kubo no espaço de Fourier

$$\tilde{\Gamma}(\omega) = m_0 k_B T 2\Re[\tilde{g}(\omega)]. \quad (6.42)$$

O resultado, após a aplicar a transformada inversa, se escreve como

$$\langle \varphi(t)\varphi(t') \rangle = m_0 k_B T g(|t - t'|), \quad (6.43)$$

que é exatamente a relação de Kubo (6.6).

Enfatizamos novamente que a relação Nyquist generalizada só é satisfeita se $g_{i \neq 0} = 0$. Em um sistema de osciladores idênticos, onde utilizaríamos o processo estocástico $u(t)$ para simplificar o tratamento matemático, esta suposição soa artificial. Em cenário mais realístico, é possível que todas partículas respondam da mesma maneira e assim $g_i = 1$ (afinal, trata-se de partículas idênticas). Outra possibilidade é que g_i decaia na medida que se considera partículas mais afastada da fonte de campo.

A rigor, a física microscópica determina completamente a matriz de covariância e a admitância, além do espectro do ruído e a fricção pelas definições (6.12) e (6.13). Mesmo no sistema de osciladores harmônicos, na prática, não é possível calcular explicitamente cada um destes termos pois isto envolve determinar cerca de 10^{23} modos normais e a matriz de similaridade U_{ij} .

Existem circunstâncias onde estes núcleos exibem uma relação de proporcionalidade e isto pode ser explorado se existir um método teórico ou experimental para determinar qualquer um destes núcleos. A relação de flutuação e dissipação (6.42) permitiria então escrever $\Sigma(t-t')$, $Y(t-t')$, $g(t-t')$ e $\Gamma(t-t')$ a partir de apenas um núcleo conhecido. Essa relação é válida no sistema de osciladores somente se $g_i = 0$.

6.3 Difusão e mistura

Em sistemas que obedecem a relação de flutuação e dissipação (6.42), podemos escrever a matriz de covariância como função de qualquer um dos outros núcleos $Y(t-t')$, $g(t-t')$ ou $\Gamma(t-t')$. Destes, o núcleo de fricção é frequentemente o mais simples de trabalhar, de modo que usaremos a relação

$$\tilde{\Sigma}(\omega) = \frac{k_B T}{m} \frac{2\Re[\tilde{g}(\omega)]}{|\tilde{g}(\omega) - i\omega|^2}, \quad (6.44)$$

válida no contexto da equação de Langevin estacionária (6.5). Com isto, é possível comparar nossos resultados com o de Morgado, et. al., que exprimem o expoente de difusão em função da densidade espectral do ruído $\varphi(t)$.

Segundo o teorema de flutuação e dissipação, o núcleo de fricção determina a a densidade espectral do ruído, $\tilde{\Gamma}(\omega) = 2mk_B T \Re[\tilde{g}(\omega)]$. A difusão normal ocorre se a integral abaixo existir

$$\gamma = \int_0^\infty dt g(t) = \tilde{g}(0), \quad (6.45)$$

e nesse caso o coeficiente de difusão reproduziria a fórmula de Einstein

$$D = \frac{1}{2} \tilde{\Sigma}(0) = \frac{k_B T}{\gamma m}. \quad (6.46)$$

Comparando com a equação anterior, fica claro que sistemas em que γ diverge possuem comportamento subdifusivo e aqueles nos quais γ se anula apresentam superdifusão.

Antes de prosseguir, convém separar $\tilde{g}(\omega)$ nas suas partes reais e imaginárias,

$$\tilde{g}(\omega) = \tilde{g}_R(\omega) + i\tilde{g}_I(\omega). \quad (6.47)$$

Uma consequência imediata de $g(t)$ ser real é que $\tilde{g}_R(\omega)$ é uma função par enquanto $\tilde{g}_I(\omega)$ é uma função ímpar. Em especial, temos $g_I(0) = 0$. Supomos que a parte real se comporta como uma lei de potência próximo à origem — $\tilde{g}_R(\omega) \sim \omega^\mu$.

O subdifusivo é caracterizado por $\mu < 0$ e $\tilde{g}_R(0) = \infty$. Nesse caso, é fácil mostrar por substituição direta na Eq. (4.69) que

$$\tilde{\Sigma}(\omega) \propto \omega^{-\mu}; \quad \langle x(\tau) \rangle \propto \tau^{1+\mu}. \quad (6.48)$$

A análise do regime de superdifusão é um pouco mais complicada porque os

termos $\tilde{g}_R(\omega)$ e $\tilde{g}_I(\omega) - \omega$ se anulam quando $\omega \rightarrow 0$, fazendo com que $\tilde{\Sigma}(0)$ se transforme em uma singularidade do tipo $0/0$. É necessário avaliar a relação entre $\tilde{g}_I(\omega)$ e $\tilde{g}_R(\omega)$ para valores muito pequenos de ω .

Devido à exigência de causalidade para a resposta linear, a função $g(t)$ se anula para $t < 0$. Deste modo, é simples constatar que a parte real da transformada de Fourier de $g(t)$ é também igual à transformada de $\frac{1}{2}g(|t|)$ e a parte imaginária corresponde à transformada de $\frac{\text{sgn}(t)}{2}g(|t|)$. Toda a informação para reconstruir $g(t)$ está portanto contida separadamente em cada função $\tilde{g}_R(\omega)$ e $\tilde{g}_I(\omega)$. De fato, podemos escrever $\tilde{g}_I(\omega)$ como função de $\tilde{g}_R(\omega)$, e vice-versa, utilizando as relações de Kramers-Kronig [63]

$$\tilde{g}_I(\omega) = \frac{2\omega}{\pi} \int_0^\infty d\omega' \frac{\tilde{g}_R(\omega')}{\omega^2 - \omega'^2}; \quad \tilde{g}_R(\omega) = \frac{2}{\pi} \int_0^\infty d\omega' \frac{\omega' \tilde{g}_I(\omega')}{\omega'^2 - \omega^2}, \quad (6.49)$$

onde a divergência em $\omega = \omega'$ é contornada utilizando o valor principal de Cauchy ao avaliar as integrais.

As relações de Kramers-Kronig implicam que $|\tilde{g}_I(\omega)| \leq \tilde{g}_R(\omega)$, como será demonstrado a seguir. Utilizamos a prescrição $\omega' \mapsto \omega' + i\epsilon^2$ ao avaliar a integral (6.49). Ao aplicar a desigualdade de Schwarz na Eq. (6.49) se obtêm

$$|\tilde{g}_I(\omega)| \leq \frac{2\omega}{\pi} \int_0^\infty d\omega' \Re \left[\frac{\tilde{g}_R(\omega')}{(\omega + \omega') |\omega - \omega' - i\epsilon|} \right] \quad (6.50)$$

$$= \frac{2\omega}{\epsilon} \int_0^\infty d\omega' \frac{\tilde{g}_R(\omega')}{(\omega + \omega')} \sqrt{\epsilon^2 + (\omega - \omega')^2} \delta_\epsilon(\omega - \omega') \quad (6.51)$$

onde $\delta_\epsilon(\omega - \omega') = \frac{\epsilon}{\pi} [\epsilon^2 + (\omega - \omega')^2]^{-1}$. Observe que $\delta_\epsilon(\omega)$ é a distribuição de Cauchy (ou Lorentziana), que converge para a delta de Dirac quando $\epsilon \rightarrow 0$.

²As relações de Kramers-Kronig estão penduradas no fato que uma extensão complexa de $\tilde{g}(\omega)$ seria analítica no semiplano superior. Deste modo, o valor principal de Cauchy pode ser implementado substituindo $\omega' \mapsto \omega' + i\epsilon$. Isto corresponde à contornar os pólos em $\omega = \omega'$ por cima, que é a prescrição utilizada para obter a Eq. (6.49).

Avaliando a integral nesse limite, a Eq. (6.51) se reduz a

$$|\tilde{g}_I(\omega)| \leq \tilde{g}_R(\omega), \quad (6.52)$$

como queríamos demonstrar.

Substituindo $\tilde{g}_R(\omega) \sim \omega^\mu$ na Eq. (6.44), onde $\mu > 0$, resulta em $\tilde{\Sigma}(\omega) \sim \omega^{-\mu}$ para $\mu \leq 1$ e $\tilde{\Sigma}(\omega) \sim \omega^{2-\mu}$ para $\mu \geq 1$. Juntando o com o resultado na Eq. (4.69), ficamos com a caracterização geral do processo de difusão

$$\tilde{g}_R(\omega) \sim \omega^\mu, \quad \langle x(\tau)^2 \rangle \sim \begin{cases} \tau^{1+\mu} & -1 < \mu < 1 \\ \tau^{3-\mu} & 1 < \mu < 3 \\ \tau^0 & \text{caso contrário} \end{cases} . \quad (6.53)$$

A restrição que o expoente de difusão seja menor ou igual a 2 aparece naturalmente.

6.4 Conclusão

A teoria de resposta linear trata de variáveis termodinâmicas sujeitas a flutuação. Essa teoria se baseia em uma forte simplificação: um sistema de resposta linear é caracterizado apenas pela covariância de equilíbrio e pelo operador de admitância. A hipótese física por trás destas idéias é que a covariância de equilíbrio não se alteraria significativamente sob a ação de uma influência externa e a resposta a esta influência deveria ser linear.

Estas hipóteses são satisfeitas pelo sistema de osciladores clássicos, que portanto admite uma descrição exata a partir da teoria de resposta linear. No entanto, outro resultado importante desta teoria pode ser violado mesmo nesse sistema simples. A relação de Nyquist, que relaciona a admitância com a covariância de uma variável estocástica, não se aplica caso a influência externa

atue em mais de um oscilador. A principal conclusão que tiramos deste exemplo é que a teoria de resposta linear não requer a validade do teorema de flutuação e dissipação.

Essa observação muito provavelmente se estende a outros sistemas físicos. O teorema de Callen e Welton efetivamente lineariza a Hamiltoniana ao tratar o campo externo como uma perturbação de primeira ordem. Como o sistema de osciladores já é exato em primeira ordem de teoria de perturbação, acreditamos que os mesmos resultados possam ser extrapolados para outros sistemas. Se isto for verdadeiro, poderíamos compreender o teorema de flutuação e dissipação como a relação decorrente da “força termodinâmica” atuar em um único grau de liberdade do sistema físico em consideração. Qualquer outra situação implicaria na violação desta relação.

Essa discussão é especialmente importante em conexão com a literatura de equações de Langevin generalizadas. A equação de Kubo, embute o teorema de flutuação e dissipação: esse é o único caso óbvio onde a densidade espectral em (6.18) seria automaticamente válida. Não sabemos se existem outras possibilidades matematicamente consistentes. De todo modo, Kubo oferece uma demonstração da relação de flutuação e dissipação que, comparada com o teorema de Callen e Welton, é extremamente econômica[38]. No entanto, essa demonstração não é satisfatória. A equação de Langevin generalizada proposta por Kubo não permite escrever sistemas sistemas de resposta linear arbitrários. Possivelmente apenas os sistemas que obedecem à relação de flutuação e dissipação seriam permitidos. Assim, ao definir o processo estocástico como (6.4), Kubo já estaria implicitamente assumindo a validade do teorema que ele queria demonstrar. (Para outra crítica à demonstração de Kubo, veja a Ref. [15]).

Ao corrigir esse problema na equação de Langevin estacionária, vemos que o teorema de flutuação e dissipação é uma hipótese adicional. O mesmo pode

ser dito sobre a segunda solução da equação de Kubo na Sec. 6.1.2. Nesse caso, no entanto, é difícil comparar diretamente a covariância com a admitância pois trata-se de um sistema não estacionário. De qualquer modo, fica claro pela Eq. (6.24) que a admitância não determina a covariância ou vice-versa.

Em resumo, a relação de Nyquist generalizada, ou teorema de flutuação e dissipação, é útil em diversas circunstâncias. Não é possível inferir disto que todo sistema de resposta linear obedece a esta relação. Conjecturamos que, para todos sistemas Hamiltonianos, se a força externa atuar em vários graus de liberdade do sistema (e não somente naquele que estamos descrevendo) a relação de flutuação e dissipação seria violada. Isto é especialmente relevante na aplicação do formalismo da equação de Langevin generalizada a sistemas microscópicos. Raramente se possui um controle tão fino sobre as influências externas, sendo que o campo externo afetaria se não todas, mas ao menos várias das partículas do sistema. Parece improvável, portanto, que a relação de Nyquist seja satisfeita nesse domínio.

Capítulo 7

Detecção de tendências em séries temporais

Apresentamos uma metodologia de inferir tendências de séries temporais com comportamento errático. Ao decompor uma série como $x(t) = \bar{x}(t) + \delta x(t)$, onde $\bar{x}(t)$ é a tendência sistemática e $\delta x(t)$ são as flutuações, as observações sobre $x(t)$ trazem informação sobre o valor da função desconhecida $\bar{x}(t)$. Atribuímos por simplicidade uma probabilidade Gaussiana às funções desconhecidas $\bar{x}(t)$ e $\delta x(t)$. A partir destas probabilidades obtemos $P(\bar{x}|x)$, que descreve a incerteza sobre \bar{x} considerando uma observação da série $x(t)$. O resultado de $\langle \bar{x}(t) \rangle$ se assemelha ao do método de médias móveis. Adicionalmente, há uma estimativa de erros e do parâmetro de suavização.

Para situar esse problema na literatura, apresentamos o método de regressão por processos Gaussianos, que já é bastante utilizado na prática. Existe uma intersecção entre os dois métodos, mas a motivação é um pouco diferente. O que mais nos distancia desta literatura é o formalismo matemático, já que utilizamos métodos de teoria de campos para desenvolver os cálculos.

A família das Gaussianas apresenta várias propriedades interessantes e apa-

rece naturalmente em muitos modelos em física e estatística. Não deve parecer surpreendente que um processo estocástico onde todas as probabilidades de N pontos são Gaussianas seja especialmente simples de lidar. Estes processos são o tema do presente capítulo.

A Gaussianidade é uma suposição comum feita nos sistemas tratados nos capítulos 4 e 6. No entanto, a hipótese de Gaussianidade não foi necessária para obtenção dos resultados expostos até o momento, de sorte reservamos o tema de processos Gaussianos para esse capítulo. Ao invés de tratar de temas diretamente relacionados à física, aqui utilizamos o formalismo de processos estocásticos para inferir tendências em séries temporais.

Uma Gaussianidade multidimensional é definida por um vetor de média $\mathbf{m} = (m_1, \dots, m_N)$ e uma matriz de covariância Σ_{ij} , $i, j \in [1, N]$, com a restrição que Σ seja positivo-definida (todos autovalores positivos). Um fato interessante sobre esta distribuição é que a marginalização sobre uma componente y_k resulta em outra Gaussianidade virtualmente idêntica à primeira: basta desconsiderar as componentes de \mathbf{m} e de Σ que possuem o índice k . Em outras palavras, a marginalização não altera a média, a covariância nem a forma Gaussianidade da distribuição.

Demonstramos esse fato. A densidade de probabilidades de uma Gaussianidade multidimensional é escrita como

$$p(\mathbf{y}) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{m})^T \Sigma^{-1}(\mathbf{y}-\mathbf{m})}, \quad (7.1)$$

o que muitas vezes é representado de maneira compacta utilizando a notação

$\mathbf{y} \sim N(\mathbf{m}, \Sigma)$. Já a função característica da Gaussiana é dada por¹

$$\phi(\mathbf{z}) = \langle e^{i\mathbf{y} \cdot \mathbf{z}} \rangle = e^{-\frac{1}{2}\mathbf{z}^T \Sigma \mathbf{z} + i\mathbf{m} \cdot \mathbf{z}}, \quad (7.2)$$

e nos permite analisar mais facilmente o que acontece no processo de marginalização. A marginalização sobre uma variável y_k corresponde à tomar o limite sobre a variável correspondente $z_k \rightarrow 0$. Como mencionado anteriormente, isto simplesmente elimina os números que fazem referência ao índice k em (7.2) e não altera o resto da distribuição.

Vamos generalizar a noção de probabilidade Gaussiana de um vetor para uma função — que significa definir um processo estocástico. Sob certos aspectos, uma função $y(t)$ sobre um parâmetro t se assemelha a um vetor com um índice contínuo. No entanto, a generalização dos resultados anteriores não é imediata pois trata-se de um número infinito e incontável de componentes. Um pouco de cautela deve ser exercitada.

Iniciamos a discussão com um exemplo concreto. Considere que $y(t)$ representa a posição de uma partícula ao longo do tempo. Uma forma de reproduzir $y(t)$ seria medir os valores y_1, \dots, y_N em vários instantes t_1, \dots, t_N . É claro que, sem conhecer $y(t)$ *a priori*, o melhor que se pode fazer é fornecer a probabilidade conjunta de se obter uma determinada seqüência de resultados, ou $P(y_1, \dots, y_N)$. Conquanto que N seja finito, esta é uma distribuição multivariada comum, que não apresenta nenhuma dificuldade matemática abstrusa.

Um experimento mais cuidadoso poderia realizar medições adicionais nos instantes t_{N+1}, \dots, t_{N+M} , e assim obter um pouco mais de informação sobre

¹O resultado decorre de um pequeno exercício de álgebra linear: 1) existe uma matriz U que diagonaliza Σ^{-1} pois Σ e Σ^{-1} são positivo-definidas. 2) Com esta matriz fazemos uma transformação linear nas variáveis de integração $\mathbf{y} \mapsto \mathbf{y}' = U(\mathbf{y} - \mathbf{m})$, onde o Jacobiano é 1 devido à unitariedade de U . 3) Na forma diagonal, (7.2) equivale ao produto de várias Gaussianas unidimensionais, $Z(\lambda_i)^{-1} \int dy'_i e^{-\frac{1}{2}\lambda_i y'_i + i(U\mathbf{z})_i y'_i} = e^{-\frac{1}{2\lambda_i}(U\mathbf{z})_i^2}$, onde λ_i são os autovalores de Σ^{-1} e $Z(\lambda_i)$ é a constante de normalização. 4) O expoente equivale à forma linear $\frac{1}{2}\mathbf{z}U^\dagger D U \mathbf{z}$, onde $D = \text{diag}(\lambda_1^{-1}, \dots, \lambda_N^{-1})$, mas como U diagonaliza Σ^{-1} , temos $U^\dagger D U = \Sigma$. 5) Voltando com a média, temos finalmente (7.2).

$y(t)$. Os possíveis resultados deste experimento são codificados na probabilidade conjunta $P(y_1, \dots, y_N, \dots, y_{N+M})$. É claro que as duas atribuições devem ser consistentes entre si: se ignorarmos o resultado dos experimentos adicionais, voltamos à situação anterior, ou seja

$$\sum_{y_{N+1}} \cdots \sum_{y_{N+M}} P(y_1, \dots, y_N, \dots, y_{N+M}) = P(y_1, \dots, y_N). \quad (7.3)$$

Note que não é necessário que os tempos t_i sejam ordenados. A equação anterior é simplesmente a condição de compatibilidade de Chapman mencionada na Sec. 1.4.

Se $P(y_1, \dots, y_N, \dots, y_{N+M})$ for uma Gaussiana com média $\{m_i\}$ e covariância $\{\Sigma_{ij}\}$ a probabilidade marginal $P(y_1, \dots, y_N)$ será uma Gaussiana virtualmente idêntica, onde apenas se desconsidera as componentes adicionais y_{N+1}, \dots, y_{N+M} . Podemos levar esse raciocínio a um extremo: dizemos que a função $y(t)$ possui uma distribuição Gaussiana com média $\mu(t)$ e covariância $\Sigma(t, t')$, onde o índice contínuo t contempla todas as suas infinitas componentes. Em analogia à notação para vetores Gaussianos, escrevemos a probabilidade da função $y(t)$ como

$$y(t) \sim \mathcal{PG}[\mu(t), \Sigma(t, t')]. \quad (7.4)$$

Em aplicações práticas, no entanto, é necessário considerar apenas um número finito de valores. Seleccionando uma malha temporal t_1, \dots, t_N , pode-se argumentar que a probabilidade conjunta para y_1, \dots, y_N seria obtida *integrando sobre todas as outras infinitas componentes*. É claro que não é necessário, nem sequer possível, fazer todas estas integrações. De qualquer forma, podemos conceber esta probabilidade como uma Gaussiana com média $m_i = \mu(t_i)$ e covariância $\Sigma_{ij} = \Sigma(t_i, t_j)$.

Os processos Gaussianos são convenientes porque, ao fim das contas, se resumem à manipulação de Gaussianas multidimensionais. Existe uma certa

flexibilidade no que $\mu(t)$ e $\Sigma(t, t')$ podem ser. Em especial, a matriz de covariância define propriedades importantes das funções geradas por estes processos como continuidade, diferenciabilidade, o caráter Markoviano, etc. Além disto, alguns processos estudados em física são exemplos de processos Gaussianos — isto geralmente acontece em sistemas com Hamiltonianos quadráticos. Antes de proceder, convém demonstrar algumas propriedades de Gaussianas multidimensionais e de processos Gaussianos.

7.1 Propriedades de processos Gaussianos

Não há como compreender adequadamente os processos Gaussianos sem um domínio razoável das propriedades de Gaussianas multidimensionais. Com isto em mente, revisamos alguns dos principais resultados e mostramos alguns desdobramentos importantes no contexto de processos Gaussianos.

Lembramos que os resultados do capítulo anterior que discutem as propriedades da matriz de covariância também se aplicam aqui.

7.1.1 Probabilidades condicionais

Considere que o argumento de uma Gaussianas multidimensional seja separado em duas partes $\mathbf{y} = (\mathbf{u}, \mathbf{v})$, onde se conhece o valor das coordenadas \mathbf{v} , mas não das coordenadas \mathbf{u} . A probabilidade condicional é dada por

$$p(\mathbf{u}|\mathbf{v}) = \frac{p(\mathbf{u}, \mathbf{v})}{p(\mathbf{v})}, \quad (7.5)$$

e se $p(\mathbf{u}, \mathbf{v})$ for Gaussianas é fácil mostrar que tanto $p(\mathbf{v})$ e $p(\mathbf{u}|\mathbf{v})$ também o são.

Para calcular os parâmetros desta distribuição, é conveniente fatorar a média e a covariância originais em uma parte que depende de \mathbf{u} e outra que

depende de \mathbf{v} . Deste modo, a matriz de covariância fica escrita como

$$\Sigma = \begin{bmatrix} A & C \\ C^T & B \end{bmatrix} \quad (7.6)$$

e a média como $\mathbf{m} = (\mathbf{a}, \mathbf{b})$. Existe uma interpretação muito simples para esta decomposição: A é a matriz de covariância para as variáveis \mathbf{u} , B cumpre o mesmo papel para \mathbf{v} , enquanto C descreve a correlação entre as variáveis \mathbf{u} com \mathbf{v} .

A inversa da matriz de covariância, que aparece na densidade de probabilidades (7.1), é dada pelo complemento de Schur, resultando em

$$\Sigma^{-1} = \begin{bmatrix} (A - CB^{-1}C^T)^{-1} & -(A - CB^{-1}C^T)^{-1}CB^{-1} \\ -B^{-1}C^T(A - CB^{-1}C^T)^{-1} & B^{-1} + B^{-1}C^T(A - CB^{-1}C^T)^{-1}CB^{-1} \end{bmatrix}, \quad (7.7)$$

fato que pode ser testado por cálculo direto. Assim, a covariância condicional é $\bar{\Sigma} = A - CB^{-1}C^T$ (corresponde ao termo quadrático em \mathbf{u} que multiplica Σ_{11}^{-1}). A média fica dada por $\bar{\mathbf{m}} = \mathbf{a} + CB^{-1}(\mathbf{v} - \mathbf{b})$.

Definimos o processo Gaussiano condicional a partir dos resultados acima. Primeiramente, identificamos $\mathbf{u} = (y(t), y(t'))$ e $\mathbf{v} = (y_1, \dots, y_N)$, onde y_i é o valor da função $y(\cdot)$ no instante t_i . Supondo que originalmente se tinha $y(t) \sim \mathcal{PG}[\mu(t), \Sigma(t, t')]$, a média e a covariância do processo condicionado ficam dadas respectivamente por

$$\bar{\mu}(t) = \mu(t) + \sum_{ij=1}^N \Sigma(t, t_i) B_{ij}^{-1} (y_i - \mu(t_i)), \quad (7.8)$$

$$\bar{\Sigma}(t, t') = \Sigma(t, t') - \sum_{ij=1}^N \Sigma(t, t_i) B_{ij}^{-1} \Sigma(t_j, t'). \quad (7.9)$$

Aqui lembramos que $B_{ij} = \Sigma(t_i, t_j)$. Adotamos a notação abaixo para representar o processo condicional,

$$y(t)|\mathbf{y}_t \sim \mathcal{PG} [\bar{\mu}(t), \bar{\Sigma}(t, t')], \quad (7.10)$$

onde \mathbf{y}_t representa os valores da função $y(\cdot)$ conhecidos em cada instante t_1, \dots, t_N .

Um fato notável sobre o processo condicional é que a operação computacionalmente mais cara, a inversão de B , só precisa ser realizada uma vez para cada conjunto de observações. Em especial, se soubermos apenas uma componente do vetor (e.g., o valor da posição no instante inicial), B torna-se um número ordinário e nenhuma operação de inversão de matrizes precisa ser realizada. Esse tipo de situação aparece quando se impõe alguma condição inicial. Em um processo condicionado a um único ponto $y_0 = y(0)$, teríamos $y(t)|y_0 \sim \mathcal{GP} [\bar{\mu}(t), \bar{\Sigma}(t, t')]$, onde a média e a matriz de covariância se escreveriam como

$$\bar{\mu}(t) = \mu(t) + \Sigma(t, t_0)/\Sigma(t_0, t_0) (y(t_0) - \mu(t_0)), \quad (7.11)$$

$$\bar{\Sigma}(t, t') = \Sigma(t, t') - \Sigma(t, t_0)\Sigma(t', t_0)/\Sigma(t_0, t_0). \quad (7.12)$$

Observe que o processo $y(t)|y_0$ pode perder algumas propriedades presentes no processo original $y(t)$, como a simetria por translação temporal. Se $y(t)$ for um processo estacionário ($\Sigma(t, t') = \Sigma(t - t')$ e $\mu(t) = \text{cte}$), então $y(t)|y_0$ não o será já que a matriz de covariância não poderia ser escrita como $\Sigma(t - t')$. É possível definir um processo estacionário $y(t)$ para representar a "verdadeira" física do sistema, como sendo aquela invariante por translação temporal. No entanto, se alguma aplicação fizer referência às condições iniciais, utilizaríamos $y(t)|y_0$ e a simetria de translação temporal se quebraria espontaneamente.

Retomamos uma discussão do capítulo anterior. A prescrição de Kubo de inserir as condições iniciais na própria equação de Langevin não é muito sólida. Fica claro que podemos definir um processo estocástico no equilíbrio, onde a

física é simples, e incorporar uma informação adicional condicionando explicitamente a probabilidade de equilíbrio às condições iniciais. Vale ressaltar que “equilíbrio” não é uma propriedade da dinâmica microscópica. A suposição de equilíbrio depende essencialmente do grau de detalhamento com que se descreve e se controla um sistema físico.

7.1.2 Números aleatórios

Eventualmente é necessário gerar vetores aleatórios a partir de uma Gaussiana multidimensional. Um exemplo deste problema consiste em sortear pontos de uma trajetória $y(t)$ para traçar em um gráfico: escolhemos a malha t_1, t_2, \dots, t_N em que se deseja marcar $y(t_i)$ e sorteamos o vetor y_1, y_2, \dots, y_N de acordo com o processo Gaussiano escolhido.

Para começar, mostramos como gerar um único número aleatório $u \sim N(0, 1)$. Um procedimento simples é baseado no método geral da transformação de quantis. A função $F(x) = \frac{1}{2} (1 + \operatorname{erf}(x/\sqrt{2}))$ descreve a probabilidade cumulativa da Gaussiana. Se invertermos $F(x)$, temos a chamada distribuição de quantis, dada por $x = F^{-1}(q)$. Ela nos fornece o valor x tal que a probabilidade de obter um resultado menor ou igual à x é dada por q . Para gerar números aleatórios Gaussianos (ou de qualquer outra distribuição unidimensional), basta sortear q distribuído uniformemente no intervalo $[0, 1]$ e aplicar a transformação de quantis apropriada. Por simples que seja, esse procedimento raramente precisa ser implementado manualmente pois qualquer plataforma de cálculo científico razoável fornece algoritmos que sorteiam números gaussianos, geralmente de maneiras mais eficientes que a aqui apresentada.

Sabendo gerar números aleatórios com $N(0, 1)$, partimos para a tarefa mais difícil de gerar *vetores* segundo uma Gaussiana multidimensional. A estratégia é gerar uma seqüência de números Gaussianos *i.i.d* $\sim N(0, 1)$ e aplicar uma transformação linear que os leve à distribuição desejada. Isto é análogo

ao caso unidimensional, onde gera-se um número $N(0, 1)$ e o multiplica pelo desvio padrão, somando o resultado à media. A diferença é que no caso multi-dimensional será necessário multiplicar o vetor *i.i.d* por uma matriz que cumpre o papel da raiz quadrada da variância e somar o resultado a um vetor.

Consideramos sem perda de generalidade que a média da distribuição é nula e a matriz de covariância é Σ . Buscamos uma transformação linear que leve o vetor gerado por $N(0, I)$, onde I é a matriz identidade, para $N(0, \Sigma)$. De certa forma, isto equivale à tirar uma raiz quadrada de Σ , já que se $\mathbf{y} = L\mathbf{u}$, onde \mathbf{u} é o vetor *i.i.d.*, a covariância de \mathbf{y} será dada por LL^\dagger , que deve se igualar à Σ . A operação de "raiz quadrada" pode ser implementada pela decomposição de Cholesky já mencionada anteriormente.

7.2 Regressão e interpolação

7.2.1 Regressão supervisionada

Regressão é a tarefa de ajustar uma relação funcional entre duas variáveis a um conjunto de dados. Seja x a variável independente e $y = f(x)$ um modelo que descreve o comportamento da segunda variável. Muitas vezes não se conhece a relação funcional exata pois existem parâmetros livres em $f(x)$. Se, por exemplo, y depende linearmente de x , teríamos $f(x) = \theta_1 x + \theta_2$, onde θ_1 e θ_2 são parâmetros a serem determinados. Logicamente, $f(x)$ pode ser qualquer função com qualquer número de parâmetros ajustáveis.

Considere um conjunto de pontos experimentais $(x_1, y_1), \dots, (x_N, y_N)$. O modelo matemático associado ao método de regressão é

$$y_i = f(x_i; \theta) + e_i, \quad (7.13)$$

onde admitimos que a medição de y_i poder ser distorcida por um ruído aleatório

e_i . A suposição comum é que os erros e_i são variáveis Gaussianas idênticas e independentes. Deste modo, é possível escrever a probabilidade de obter uma certa seqüência de dados $\mathbf{y} = (y_1, \dots, y_N)$ fazendo as mensurações nos pontos $\mathbf{x} = (x_1, \dots, x_N)$. Denotamos $\mathbf{y}_{\mathbf{x}}$ como o conjunto completo de dados experimentais. Assim, a probabilidade de obter um certo vetor $\mathbf{y}_{\mathbf{x}}$ é

$$P(\mathbf{y}_{\mathbf{x}}|\theta) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^N [y_i - f(x_i; \theta)]^2} d^N \mathbf{y}. \quad (7.14)$$

A função $P(\mathbf{y}_{\mathbf{x}}|\theta)$ é conhecida como verossimilhança e define o modelo estatístico em questão.

A partir deste ponto, existe uma divergência de metodologia sobre como proceder. Um critério muito utilizado, o método da máxima verossimilhança, prescreve que o melhor valor de θ é o que maximiza $P(\mathbf{y}_{\mathbf{x}}|\theta)$. O resultado disto é o bem conhecido método dos mínimos quadrados: maximizar a Eq. (7.14) obviamente equivale à minimizar $\sum_{i=1}^N (y_i - f(x_i; \theta))^2$.

O método da máxima verossimilhança, no entanto, pode ser facilmente enganado por um modelo desnecessariamente complicado: uma parametrização de $f(x; \theta)$ o suficientemente flexível, eventualmente reproduzirá os dados nos mínimos detalhes, inclusive ajustando o erro, o que obviamente é indesejável. De um modo geral, o método da máxima verossimilhança somente fornece resultados adequados se o número de exemplos N for muito maior que o número de parâmetros: assim os resultados atípicos (associados a erros muito grandes) podem se cancelar e é muito improvável que um vetor grande $\mathbf{y}_{\mathbf{x}}$ contenha um número expressivo destes.

De fato, o método da máxima verossimilhança é justificado rigorosamente apenas no limite de uma série de dados infinita. Nesse contexto, admitimos que o modelo da Eq. (7.13) realmente descreve os pontos experimentais, mas

²Incidentalmente, esta é a razão porque usamos quadrados e não outra potência da diferença entre y_i e $f(x_i)$: quadrados estão associados a erros Gaussianos, que é uma suposição bastante robusta.

o valor de θ é desconhecido. A verossimilhança então ficaria cada vez mais concentrada em um pico em torno do valor correto de θ . No limite em que o número de dados torna-se infinito, a incerteza estatística desaparece e a verossimilhança se anula em todos os pontos fora uma vizinhança pequena em torno do valor correto de θ .

No modelo de regressão exposto anteriormente, é necessário prover a relação funcional $f(x; \theta)$. Isto caracteriza o que chamamos de método de inferência supervisionado: o algoritmo de inferência simplesmente se encarrega de ajustar os parâmetros livres de um modelo rígido. É claro que há várias situações onde isto é desejável: em bons modelos, os parâmetros possuem uma interpretação clara, e encontrar o valor destes muitas vezes é mais importante que oferecer um ajuste perfeito ao conjunto de dados ³.

No entanto, há várias situações onde os bons modelos simplesmente não existem. Para citar apenas dois exemplos evocativos, lembramos da bolsa de valores e o resultado das partidas de futebol. Não se conhece $f(x; \theta)$ nestes casos (e em muitos outros), mas talvez seja possível descrevê-la de uma forma genérica — um polinômio de alta ordem, os primeiros termos de uma série de Fourier, etc. É claro que sem conhecer a “verdadeira” relação que descreve os dados, é possível que a representação genérica seja inadequada (e.g., nem toda função possui uma representação em série de Taylor, além do que convergência pode ser muito lenta). Mais ainda, quanto maior o número de parâmetros livres, maior é a quantidade de dados necessária para obter uma inferência confiável. Deste modo, existe um limite prático para a complexidade do modelo. O tratamento Bayesiano consegue minimizar alguns destes problemas, como veremos adiante.

³Aqui lembramos a citação muitas vezes creditada a Dirac: “Se me derem um parâmetro, ajusto um elefante. Se me derem dois, faço ele mexer a tromba”.

7.2.2 Inferência Bayesiana

O método da máxima verossimilhança corresponde à prescrição *clássica* de inferência. Em seu favor, destaca-se a simplicidade matemática. Com o advento de computadores velozes, é possível sacrificar um pouco a eficiência numérica em prol de um método estatístico mais robusto. Os métodos Bayesianos geralmente se enquadram nesta categoria.

Faremos uma pequena digressão sobre como os métodos Bayesianos funcionam usando o modelo de regressão como exemplo. A mesma abordagem pode ser traduzida *ipsis literis* para modelo de regressão estocástico que será formulado a seguir. Para a maior conveniência do leitor que não esteja familiarizado com o formalismo de inferência Bayesiana, esta discussão é repetida de maneira mais detalhada no Apêndice. II, na página 203.

A inferência Bayesiana deriva da regra de Bayes,

$$P(\theta|\mathbf{y}_x) = P(\theta) \frac{P(\mathbf{y}_x|\theta)}{P(\mathbf{y}_x)} = \frac{P(\theta) P(\mathbf{y}_x|\theta)}{\sum_{\theta} P(\theta) P(\mathbf{y}_x|\theta)}, \quad (7.15)$$

que é uma consequência trivial do cálculo de probabilidades condicionais. Para obtê-la, basta reorganizar a identidade $P(A, B) = P(A)P(B|A) = P(B)P(A|B)$ substituindo A por θ e B por \mathbf{y}_x . A pergunta original, "*quais são os valores dos parâmetros?*", é substituída por uma versão mais modesta "*qual a probabilidade dos parâmetros possuírem um certo valor θ ?*".

Enfatizamos que a probabilidade atribuída à variável θ descreve a ignorância sobre o valor dos parâmetros e não resulta de uma aleatoriedade intrínseca no sentido de um jogo de dados. Não existe nenhum experimento aleatório para θ . De certa forma, a regra de Bayes ilustra como aprendemos com os exemplos: a probabilidade *a priori* $P(\theta)$ (para encurtar chamamos apenas de *a priori*) se "atualiza" para considerar as observações \mathbf{y}_x , e assim, após contemplar os dados experimentais, ficamos com um "grau de ignorância" revisado

descrito por $P(\theta|\mathbf{y}_x)$.

Existe uma dificuldade prática na aplicação do teorema de Bayes: é explicitar o *priori*. Pessoas racionais podem ter diferentes predisposições, e quem sabe somente a impessoalidade dos dados experimentais poderia forçá-las a concordarem entre si. Mas, o *priori* independe dos dados experimentais por construção. É fácil imaginar acadêmicos se digladiando por minúcias sobre o que representa o "verdadeiro" *priori*. Nestas circunstâncias, como proceder?

Não existe um verdadeiro *priori* e portanto não há solução incontroversa para esta questão. No entanto, uma série de métodos e critérios ao menos tentam restringir a forma do *priori* e, com sorte, chegam a uma solução unívoca. Em vários casos, a escolha não influencia muito os resultados finais no limite em que há muitos dados experimentais. Normalmente a função $P(\mathbf{y}_x|\theta)$ converge para um pico muito concentrado em torno de um máximo e esse pico acaba carregando a maior parte da massa de probabilidade de $P(\theta|\mathbf{y}_x)$. Esse comportamento recebe o nome de *normalidade assintótica*⁴ e é um pré-requisito básico sobre a maioria dos métodos não-Bayesianos de inferência. A normalidade assintótica não é necessária no contexto Bayesiano, apesar de simplificar bastante as coisas quando se deseja fazer um cálculo aproximado.

Em linguajar Bayesiano, ao invés de falar em normalidade assintótica, diremos que há modelos em que os dados “lavam” o *priori*. Isto nem sempre acontece: pode ser que o conjunto de dados seja pequeno e a informação codificada no *priori* ainda consiga se manifestar em $P(\theta|\mathbf{y}_x)$. Em alguns modelos complicados, a probabilidade $P(\theta|\mathbf{y}_x)$ é multimodal: a convergência se dá para mais de um pico. Além disto, o *priori* pode incorporar restrições fortes sobre θ , por exemplo, restringindo o domínio de possíveis valores dos parâmetros. Mesmo quando a informação experimental apontar para uma região proibida do espaço de parâmetros, a probabilidade $P(\theta|\mathbf{y}_x)$ continuará sendo nula nesta

⁴Na realidade, o conceito de normalidade assintótica é um pouco mais restritivo: assume-se que o pico formado por $P(\theta|\mathbf{d})$ é (assintoticamente) Gaussiano.

região, como o esperado.

Uma vez escolhido o *priori*, a tarefa de inferência está praticamente completa: toda informação sobre θ está codificada na probabilidade posterior para os parâmetros $P(\theta|\mathbf{y}_x)$. Normalmente, é conveniente escolher alguns valores representativos que forneçam uma noção mais imediata sobre o parâmetro desconhecido. Isto muitas vezes é feito calculando a média e o desvio padrão de θ sobre $P(\theta|\mathbf{y}_x)$ (a mediana ou a moda também são muito utilizadas). Assim se reporta um valor ótimo de $\bar{\theta} = \langle \theta \rangle$ junto ao erro associado $\sigma_{\theta}^2 = \langle (\theta - \bar{\theta})^2 \rangle$.

O tratamento Bayesiano do problema de regressão exposto anteriormente envolve duas etapas: (1) escolher um *priori* adequado e (2) estimar $\bar{\theta}$ e σ_{θ} a partir da probabilidade $P(\theta|\mathbf{y}_x)$. É claro que isto é muito mais trabalhoso que o método da máxima verossimilhança. Primeiro, ele envolve a escolha do *priori* que antes não era necessária. Além disto, a máxima verossimilhança é um método de maximização (que geralmente envolve apenas subir o gradiente de uma função) enquanto a estimativa Bayesiana envolve uma integral. Por outro lado, o tratamento Bayesiano é geralmente mais robusto a conjuntos pequenos de dados experimentais e à modelos excessivamente complicados. Em vários casos, o método da máxima verossimilhança pode ser entendido como a aproximação de Laplace do resultado Bayesiano, que consiste em substituir $P(\theta|\mathbf{y}_x)$ por uma Gaussiana centrada em torno do máximo.

Probabilidade preditiva

A probabilidade posterior $P(\theta|\mathbf{y}_x)$ possibilita vários tipos de inferência. Voltando ao algoritmo de regressão, podemos estimar uma faixa de valores adequados para os parâmetros da função $f(x; \theta)$. Melhor que isto, também é possível estimar o resultado de medições futuras para $y(x)$. Definimos $P(y_x|\mathbf{y}_x)$ como a probabilidade de obter o valor y no ponto x uma vez que se observou previamente o conjunto de dados \mathbf{y}_x . Em princípio, as medidas de y são

eventos independentes. No entanto, a correlação entre a nova medição y e os resultados anteriores \mathbf{y}_x é intermediada por $f(x; \theta)$: os dados refinam o conhecimento sobre a função que posteriormente é utilizada para melhor calcular $y \simeq f(x)$.

O procedimento é bastante simples. Lembramos que

$$P(y_x|\mathbf{y}_x) = \sum_{\theta} P(y_x \theta|\mathbf{y}_x), \quad (7.16)$$

$$= \sum_{\theta} P(\theta|\mathbf{y}_x) P(y_x|\theta \mathbf{y}_x). \quad (7.17)$$

O termo $P(\theta|\mathbf{y}_x)$ é a probabilidade posterior e o segundo termo é, na realidade, a verossimilhança para $y_x : \mathbf{y}_x$ influencia y_x apenas para ajustar os parâmetros θ , e uma vez que estes são conhecidos, \mathbf{y}_x se torna desnecessário fazendo com que $P(y_x|\theta \mathbf{y}_x) = P(y_x|\theta)$.

A função $P(y_x|\mathbf{y}_x)$ é conhecida como probabilidade preditiva. Usamos o que se aprendeu com o conjunto de exemplos em \mathbf{y}_x para extrapolar a relação $y = f(x)$ em pontos que não haviam sido observados anteriormente.

Fatores de Bayes

Uma parte importante da análise Bayesiana consiste no cálculo do chamado *fator de Bayes*, também conhecido como *evidência*, *fator de Occam*, entre outros nomes. O fator de Bayes corresponde ao termo $P(\mathbf{y}_x) = \sum_{\theta} P(\theta)P(\mathbf{y}_x|\theta)$ na Eq. (7.15). A interpretação deste número fica escondida pela notação: as probabilidades na regra de Bayes (7.15) dependem de uma certa escolha de parametrização de $f(x) = f(x; \theta)$ que chamamos de modelo M_{θ} . É lógico que outras parametrizações também são possíveis e, portanto, poderíamos ter sido mais explícitos e condicionar todas as probabilidades em (7.15) ao modelo

específico M_θ . A regra de Bayes se reescreveria como

$$P(\theta|\mathbf{y}_x, M_\theta) = P(\theta|M_\theta) \frac{P(\mathbf{y}_x|\theta, M_\theta)}{P(\mathbf{y}_x|M_\theta)}. \quad (7.18)$$

Aplicamos novamente a regra de Bayes para inverter a probabilidade $P(\mathbf{y}_x|M_\theta)$.

O resultado é

$$P(M_\theta|\mathbf{y}_x) = \frac{P(M_\theta)P(\mathbf{y}_x|M_\theta)}{\sum_M P(M_\theta)P(\mathbf{y}_x|M_\theta)}, \quad (7.19)$$

onde $P(M_\theta)$ é a probabilidade *a priori* do modelo M_θ ser o correto e a somatória no fator de normalização corre sobre todos modelos considerados. A Eq. (7.19) determina a probabilidade do modelo M_θ estar correto de acordo com um certo conjunto de dados. O fator de Bayes $P(\mathbf{y}_x|M_\theta)$ é proporcional à esta probabilidade.

O fato que é possível antever o fracasso ou sucesso de um modelo por um critério matemático simples é um tanto quanto surpreendente. No entanto, como é necessário atribuir um *priori* $P(M_\theta)$, esta análise é útil apenas para *comparar* dois ou mais modelos entre si. Ou seja, não é possível fazer juízos de valor absoluto sobre a validade de um determinado modelo (em estatística tradicional estes testes corresponderiam aos critérios de *goodness-to-fit*⁵), mas podemos selecionar entre dois ou mais modelos qual deles é o melhor.

Um exemplo típico para o uso deste formalismo seria comparar modelos com diferentes graus de complexidade. Por exemplo, M_1 , M_2 e M_3 poderiam denotar respectivamente uma parametrização linear, quadrática e cúbica. É claro que M_3 sempre oferecerá o melhor ajuste a qualquer conjunto de dados experimentais. No entanto, isto não significa que M_3 seja um modelo melhor e nem que $P(M_3|\mathbf{y}_x)$ será maior.

Mesmo se atribuirmos o mesmo *priori* a todos os modelos, o valor de $P(M_i|\mathbf{y}_x)$ depende de fatores diversos. Se o ajuste aos dados experimentais

⁵Existem alguns critérios Bayesianos do tipo *goodness-to-fit*. Em geral, eles envolvem a comparação com um modelo abstrato que representaria o modelo correto.

for bom, a verossimilhança $P(\mathbf{y}_x|\theta_i M_i)$ terá um valor alto, contribuindo para aumentar o fator de Bayes que é a média desta função segundo $P(\theta)$. Outro ponto importante é a complexidade do modelo. Geralmente a verossimilhança só atinge valores muito altos em uma região relativamente pequena do espaço de parâmetros e as outras regiões pouco contribuem. A razão entre o volume do espaço de parâmetros onde a verossimilhança é apreciável e o volume onde o priori é apreciável normalmente decresce com a dimensionalidade do espaço de parâmetros⁶. Se a melhora no ajuste aos dados for apenas marginal, uma maior complexidade (modelos com mais parâmetros) é naturalmente penalizada no cálculo dos fatores de Bayes.

7.2.3 Aprendizado não supervisionado

Agora voltamos para a linha raciocínio principal do capítulo e discutimos o modelo de regressão não supervisionado. A principal característica que o diferencia do modelo discutido na Sec. 7.2.1 é que não é necessário fornecer uma expressão fechada para $f(x; \theta)$. O objetivo da regressão não-supervisionada é reconstruir a função $f(x)$ a partir dos dados experimentais. Não se trata, portanto, de um método de ajuste de parâmetros.

Via de regra, os dois modelos de regressão se destacam em situações distintas. Na regressão supervisionada, a ênfase é o modelo. Os parâmetros livres normalmente possuem uma interpretação física e queremos determinar o valor de alguns destes parâmetros. Já para a regressão não-supervisionada, a ênfase está em fazer boas previsões. A idéia é que não importa o modelo que está por trás dos dados, desde que suas previsões se confirmem. O método

⁶Para entender melhor esse argumento considere a seguinte situação. O modelo M_1 possui apenas um parâmetro θ_1 que pode variar na faixa $[0, 1]$ e o modelo mais complexo M_2 possui um parâmetro adicional θ_2 que varia no mesmo intervalo. Digamos que, em ambos casos, a verossimilhança só é apreciável para tanto θ_1 quanto θ_2 no intervalo $[0.25, 0.75]$. Para o modelo M_1 , esta faixa representa metade do volume do espaço de parâmetros, mas, devido à dimensão adicional, esta razão cai para um quarto em M_2 . O modelo M_2 é então penalizado pela complexidade adicional e precisa compensar com um valor numérico pelo menos duas vezes maior em sua verossimilhança.

não-supervisionado é indicado para as situações onde o mecanismo que gera os dados é essencialmente desconhecido, mas mesmo assim é necessário extrapolar um certo conjunto de dados para situações ainda não observadas.

Nos casos mais simples de modelos não-supervisionados, os quais iremos formular aqui, a função $f(x)$ recebe um argumento real e retorna outro número real. Um exemplo seria a tentativa de extrapolar alguma tendência com base nos dados de uma série temporal: pode ser o preço de um ativo no mercado de ações ou a concentração de CO_2 na atmosfera; os detalhes não importam muito. Existem, ainda, problemas mais complicados onde x e $f(x)$ podem representar vetores de alta dimensionalidade ou mesmo algum objeto mais exótico como uma palavra, um gráfico, ou uma imagem [57]. Em resumo, as possibilidades de complicar o esquema geral que apresentaremos são muito amplas.

Começamos montando o problema da mesma maneira que no caso anterior

$$y_i = f(x_i) + e_i,$$

mas sem assumir uma forma paramétrica para $f(x)$. A função de verossimilhança $P(\mathbf{y}_x|\theta)$ é a mesma que (7.14), mas agora quem cumpre o papel dos parâmetros a serem ajustados é a própria função $f(x)$. Os parâmetros livres representam a informação necessária para reconstruir a relação funcional $f(x)$, mas como não assumimos qualquer forma *a priori*, isto consiste numa quantidade infinita de informação: o valor da função em x_0 , em $x_0 + \delta x$, $x_0 + 2\delta x$, e assim por diante.

Escrevemos a regra de Bayes para esse modelo

$$P[f|\mathbf{d}] = P[f] \frac{P(\mathbf{y}_x|f)}{P(\mathbf{y}_x)} = \frac{P[f] P(\mathbf{y}_x|f)}{\sum_f P[f] P(\mathbf{y}_x|f)}, \quad (7.20)$$

onde distinguimos as probabilidades usuais atribuídas a números de probabi-

lidades atribuídas a funções utilizando colchetes nas segundas. Nesse ponto é necessário fornecer a probabilidade *a priori* para as funções $f(x)$, que aqui supomos ser um processo Gaussiano. O motivo é pragmático: processos Gaussianos são bastante flexíveis e, ao mesmo tempo, consistem em uma das poucas situações tratáveis sem muitas complicações matemáticas.

Deste modo, seja $f(x) \sim \mathcal{PG}[\mu(x), \Sigma(x, x')]$. Supomos que esse processo descreve o *priori* $P[f]$. A verossimilhança dada pela Eq. (7.14) afeta o processo Gaussiano alterando a matriz de covariância e a média. Podemos determinar o processo Gaussiano posterior calculando a média e a covariância de $f(\cdot)$ em dois pontos arbitrários x e x' . O resultado é um novo processo Gaussiano que denominamos

$$f(x)|\mathbf{y}_{\mathbf{x}} \sim \mathcal{PG}[\bar{\mu}(x), \bar{\Sigma}(x, x')], \quad (7.21)$$

para o qual

$$\bar{\mu}(x) = \mu(x) + \mathbf{s}(x)^T (K + I\sigma^2)^{-1} \mathbf{y} \quad (7.22)$$

$$\bar{\Sigma}(x, x') = \Sigma(x, x') - \mathbf{s}(x)^T (K + I\sigma^2)^{-1} \mathbf{s}(x') \quad (7.23)$$

onde $K_{ij} \equiv \Sigma(x_i, x_j)$, $s_i(x) \equiv \Sigma(x, x_i)$ e I é a matriz identidade. Note que estes resultados não são os mesmos que as Eqs. (7.8-7.9) porque levam em conta a existência do erro experimental. É claro que, quando $\sigma = 0$, tanto a média quanto a covariância em (7.8-7.9) são recuperadas.

O desvio padrão em um ponto experimental x_i é simplesmente

$$\sigma(x_i) = \sigma \sqrt{(I + \sigma^2 K^{-1})_{ii}^{-1}}. \quad (7.24)$$

No limite de pequeno erros experimentais σ , o desvio padrão se torna essencialmente igual ao próprio σ . Em todo caso, o comportamento típico é que o desvio padrão seja menor nos pontos experimentais e cresça na medida em que

nos afastamos deles. Tal comportamento é ilustrado pela na Fig. 7.1.

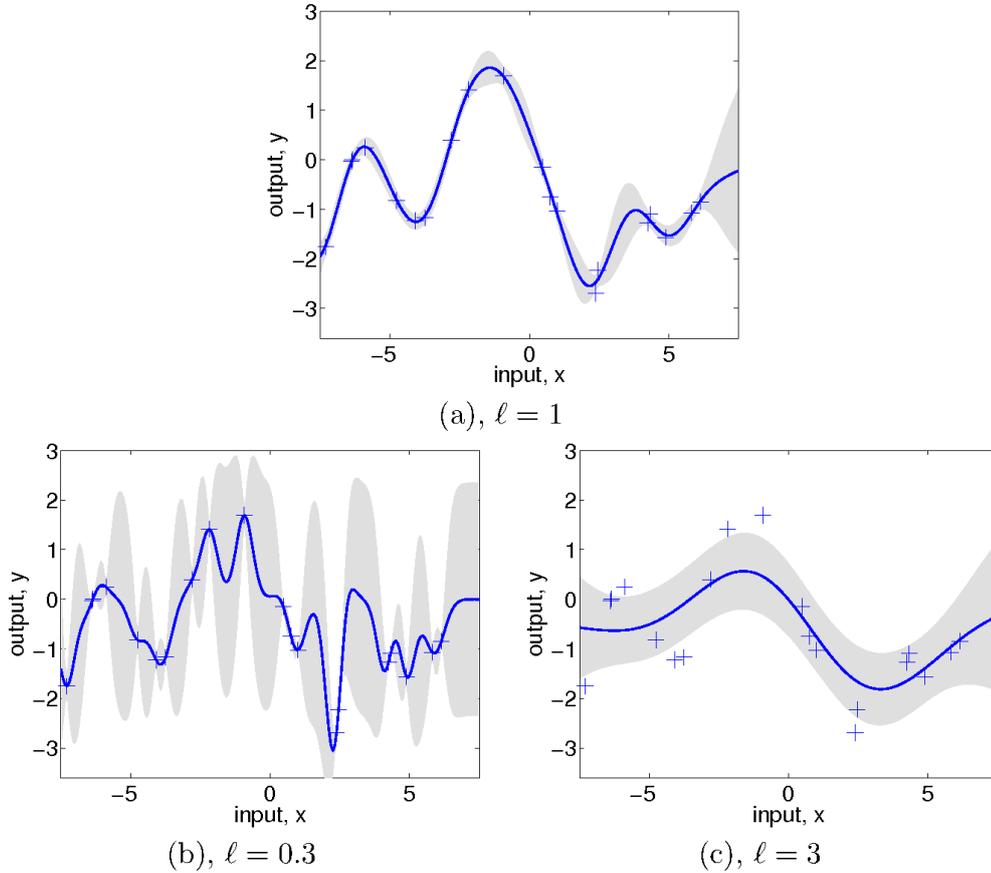


Figura 7.1: A figura mostra média e a região de 95% de confiança para o processo Gaussiano condicionado a um certo conjunto de pontos experimentais. Essa figuras foram retiradas de [57], para o processo $y(t) \sim \mathcal{PG} \left[0, \Sigma_0 e^{-|t-t'|^2/2\ell^2} \right]$: (a) descreve a escolha $(\ell; \Sigma_0; \sigma) = (1; 1; 0, 1)$, enquanto (b) e (c) representam respectivamente $(0, 3; 1, 08; 5 \times 10^{-4})$ e $(3; 1, 16; 0, 89)$.

Probabilidade preditiva

Em um processo Gaussiano, a probabilidade posterior se confunde com a probabilidade preditiva. De fato, o posterior fornece a probabilidade de obter uma função $f(x)$ quando se mediu um conjunto de dados \mathbf{y}_x . A probabilidade preditiva, no entanto, é algo um pouco diferente. Ela representa a chance de obter um novo conjunto de resultados y_j^* conhecendo os valores antigos $\{y_i\}$ e $\{x_i\}$. Deste modo, a única diferença com relação ao processo definido nas

Eqs. (7.22-7.23) é o fato que devemos considerar o erro experimental.

Em especial, se estivermos interessados na probabilidade marginal para um único valor de $y(x)$, o resultado é uma Gaussiana com média igual à Eq. (7.22) e variância

$$\sigma^*(x)^2 = \Sigma(x, x) - \mathbf{s}(x)^T (K + I\sigma^2)^{-1} \mathbf{s}(x) + \sigma^2. \quad (7.25)$$

O resultado se generaliza facilmente para qualquer probabilidade preditiva a N -pontos, onde a contribuição adicional σ^2 é adicionada apenas à diagonal da matriz de covariância.

Fator de Bayes

Calculamos agora o fator de Bayes para o processo Gaussiano. Lembramos que esse corresponde à normalização da probabilidade posterior que aparece no teorema de Bayes (Eq. [7.20]). Deste modo, temos

$$p(\mathbf{y}_x) = \int d\mathbf{f} p(\mathbf{f}) p(\mathbf{y}_x | \mathbf{f}), \quad (7.26)$$

onde a integral corre sobre os valores f_i da função avaliada nos pontos x_i ⁷. As densidades de probabilidade em (7.26) foram definidas da forma usual, $P(\mathbf{y}_x) = p(\mathbf{y}_x) dy_1 \dots dy_N$, $P(\mathbf{f}) = p(\mathbf{f}) df_1 \dots df_N$ e $P(\mathbf{y}_x | \mathbf{f}) = p(\mathbf{y}_x | \mathbf{f}) dy_1 \dots dy_N$.

Notação à parte, o cálculo de $p(\mathbf{y}_x)$ é bastante simples já que todas as integrais, assim como o resultado, são Gaussianas. De fato, é possível obter $p(\mathbf{y}_x)$ diretamente sem realizar nenhuma integral. Lembramos que $y_i = f_i + e_i$, onde e_i é o erro experimental independente para cada observação e distribuído

⁷É possível pensar informalmente que a integração deveria ser feita sobre a função $f(x)$ completa, e não apenas os pontos $f_i = f(x_i)$. Deste modo, escreveríamos $p(\mathbf{d}) = \int Df p[f] p(\mathbf{d} | \mathbf{f})$, onde o elemento de integração Df é uma medida sobre todo o espaço de funções. É conveniente avaliar esta integral integrando primeiro sobre todas as coordenadas de $f(x)$, exceto aquelas em que $x = x_i$. O resultado desta integração é justamente $p(\mathbf{f}) p(\mathbf{d} | \mathbf{f})$. A partir deste ponto, teríamos simplesmente que computar o resultado da Eq. (7.26).

como uma Gaussiana de média nula e desvio padrão σ . Deste modo, y_i é a soma de duas variáveis Gaussianas, e sua distribuição é

$$\mathbf{y}_x \sim N(\mathbf{m}, K + I\sigma^2), \quad (7.27)$$

onde usamos a regra que a média da soma é a soma das médias e a variância da soma é a soma das variâncias. (De um modo geral, cada cumulante da soma é a soma dos cumulantes.)

Reescrevemos o resultado anterior como

$$\ln p(\mathbf{y}_x) = -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{m})^T (I + \sigma^{-2}K)^{-1} (\mathbf{y} - \mathbf{m}) - \frac{1}{2} \ln |I + \sigma^{-2}K| - \frac{N}{2} \ln (2\pi\sigma^2), \quad (7.28)$$

onde $|\cdot|$ denota o determinante de uma matriz. Lembramos que o fator de Bayes descreve, fora um termo de proporcionalidade, a probabilidade de um modelo ser correto frente aos dados \mathbf{y}_x . Ele é útil para comparar diferentes modelos entre si, onde o modelo com o maior fator de Bayes é geralmente o preferido.

A Eq. (7.28) descreve um jogo de força entre dois fatores. O primeiro descreve o quão bem os dados \mathbf{y}_x se ajustam ao modelo: um modelo oferece um ajuste excelente se $\mathbf{y} \simeq \mathbf{m}$ ou se $\mathbf{y} - \mathbf{m}$ for uma composição linear de autovetores de $I + \sigma^{-2}K$ com autovalores muito grandes. Note que se todos autovalores de K forem tais que $\lambda_i \gg \sigma^2$, qualquer conjunto de dados se ajustaria bem ao modelo no sentido que o primeiro termo da Eq. (7.28) seria um número negativo pequeno. No entanto existe uma penalidade para isto. Lembramos que

$$\ln |I + \sigma^{-2}K| = \sum_{i=1}^N \ln (1 + \sigma^{-2}\lambda_i). \quad (7.29)$$

Deste modo, se todos os autovalores de K forem grandes em comparação com σ^2 o segundo termo implicará em uma penalidade grande sobre o fator de

Bayes. Dizemos que o modelo é excessivamente complexo no sentido que ajusta qualquer conjunto de dados, mas sem explicar bem nenhum deles. A situação ideal seria um modelo em que somente os autovetores de K que participam da expansão de $\mathbf{y} - \mathbf{m}$ possuam autovalores grandes enquanto os outros devem ser comparativamente pequenos.

Utilizando os fatores de Bayes, podemos confrontar modelos distintos para escolher aquele que melhor descreve os dados experimentais. Um uso comum desta análise consiste no ajuste de parâmetros livres na função de covariância. Considere, por exemplo, que o *priori* consiste no processo Gaussiano abaixo

$$f(x) \sim \mathcal{PG} \left[0, v_0 e^{-\frac{1}{2\ell^2}(x-x')^2} \right]. \quad (7.30)$$

A função de covariância Gaussiana implementa um processo estocástico onde as funções $f(x)$ são analíticas (possuem derivadas de todas as ordens). O resultado depende do valor numérico dos hiperparâmetros v_0 e ℓ . (Chamamos de hiperparâmetros qualquer parâmetro livre que apareça no *priori* para distinguir dos parâmetros da verossimilhança.) Se acreditarmos que todos os valores de v_0 e ℓ são igualmente prováveis, a análise dos fatores de Bayes prescreveria o valor que maximiza a Eq. (7.28). Caso se deseje atribuir um *priori* para v_0 e ℓ , a Eq. (7.28) seria somada ao termo $\ln P(v_0, \ell)$ antes de se calcular o máximo.

O resultado desta análise pode ser visto nas figuras abaixo.

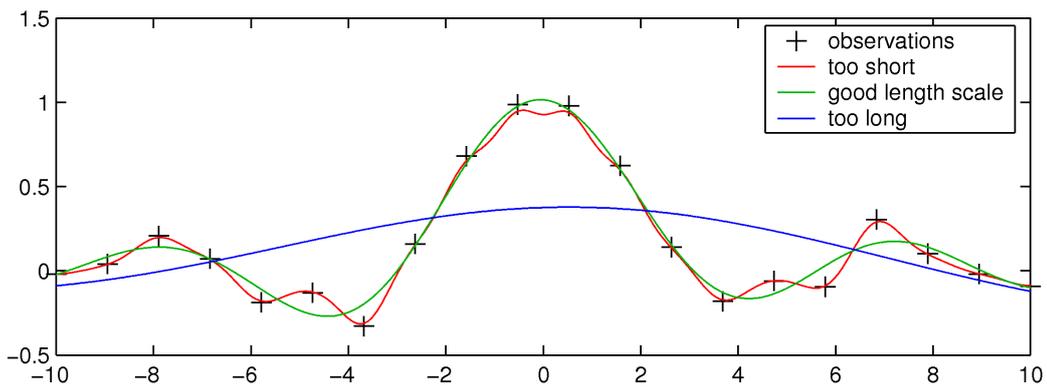


Figura 7.2: A figura foi tirada de uma apresentação de Rasmussen e descreve um processo idêntico ao considerado na Fig. 7.1. A linha vermelha oferece um ajuste quase perfeito aos dados experimentais, mas no entanto, a escolha que otimiza o fator de Bayes é a linha verde, que possui um ajuste um pouco pior, mas descreve uma curva intuitivamente mais simples. A curva azul sofre do efeito oposto ao da curva vermelha.

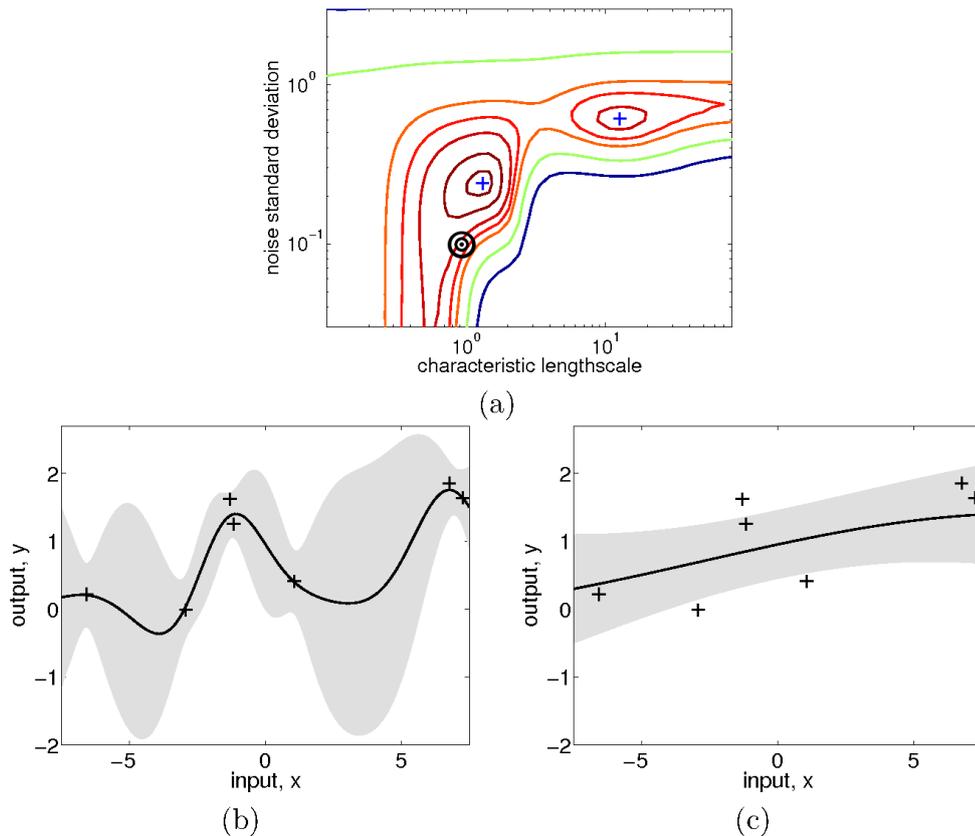


Figura 7.3: A figura foi tirada de [57] e descreve um processo idêntico ao considerado na Fig. 7.1. Os dados foram gerados por um processo com $(\ell; \Sigma_0; \sigma) = (1; 1; 0, 1)$ e a figura mostra a inferência para $P(\ell, \sigma)$, assumindo $\Sigma_0 = 1$ já conhecido. O valor correto está marcado no gráfico como um alvo. Note que existe uma concordância razoável se consideramos que se trata de apenas sete pontos experimentais. A figura à esquerda mostra a inferência da média e da região de 95% de confiança de $y(t)$ para o máximo global e a da direita mostra as mesmas características para o máximo local. Os dois máximos estão marcados com um “+” no gráfico.

7.3 Suavização de curvas

Desenvolvemos aqui uma aplicação original de processos Gaussianos para a detecção de tendências em séries temporais. Algumas idéias foram emprestadas do problema exposto na seção anterior, que foi incluída no presente trabalho justamente por esse motivo. No entanto, acrescentamos elementos novos que tornam o método mais adequado para o tratamento de séries temporais e de

outros problemas onde o ruído é intrínseco e não resulta necessariamente de um erro experimental.

Muitos fenômenos exibem um comportamento errático, ainda que se admita a existência de tendências escondidas. Um exemplo típico são as séries financeiras (bastante em voga nesta época de crise): os preços flutuam ao sabor dos mercados, deixando uns um pouco mais ricos e outros um pouco mais pobres, mas há tendências menos volúveis que operam em períodos longos: estas dependem da saúde geral da economia, da liquidez de uma empresa, da disposição dos *subprime* em pagarem seus empréstimos, etc. O comportamento errático destas séries temporais não se deve à presença de erros experimentais: o valor é obtido de forma exata, mas a tendência é distorcida por uma influência aleatória.

De fato, uma das maneiras mais simples de inferir as tendências escondidas é pelo método das médias móveis. A idéia é que as flutuações de curto prazo se cancelam na média e o resultado da suavização filtraria apenas as tendências de mais longo prazo. Ainda que o argumento para o uso de médias móveis pareça razoável, é difícil dizer se as flutuações de fato se cancelam numa série real. Seria possível justificar esse argumento de maneira rigorosa?

Matematicamente, o método das médias móveis separa uma série temporal $x(t)$ em duas partes $x(t) = \bar{x}(t) + \delta x(t)$: $\delta x(t)$ corresponde à flutuação a ser eliminada e $\bar{x}(t)$ a curva suavizada. A suavização de $x(t)$ realmente reconstrói o valor verdadeiro de $\bar{x}(t)$? A resposta obviamente depende da intensidade do ruído, pois sem ruído não teríamos trabalho algum e $\bar{x}(t) = x(t)$, mas um ruído forte certamente comprometeria a qualidade da inferência.

Conceitualmente, o trabalho é simples: basta definir uma distribuição de probabilidades para as tendências $\bar{x}(t)$ que considere tanto os dados mensurados $x(t)$ quanto um modelo para as flutuações — em símbolos, escrevemos simplesmente $P[\bar{x}|x]$. É claro que damos sentido à noção de "probabilidade de

uma função" associando um processo estocástico à $\bar{x}(t)$. Utilizaremos processos Gaussianos. Posteriormente atacamos o problema de amostragem contínua onde o processo Gaussiano essencialmente se torna uma teoria de campo livre.

7.3.1 Inferência com processos Gaussianos

A suavização de curvas está baseada no seguinte modelo

$$x(t) = \bar{x}(t) + \delta x(t), \quad (7.31)$$

onde $x(t)$ é a trajetória acessível à observação e $\bar{x}(t)$ e $\delta x(t)$ são funções desconhecidas que descrevem respectivamente as tendências e as flutuações.

Em diversas situações, é razoável supor que a derivada de $\delta x(t)$ seja aproximadamente um ruído branco. Isto corresponde a dizer que as correlações de $\delta x(t)$ decaem em uma escala de tempo muito mais rápida que $\bar{x}(t)$, o que é o comportamento esperado das flutuações frente às tendências. Nesse caso, a Eq. (7.31) pode ser reescrita na forma diferencial,

$$dx = a dt + b dW_t, \quad (7.32)$$

onde $a(t) = d\bar{x}/dt$ e dW_t é o ruído de Wiener (possui uma distribuição Gaussiana com média nula e variância \sqrt{dt}).

O modelo descrito pela Eq. (7.32) resulta em trajetórias $x(t)$ contínuas, mas não diferenciáveis. A inexistência da derivada temporal de $x(t)$ reflete na aparência errática das séries temporais consideradas. Caso a série considerada for suave o suficiente para se falar de sua derivada com relação ao tempo, seria necessário substituir $\delta x(t)$ por outro processo Gaussiano. Os argumentos que se seguem assumem o ruído branco, mas podem ser facilmente adaptados a outros casos.

Assumimos que o modelo para $\bar{x}(t)$ corresponde a um processo de média

nula,

$$\bar{x}(t) \sim \mathcal{PG}[0, \Sigma(t, t')]. \quad (7.33)$$

Já o processo Gaussiano para $\delta x(t)$, que decorre da suposição de ruído de Wiener em (7.32), é dado explicitamente por

$$\delta x(t) \sim \mathcal{PG}[0, b^2 \min(t, t')]. \quad (7.34)$$

Obtemos o processo condicionado $\bar{x}(t)|x(t)$ utilizando o teorema de Bayes,

$$P[\bar{x}|\mathbf{y}_t] = \frac{P[\bar{x}]P(\mathbf{y}_t|\bar{x})}{P(\mathbf{y}_t)}, \quad (7.35)$$

onde vale a observação que $P(\mathbf{y}_t|\bar{x})$ é uma realização sobre a malha t_1, \dots, t_N do processo estocástico $x(t)|\bar{x}(t) \sim \mathcal{PG}[\bar{x}(t), b^2 \min(t, t')]$.

A determinação do processo Gaussiano para $\bar{x}(t)|\mathbf{y}_t$ é um pouco mais complicada. A probabilidade $P(\mathbf{y}_t|\bar{x})$ depende somente do valor da função desconhecida $\bar{x}(t)$ nos pontos t_1, \dots, t_N . Deste modo, escrevemos $P(\mathbf{y}_t|\bar{x}) = p(\mathbf{y}_t|\bar{\mathbf{x}})d\mathbf{y}$, onde definimos o vetor $\bar{\mathbf{x}} = (\bar{x}_1, \dots, \bar{x}_N)$ com $\bar{x}_i = \bar{x}(t_i)$. Para calcular a média e a covariância do processo Gaussiano é necessário considerar duas variáveis \bar{x} e \bar{x}' escolhidas em instantes arbitrários t e t' além de todas as variáveis \bar{x}_i que aparecem em $p(\mathbf{y}_t|\bar{\mathbf{x}})$. As integrais envolvidas no cálculo da média e covariância de \bar{x} e \bar{x}' são da forma abaixo

$$\frac{\int d\bar{x} d\bar{x}' d\bar{\mathbf{x}} p(\bar{x}, \bar{x}', \bar{\mathbf{x}}) p(\mathbf{y}_t|\bar{\mathbf{x}}) \bar{x}^a \bar{x}'^b}{\int d\bar{x} d\bar{x}' d\bar{\mathbf{x}} p(\bar{x}, \bar{x}', \bar{\mathbf{x}}) p(\mathbf{y}_t|\bar{\mathbf{x}})}, \quad (7.36)$$

que pode ser simplificada como

$$\frac{\int d\bar{\mathbf{x}} p(\bar{\mathbf{x}}) p(\mathbf{y}_t|\bar{\mathbf{x}}) \int d\bar{x} d\bar{x}' p(\bar{x}, \bar{x}'|\bar{\mathbf{x}}) \bar{x}^a \bar{x}'^b}{\int d\bar{\mathbf{x}} p(\bar{\mathbf{x}}) p(\mathbf{y}_t|\bar{\mathbf{x}})} = \frac{\int d\bar{\mathbf{x}} p(\bar{\mathbf{x}}) p(\mathbf{y}_t|\bar{\mathbf{x}}) \langle \bar{x}^a \bar{x}'^b | \bar{\mathbf{x}} \rangle}{p(\mathbf{y}_t)}. \quad (7.37)$$

Para passar da Eq. (7.36) para o lado esquerdo da Eq. (7.37), usamos o fato

que $p(\bar{x}, \bar{x}', \bar{\mathbf{x}}) = p(\bar{\mathbf{x}})p(\bar{x}, \bar{x}'|\bar{\mathbf{x}})$. O termo $\langle \bar{x}^a \bar{x}'^b | \bar{\mathbf{x}} \rangle$ corresponde à média de $\bar{x}^a \bar{x}'^b$ sobre a distribuição condicional $p(\bar{x}, \bar{x}'|\bar{\mathbf{x}})$.

A integração no lado direito da Eq. (7.37) corresponde à tomar a média de $\langle \bar{x}^a \bar{x}'^b | \bar{\mathbf{x}} \rangle$ sobre a distribuição $q(\bar{\mathbf{x}}|\mathbf{y}_t) \equiv p(\bar{\mathbf{x}})p(\mathbf{y}_t|\bar{\mathbf{x}})/p(\mathbf{y}_t)$. Deste modo, é interessante explicitar esta distribuição para facilitar os cálculos posteriores. Começamos com o fator de normalização $p(\mathbf{y}_t) = \int d\bar{\mathbf{x}} p(\bar{\mathbf{x}})p(\mathbf{y}_t|\bar{\mathbf{x}})$. Uma vez que todas as probabilidades envolvidas são Gaussianas, o resultado também é uma Gaussiana para \mathbf{y}_t . Obtemos por cálculo direto que a média de $p(\mathbf{y}_t)$ é nula e a matriz de covariância é dada por

$$H = M^{-1} - M^{-1} (K^{-1} + M^{-1})^{-1} M^{-1}, \quad (7.38)$$

onde $M_{ij} = b^2 \min(t_i, t_j)$ e $K_{ij} = \Sigma_{\bar{x}}(t_i, t_j)$.

Já a distribuição $q(\bar{\mathbf{x}}|\mathbf{y}_t)$, é uma Gaussiana com covariância $(K^{-1} + M^{-1})^{-1}$ e média

$$\langle \bar{\mathbf{x}} \rangle = (K^{-1} + M^{-1})^{-1} M^{-1} \mathbf{y}. \quad (7.39)$$

Primeiramente, calculamos a média do processo Gaussiano $\bar{x}(t)|\mathbf{y}_t$. Antes de proceder, é conveniente definir o vetor $\mathbf{s}(t)$ com componentes

$$s_i(t) = \Sigma_{\bar{x}}(t, t_i). \quad (7.40)$$

Utilizando esta notação, o cálculo da média $\mu(t)$ de $\bar{x}(t)|\mathbf{y}_t$ se reduz a

$$\mu(t) = \int d\bar{\mathbf{x}} q(\bar{\mathbf{x}}|\mathbf{y}_t) \langle \bar{x} | \bar{\mathbf{x}} \rangle, \quad (7.41)$$

$$= \mathbf{s}(t) K^{-1} \int d\bar{\mathbf{x}} q(\bar{\mathbf{x}}|\mathbf{y}_t) \bar{\mathbf{x}}, \quad (7.42)$$

$$= \mathbf{s}(t) K^{-1} (I + MK^{-1})^{-1} \mathbf{y}, \quad (7.43)$$

onde I representa a matriz identidade. Por um procedimento semelhante, a

matriz de covariância se reduz a

$$\Sigma(t, t') = \Sigma_{\bar{x}}(t, t') - \mathbf{s}(t)K^{-1}\mathbf{s}(t') + \mathbf{s}(t)(K + KM^{-1}K)^{-1}\mathbf{s}(t'). \quad (7.44)$$

Observe que as Eqs. (7.8-7.9) são recuperadas se a variância da flutuação δx tender a zero. Nesse caso, teríamos $M \rightarrow 0$, fazendo com que $K + KM^{-1}K \sim KM^{-1}K$ e conseqüentemente o último termo ficaria $K^{-1}MK^{-1} \rightarrow 0$. Pela definição de operadores positivo-definidos, o segundo termo sempre fornece uma contribuição maior que zero, tornando a variância maior na presença de flutuações. A média, na ausência de flutuações, se reduziria à $\mathbf{s}(t)K^{-1}\mathbf{y}$, que é o resultado da Eq. (7.8).

7.3.2 Amostragem contínua: teoria de campos

A suavização de curvas foi previamente formulada como um processo Gaussiano para as tendências $\bar{x}(t)$. No entanto, existe uma limitação de ordem prática no cálculo das Eqs. (7.43-7.44). É necessário inverter matrizes de correlação cuja dimensionalidade depende do número de pontos existentes na série temporal. É comum encontrar séries de dados com milhares, milhões ou até mesmo bilhões de pontos. A partir de um certo ponto, o procedimento exposto anteriormente torna-se impraticável já que requer inverter uma matriz de altíssima dimensionalidade.

Os cálculos podem ser simplificados assumindo uma amostragem contínua. O conjunto finito de pontos amostrais é então substituído por uma função sobre o intervalo contínuo de mensuração $t \in [0, T]$. Em aplicações, esta função pode ser definida simplesmente ligando os pontos da série temporal por retas, ou outra prescrição semelhante. Na nova formulação, o que antes era um processo Gaussiano legítimo, torna-se agora uma teoria de campo livre. Ainda que a formulação matemática seja um pouco mais sofisticada que aquela

exposta na seção anterior, a computação envolvida não é tão intensa⁸ e em vários problemas, esse ganho de eficiência é fundamental.

Função de verossimilhança

A amostragem contínua introduz uma séria dificuldade matemática: é necessário definir a verossimilhança sobre um espaço \mathbf{y}_t com um número infinito e não-enumerável de variáveis. O que fazemos, efetivamente, é explorar o limite de uma frequência amostral infinita, ainda que na prática se utilize séries temporais com um número finito de amostras.

No que se segue, adotamos o modelo descrito na Eq. (7.32), já assumindo o ruído de Wiener. O cálculo da verossimilhança torna-se trivial quando observamos que para intervalos de tempo suficientemente pequenos, o ruído de Wiener prescreve uma Gaussiana com média $a(t) dt$ e variância $b^2 dt$. Deste modo, considere uma discretização que divide o intervalo de mensuração $t \in [0, T]$ em N sub-intervalos idênticos de tamanho $dt = T/N$. A verossimilhança $P[x|a]$ torna-se

$$P[x|a] \propto \exp \left[- \sum_{i=1}^N \frac{(x_i - x_{i-1} - a_{i-1} dt)^2}{2b^2 dt} \right], \quad (7.45)$$

onde $t_i = i dt$, $x_i \equiv x(t_i)$ e $a_i \equiv a(t_i)$. No limite $N \rightarrow \infty$, o expoente vira formalmente uma integral,

$$P[x|a] \propto e^{-S_V}; \quad S_V = \int_0^T dt \frac{1}{2b^2} (\dot{x} - a)^2. \quad (7.46)$$

Chamamos S_V de integral de ação para o processo devido à semelhança (ainda que superficial) com o mesmo objeto da física clássica.

A representação de probabilidades Gaussianas como integrais de ação é

⁸Adiantando um pouco a discussão, o passo computacionalmente mais demorado na formulação como processo Gaussiano é inverter a matriz de covariância sobre as observações. Uma vez que as matrizes de covariância são positivo-definidas, esta operação pode ser realizada pela decomposição de Cholesky, que se comporta com $O(N^3)$. A computação mais cara na formulação que apresentaremos é o cálculo dos coeficientes da série de Fourier da função $f(t)$. Isto é feito pelo algoritmo de transformadas rápidas de Fourier em $O(N \ln N)$.

bastante conveniente tanto do ponto de vista prático quanto para a economia de notação. É importante lembrar, no entanto, que a Eq. (7.46) não corresponde à definição de integral de Riemann. Aqui adotamos a convenção de Itô que avalia o integrando no ponto inicial de cada partição do domínio de integração. Nas integrais de Riemann, esta prescrição adicional não é necessária pois qualquer escolha convergiria para o mesmo valor no limite $N \rightarrow \infty$.

Outra observação importante é que a discretização na Eq. (7.45) é apenas um artifício matemático. Em uns poucos casos, é possível dizer que amostragem é, de fato, contínua. Assim, uma série temporal real normalmente consiste em um conjunto de pontos x_1, \dots, x_n avaliados em t_1, \dots, t_n . Obtemos uma aproximação contínua ligando estes pontos por retas (ou outros tipos de curva). No entanto, a discretização na Eq. (7.45) nada tem a ver com a discretização original da série, i.e., N e n são números independentes. Em especial, iremos sempre trabalhar no limite onde $N \rightarrow \infty$.

Priori

É conveniente, antes de obter $P[\bar{x}|x]$, calcularmos a probabilidade $P[a|x]$.

O priori $P[a]$ descreve o que se sabe sobre as tendências do coeficiente de drift $a(t)$ antes da observação de $x(t)$. Assumimos que $a(t)$ é contínuo e estacionário. A continuidade é esperada de um processo que descreva uma tendência suave, onde a continuidade de $a(t)$ garante a diferenciabilidade de $\bar{x}(t)$. A estacionariedade, por outro lado, diz que $a(t)$ e $a(t + \Delta t)$ são avaliados com a mesma probabilidade para qualquer deslocamento temporal Δt . Essa simetria é esperada na ausência de informação que selecione algum instante de tempo específico.

Uma das escolhas mais simples que satisfaz as propriedades anteriores é a

ação da partícula livre

$$S_0 = \frac{\tau^2}{2b^2} \int_{-\infty}^{\infty} dt (\partial_t a)^2, \quad (7.47)$$

onde τ é uma constante de tempo que torna S_0 adimensional. A Eq. (7.47), assim como outras alternativas envolvendo apenas as derivadas de $a(t)$ penalizam as variações de $a(t)$ em uma escala de tempo comparável com τ . Estas são invariantes por uma transformação do tipo $a(t) \mapsto a'(t) = a(t) + a_0$, que é uma simetria desejável em vários contextos (ela corresponderia ao *boost* Galileano em física clássica). Além disto, S_0 é unívoca se, junto às simetrias previamente mencionadas, o processo Gaussiano associado com $a(t)$ for Markoviano.

Do ponto de vista pragmático, a Eq. (7.47) fornece um modelo facilmente solúvel. É possível adicionar termos do tipo a^2 ou $(\partial_t^n a)^2$, mas estes podem resultar numa proliferação indesejável de parâmetros adicionais. Ao tratar o problema de interpolação de curvas com um formalismo semelhante ao nosso, Skilling [60] considera apenas um termo com a derivada segunda e obtêm uma resposta escrita como *splines* cúbicas.

Probabilidade posterior

De acordo com a regra de Bayes,

$$P[a|x] = P[a] \frac{P[x|a]}{P[x]}, \quad (7.48)$$

a integral de ação que define $P[a|x]$ é simplesmente a soma das ações do *priori* e da verossimilhança, $S = S_0 + S_D$. No entanto, esse resultado não pode ser utilizado diretamente: S_0 e S_V estão definidas sobre intervalos de integração diferentes e é necessário eliminar esta discrepância. Essa situação é esperada porque o intervalo de mensuração T não deveria aparecer no *priori*. Afinal, se nenhuma observação foi feita, não saberíamos o valor de T .

Note que esta é uma peculiaridade da ação da partícula livre, mas também vale para outros processos onde a ação S_0 é escrita como uma integral simples sobre a Lagrangiana. (Em contraposição, uma integral de ação do tipo $S = \int_{-\infty}^{\infty} dt dt' (\dots)$ não possuiria esta propriedade e se modificaria de uma maneira não trivial no procedimento de marginalização.)

Determinamos, agora, a média e covariância de $a(t)|x$. Para isto, escrevemos explicitamente a ação posterior,

$$S = \frac{1}{2b^2} \int_0^T dt [\tau^2 (\partial_t a)^2 + (\dot{x} - a)^2]. \quad (7.50)$$

$$= \frac{1}{2b^2} \left[\mathbf{a} \left(\frac{\tau^2}{dt} \gamma + dt I \right) \mathbf{a} + 2 dt \mathbf{a} \cdot \dot{\mathbf{x}} + dt \dot{\mathbf{x}} \cdot \dot{\mathbf{x}} \right], \quad (7.51)$$

onde, na segunda linha, temos a versão discretizada ($\dot{\mathbf{x}}$ possui componentes $\dot{x}_i = (x(t_i) - x(t_{i-1})) / dt$). Reorganizamos estes termos como o expoente de uma Gaussiana,

$$S = \frac{1}{2b^2} [(\mathbf{a} - \mathbf{m}) K^{-1} (\mathbf{a} - \mathbf{m})] + \text{cte}, \quad (7.52)$$

onde identificamos $K^{-1} \equiv \left(\frac{\tau^2}{dt} \gamma + dt I \right)$ e $\mathbf{m} = dt K \dot{\mathbf{x}}$. A média no ponto t_i é dada por $(\mathbf{m})_i$ e a covariância entre t_i e t_j é simplesmente $b^2 K_{ij}$.

Para encontrar K , substituímos $K^{-1} = \tau^2 \gamma / \delta t + \delta t I$ na relação $K^{-1} K / dt = I dt^{-1}$. Tomando o limite $dt \rightarrow 0$, a relação anterior vira uma equação diferencial,

$$(-\tau^2 \partial_t^2 + 1) K(t, t') = \delta(t - t'). \quad (7.53)$$

valida para todo par t, t' no intervalo de 0 à T , exceto nas bordas. Para determinar as condições de contorno, inspecionamos a primeira e a última linha de K^{-1} : comparando com (7.49) vemos que estas correspondem à $\partial_t K(t, t')|^{t=0} =$

$\partial_t K(t, t')|^{t=T} = 0$. A solução de (7.53) é portanto

$$K(t, t') = \omega \frac{\cosh(\omega t^-) \cosh(\omega(T - t^+))}{\sinh(\omega T)}, \quad (7.54)$$

onde $\omega \equiv \tau^{-1}$, $t^+ = \max(t, t')$ e $t^- = \min(t, t')$. A expressão anterior se reduz a um núcleo exponencial $\omega e^{-\omega|t-t'|}$ longe das fronteiras. O núcleo exponencial também é solução de (7.53), mas, sem levar em conta as condições de contorno corretas, ele apresenta um comportamento inadequado perto de $t = 0$ e $t = T$.

Probabilidade para $\bar{x}(t)$

Utilizamos o fato que $a(t) = d\bar{x}/dt$ para calcular $P[\bar{x}|x]$ a partir de $P[a|x]$ obtido na seção anterior. Qualquer combinação linear de variáveis Gaussianas também é Gaussianas. Assim, a média e a covariância de $P[\bar{x}|x]$ são calculadas aplicando-se a mesma transformação linear ($\bar{x}(t) = x_0 + \int_0^t dt a(t)$) nas quantidades correspondentes de $P[a|x]$. Isto resulta em

$$\langle \bar{x}(t) \rangle = x_0 + \int_0^t ds \langle a(s) \rangle \quad (7.55)$$

e

$$\text{Cov}[\bar{x}(t), \bar{x}(t')] = \int_0^t ds \int_0^{t'} ds' \text{Cov}[a(s), a(s')]. \quad (7.56)$$

Os resultados anteriores são mostrados na figura (7.4). Diferentemente do comportamento de $\bar{x}(t)$, o desvio padrão de $a(t)$ mostrado em (7.5) não cresce com o tempo, sendo aproximadamente igual à $b/\sqrt{2\tau}$ (for $T \gg \tau$). Ele cresce próximo às fronteiras já que uma quantidade menor de pontos é utilizada para estimar $a(t)$ (a estimativa de $\bar{x}(t)$ usa, efetivamente, os pontos em torno de um intervalo da ordem de τ em volta de t). É claro que as estimativas para $\bar{x}(t)$ e $a(t)$ não se casam perfeitamente com o valor verdadeiro. No entanto, a diferença raramente é maior que um desvio padrão.

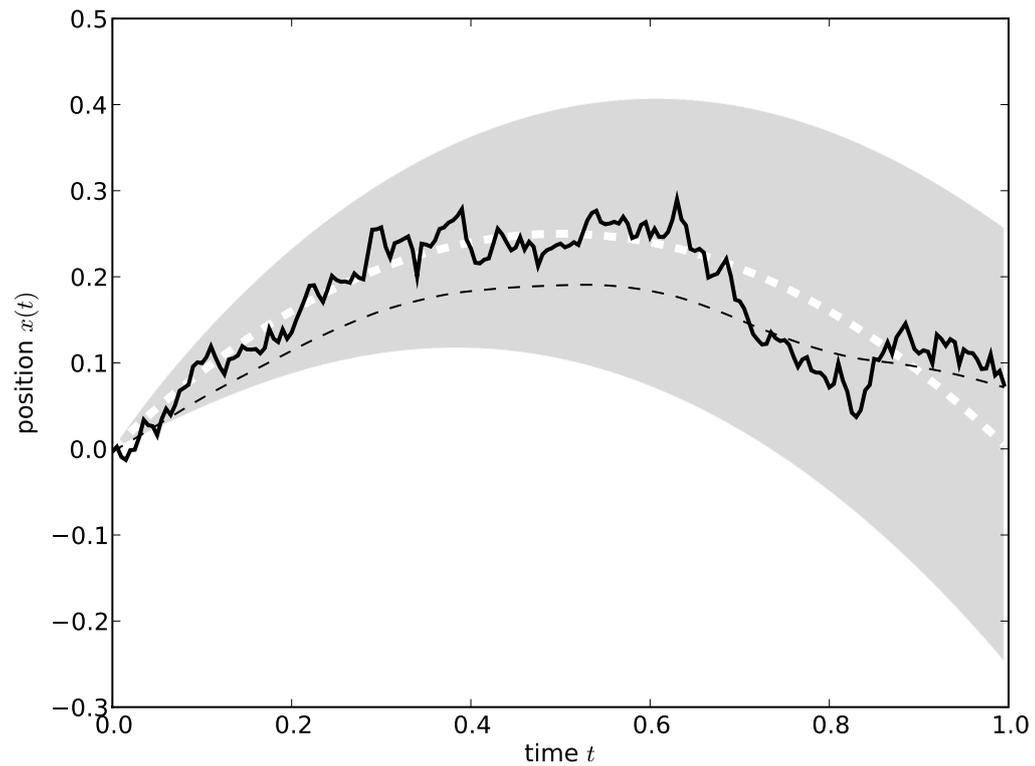


Figura 7.4: Posição $x(t)$ como função do tempo para dados sintéticos gerados a partir da Eq. (7.32). A linha negra tracejada representa a estimativa $\langle \bar{x}(t) \rangle$, enquanto a região cinza descreve o intervalo de um desvio padrão em torno do valor verdadeiro de $\bar{x}(t)$ (tracejado branco) que resultaria das flutuações serem completamente suprimidas.

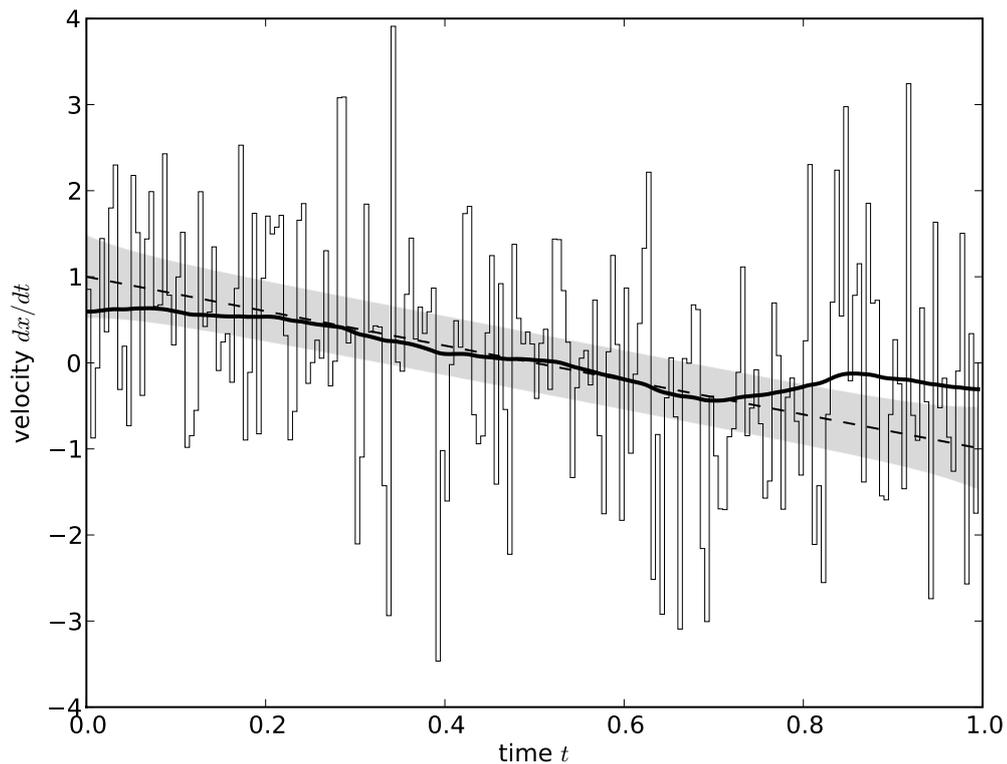


Figura 7.5: A estimativa da tendência $\langle a(t) \rangle$ a partir dos dados considerados na Fig. 7.4. A linha tracejada corresponde ao valor conhecido e a região cinza é o intervalo de um desvio padrão. Os dados na série temporal de $\dot{x}(t)$ foram reescalados pela metade para permitir uma melhor visualização.

Intensidade do ruído

Consideramos, até o momento, que a intensidade do ruído de Wiener, descrita pelo coeficiente b era conhecida de antemão. Normalmente isto não é verdade, mas existe a possibilidade de estimá-la utilizando a noção de variação quadrática da trajetória $x(t)$.

Escrevemos o ruído de Wiener dW_t como $\beta \sqrt{dt}$ onde $\beta \sim N(0, 1)$ para enfatizar o fato que ele é $O(dt^{1/2})$. Assim, ao avaliarmos dx^2 , resta um termo de $O(dt)$ dado por $b^2 \beta^2 dt$. Definimos a variação quadrática tomando a média

deste termo sobre uma trajetória muito curta,

$$\frac{d[x(t)]}{dt} \equiv \lim_{\delta t \rightarrow 0} \frac{1}{2 \delta t} \int_{t-\delta t}^{t+\delta t} dx^2 \quad (7.57)$$

$$= \lim_{\delta t \rightarrow 0} \frac{b^2}{2 \delta t} \int_{t-\delta t}^{t+\delta t} dt \beta^2 = b^2. \quad (7.58)$$

Para o processo de Wiener, b^2 é simplesmente a derivada da variação quadrática.

Em termos práticos, aproximamos a variação quadrática de uma série de alta frequência como $d[x(t_i)] = (x_i - x_{i-1})^2$. Para uma amostragem finita, $d[x(t)]$ calculado desta maneira não será exatamente igual à variação quadrática e portanto flutuará em torno de b^2 . Uma maneira simples de eliminar estas flutuações seria calcular a média temporal $\overline{d[x]/dt}$ sobre todo o conjunto de dados, fornecendo $b^2 \simeq \sum_{i=1}^n (x_i - x_{i-1})^2 / (t_i - t_{i-1})$. Usamos esta aproximação para produzir as figuras 7.4 e (7.5).

Os resultados obtidos anteriormente se baseiam na premissa que b é constante. Caso b varie no tempo, é possível redefinir a escala temporal de tal maneira que o coeficiente de difusão se torne constante. Essa nova escala de tempo é conhecida como o tempo estocástico e é definida por $ds \equiv b^2 dt$. O tempo estocástico restaura a constância do coeficiente de difusão, já que resulta na equação estocástica

$$dx = a dt + b dW_t, \quad (7.59)$$

$$= \frac{a}{b^2} ds + dW_s, \quad (7.60)$$

onde dW_s é o ruído de Wiener na nova escala de tempo $s(t) = \int_0^t dt b^2$.

Constante de tempo τ

O operador $K(t, t')$ depende de uma constante de tempo τ que regula a velocidade das variações de $a(t)$. Determinamos o melhor valor para τ pela análise dos fatores de Bayes.

O primeiro passo consiste em calcular o fator de Bayes $P[x] = \int Da P[a]P[x|a]$, onde explicitamos a dependência em τ . O resultado é facilmente obtido pois se trata de uma integral Gaussiana,

$$P[x|\tau] \propto \sqrt{\frac{\omega T}{\sinh(\omega T)}} e^{-\frac{1}{2b^2} \int dt \dot{x}(I-K)\dot{x}}. \quad (7.61)$$

O termo $\omega T / \sinh(\omega T)$ vem da normalização de $P[x|\tau]$ e está relacionado ao determinante de $I - K$ (lembramos que $\omega \equiv \tau^{-1}$). Esse operador possui um autovalor nulo associado ao subespaço em que $(I - K)u = 0$, para a autofunção $u(t) = \text{cte}$. O resultado (7.61) pressupõe uma regularização onde o autovalor nulo é simplesmente omitido. Vale notar que as outras probabilidades $P[a|\tau]$, $P[x|a]$, etc, sofrem exatamente do mesmo problema, mas não tínhamos nos preocupado em calcular a constante de normalização até agora.

Existe um problema de ordem prática com o fator de Bayes calculado em (7.61): $P[x|\tau]$ começa em zero e cresce até atingir um valor constante sem antes definir um pico de alta probabilidade para τ . Ou seja, o fator de Bayes não seleciona nenhum valor em particular, mas apenas diz que τ é não nulo, dando uma certa preferência a valores maiores de τ . Deste modo, não há uma escolha definitiva, tornando necessário especificar o priori $P(\tau)$ que compara modelos com diferentes valores de τ entre si.

De fato, a necessidade de se especificar o *priori* é comum a quase todos modelos com parâmetros contínuos. A prescrição de “maximizar o fator de Bayes para encontrar o modelo mais provável” assume implicitamente um priori equiprovável para todos os modelos (a probabilidade atribuída à cada modelo é “fa-

tor de Bayes” \times “*priori do modelo*”). No entanto, a noção de equiprobabilidade é problemática em um espaço de parâmetros contínuos: ela depende de uma escolha arbitrária de parametrização. Por exemplo, um *priori* equiprovável em τ não o será em ω já que $P(\tau) = \text{cte } d\tau = \text{cte } \left| \frac{d\tau}{d\omega} \right| d\omega = \frac{\text{cte}}{\omega^2} d\omega = P(\omega)$.

Para contornar esse problema, trocamos equiprobabilidade por uma noção geométrica: volumes iguais do espaço amostral carregam a mesma massa de probabilidade. A noção de volume independe da parametrização, mas requer que se especifique uma métrica para o espaço de parâmetros. Deste modo, trocamos o problema de especificar o *priori* para o problema de especificar a métrica do espaço de parâmetros. A segunda abordagem é promissora porque existe uma maneira sistemática de definir uma métrica associada aos parâmetros livres de uma distribuição: é a métrica de Fisher, definida como

$$g_{ij}(\theta) = \left\langle \frac{\partial}{\partial \theta_i} \ln P(x|\theta) \frac{\partial}{\partial \theta_j} \ln P(x|\theta) \right\rangle, \quad (7.62)$$

onde $\{\theta_i\}$ representa o conjunto de parâmetros livres em $P(x|\theta)$.

A interpretação deste resultado é que o *priori* definido como o elemento de volume deste espaço ($dV = \sqrt{\det g(\theta)} d\theta$) representa um estado de total ignorância sobre o valor dos parâmetros. Isto é análogo à atribuição equiprovável no caso de variáveis discretas: postulamos que todas as regiões no espaço de parâmetros são equivalentes, mas em geral trata-se de um espaço curvo. Qualquer informação adicional sobre os parâmetros θ distorceria o *priori*. De certo modo, muitos modelos *exigem* uma especificação adicional: o *priori* geométrico não é normalizável em um espaço de parâmetros com volume infinito e isto pode ocasionar uma série de problemas técnicos.

Voltando à tarefa de definir o *priori* para $P(\tau)$, o critério de *priori* de volume prescreve que $P(\tau) = \sqrt{g(\tau)} d\tau$. Observe que a métrica de Fisher utiliza o fator de Bayes $P[x|\tau]$ mostrado na Eq. (7.61), que é uma Gaussiana em x . A

métrica de Fisher associada a uma Gaussiana com média nula depende apenas dos autovalores da matriz de covariância. Um cálculo relativamente longo nos leva a

$$P(\tau) \propto \frac{1}{\tau} \sqrt{\left(\frac{\omega T}{\sinh(\omega T)}\right)^2 + \frac{\omega T}{\tanh(\omega T)} - 2}, \quad (7.63)$$

que possui uma dependência indesejável em T . A fórmula anterior pressupõe que o estado de ignorância a respeito de τ depende da duração T de uma observação posterior, o que é um contra-senso.

Uma maneira de contornar o problema anterior é interpretar (7.63) como o resultado de um *gedankenexperiment* onde T joga o papel de um parâmetro livre. Consideramos o limite $T \rightarrow 0$ para representar a situação de ignorância total sobre τ : esta seria a informação codificada em uma observação que dura apenas um lapso infinitesimal de tempo. Esse limite resulta em $P(\tau) \propto \tau^{-3}$, que é um priori não-normalizável assim como a Eq. (7.63).

Após definir o priori $P(\tau)$, calculamos a probabilidade $P(\tau|x) \propto P(\tau)P[x|\tau]$ para o valor correto de τ condicionado aos dados experimentais. Ela nos permite escolher um valor para τ segundo um critério de optimalidade. Ao discutir o problema de regressão anteriormente, consideramos a prescrição de selecionar o modelo mais provável (o que maximiza $P(\tau|x)$). Essa não é a única alternativa: é comum utilizar estimativas pontuais de τ como a mediana ou a média (mas no presente caso, a média de τ segundo $P(\tau|x)$ é infinita).

Existe outra opção que é mais adequada no nosso caso. A distribuição de $P(\tau|x)$ é bastante espalhada e nenhum valor de τ sozinho é representativo de toda a distribuição. Assim, ao invés de utilizar a probabilidade posterior $P[a|x, \tau]$ para um valor específico de τ , podemos utilizar

$$P[a|x] = \int d\tau p[a, \tau|x], \quad (7.64)$$

$$= \int d\tau p(\tau|x) P[a|x, \tau]. \quad (7.65)$$

A média de $a(t)$ é

$$\langle a(t) \rangle = \int d\tau p(\tau|x) \mu_\tau(t), \quad (7.66)$$

onde $\mu_\tau(t) \equiv \int_0^T ds K_\tau(t, s) \dot{x}(s)$ é a média de $a(t)$ segundo $P[a|x, \tau]$. Já a covariância é dada por

$$\Sigma(t, t') = \int d\tau p(\tau|x) [b^2 K_\tau(t, t') + \mu_\tau(t) \mu_\tau(t')] - \langle a(t) \rangle \langle a(t') \rangle. \quad (7.67)$$

As integrais anteriores podem ser implementadas numericamente sorteando vários valores de τ a partir da probabilidade $P(\tau|x)$, onde tomamos a média do integrando sobre o *ensemble* de τ 's.

Os resultados mostrados na Fig. 7.4 e na Fig. 7.5 utilizam esse método. Para produzir estas figuras, utilizamos 50 amostras onde a região de 50% da massa de probabilidade de $p(\tau|x)$ se situa em $\tau \in [0, 10, 0, 23]$, com mediana 0,15. O coeficiente de difusão computado utilizando a variação quadrática foi de 0,18 enquanto o valor verdadeiro era 0,20.

Decomposição espectral de K

Em várias etapas durante as seções anteriores, foi necessário utilizar os autovalores da matriz $I - K$, que podem ser facilmente computados a partir da decomposição espectral de K . Tanto K quanto K^{-1} compartilham os mesmos autovetores com autovalores recíprocos. Deste modo, consideramos a equação de autovalor $K^{-1} \phi_n = \lambda_n^{-1} \phi_n$, que pode ser resolvida trivialmente. O resultado é $\phi_n(t) = \cos(\pi n t / T)$ e $\lambda_n^{-1} = 1 + (\pi n \tau / T)^2$, que é bastante fortuito já que, ao expandir $\dot{x}(t)$ em uma série de cossenos, a integral dupla no expoente de $P[x|\tau]$ (veja a Eq. [7.61]), se torna uma somatória simples. Esse procedimento reduz o tempo de computação de $O(N^2)$ para $O(N)$, onde N é o número de coeficientes de Fourier na expansão de $\dot{x}(t)$. Para maior eficiência, a expansão

de $x(t)$ em série de cossenos pode ser feita utilizando o algoritmo de FFT (*Fast Fourier Transforms*), com a vantagem de obter uma melhor acuidade numérica e um tempo de computação de $O(N \ln N)$ ao invés de $O(N^2)$.

7.4 Conclusão

O método das médias móveis é frequentemente justificado com o argumento que as flutuações se cancelam na média. Por mais que isto pareça intuitivamente correto, não se pode tomar esta afirmação como uma justificativa séria. Antes de tudo, é necessário responder várias questões: quais são as propriedades das tendências e flutuações? qual é a melhor calibragem para a média móvel? existe uma incerteza inerente ao método? até que ponto podemos confiar que as flutuações foram completamente suprimidas?

Tratamos todas estas questões ao formular a média móvel como um método de inferência estatística. O uso da estatística Bayesiana é fundamental pois a inferência é dominada pelas características *a priori* das tendências expressas na ação (7.47). Deste modo, o objetivo consiste em obter a distribuição de probabilidade para as tendências, levando em conta a série temporal observada. O resultado foi bem sucedido e permite responder a todas as questões apresentadas anteriormente.

Utilizamos métodos de integrais de trajetória e teoria de campos em física para a análise do problema de inferência. Existe outros trabalhos com esse viés, como Bialek, *et. al.* [5] e Holy [22], mas que não são aplicados à análise de séries temporais. Acreditamos que a relação entre métodos estatísticos e teoria de campos seja bastante profícua e merece ser investigada no futuro.

Capítulo 8

Perspectivas e conclusão

Como foi dito anteriormente, o presente trabalho consiste em um esforço de aplicar alguns resultados da teoria de probabilidades a alguns problemas de física. No caso específico do último capítulo, também aplicamos métodos de física a um problema de inferência estatística. Uma vez que este texto não trata de um tema único e uma breve conclusão foi feita ao final de cada capítulo, aqui nos restringimos apenas a considerações gerais.

Alguns temas recorrentes foram o uso do formalismo de função características, processos estocásticos e teoria Bayesiana de probabilidades. Nenhuma destas idéias é totalmente estranha aos físicos. Seria possível, então, encontrar resultados originais tentando preencher as lacunas? O presente trabalho se equiparou às expectativas produzindo algo de novo?

Acredito que sim.

O formalismo de funções características é usado de modo bastante incipiente em física e raramente vai muito além da constatação (um pouco imprecisa) que a função característica é a transformada de Fourier da densidade. No entanto, existem vários resultados importantes que podem ser utilizados em problemas de física. Mencionamos, por exemplo, o teorema de De Finetti que utilizamos para caracterizar as “Hamiltonianas” dos processos Markovia-

nos descritos por integrais funcionais. No conhecimento do autor, esta conexão não havia sido percebida anteriormente, o que é um pouco surpreendente, pois trata-se de um resultado bastante fundamental.

Esse é apenas um exemplo de casos onde um resultado bem estabelecido da teoria de probabilidade não é amplamente conhecido pela comunidade de física, ainda que seja potencialmente útil. Como não pretendo listar detalhadamente cada uma das contribuições (isto foi feito no fim de cada capítulo), tentarei delinear uma direção geral para seguir a partir daqui. Um tema interessante omitido ao longo do texto é a possibilidade de definir processos estocásticos para sistemas quânticos. Não existe uma prescrição universalmente aceita que define a probabilidade conjunta de observáveis que não comutam (como geralmente é o caso de um operador de Heisenberg avaliado em tempos distintos). Mais ainda, é possível provar que em alguns casos é impossível construir tal probabilidade. Isto não impede o aparecimento de conceitos como *correlações quânticas*, ainda que a rigor elas possam não ter a menor conexão com o conceito clássico. Desta forma, é interessante descobrir quais resultados aqui obtidos podem ser formulados em um domínio quântico e se isto resulta em novas propriedades. (Por exemplo: o que se poderia dizer sobre um processo de saltos quântico?)

Para finalizar, ressalto que este texto foi concebido em torno de uma idéia central muito abrangente: que as teorias de processos estocásticos, probabilidades e funções características *como feitas pelos matemáticos* enriquecem a física. Certamente não é possível explorar esta idéia até suas últimas conseqüências e também não defendo a posição que um físico deve se candidatar aos mesmos padrões de rigor que um matemático. No entanto, se o leitor se convenceu que pelo menos uma das idéias matemáticas aqui apresentadas é importante para sua pesquisa e merece ser aprofundada, este trabalho cumpriu seu objetivo.

Apêndice I: Funções especiais

Apresentamos algumas funções especiais utilizadas ao longo do texto nos casos onde pode haver alguma ambigüidade quanto à definição.

Função de Bessel

As funções de Bessel são as soluções da equação diferencial

$$x^2 \frac{d^2}{dx^2} y(x) + x \frac{d}{dx} y(x) + (x^2 - \nu^2) y(x) = 0. \quad (8.1)$$

Como toda função definida desta maneira, é necessário estabelecer as condições de contorno e a normalização das soluções para eliminar qualquer ambigüidade. Definimos as funções de Bessel do primeiro tipo como as soluções de (8.1) finitas na origem. Estas funções podem ser expressas em série de potências como

$$J_\nu(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n! \Gamma(n + \nu + 1)} \left(\frac{x}{2}\right)^{2n+\nu}. \quad (8.2)$$

A segunda solução linearmente independente de $J_\nu(x)$ é a chamada função de Bessel do segundo tipo, definida como

$$Y_\nu(x) = \begin{cases} \frac{J_\nu(x) \cos(\nu\pi) - J_{-\nu}(x)}{\sin(\nu\pi)} & \nu \notin \mathbb{N} \\ \lim_{\nu \rightarrow n} Y_\nu(x) & \nu \in \mathbb{N} \end{cases}. \quad (8.3)$$

No caso em que ν é um número inteiro, definimos a função de Bessel pelo limite mostrado anteriormente.

Finalmente, temos as funções de Bessel modificadas que são as soluções da equação de Bessel após a troca de variáveis $x \mapsto ix$. Observe que o efeito de realizar esta troca de variáveis corresponde somente à mudar o sinal do termo com $x^2 y(x)$ na equação diferencial (8.1). Definimos o par de funções

$$I_\nu(x) = i^{-\nu} J_\nu(ix), \quad (8.4)$$

$$K_\nu(x) = \frac{\pi}{2} \frac{I_{-\nu}(x) - I_\nu(x)}{\sin(\nu\pi)}, \quad (8.5)$$

conhecidas como as funções de Bessel modificadas do primeiro e segundo tipo

Função hipergeométrica

Diversas funções especiais, assim como boa parte das funções analíticas mais comuns podem ser escritas como casos especiais de funções hipergeométricas.

A definição geral,

$${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; z) = \sum_{n=0}^{\infty} \frac{(a_1)_n \dots (a_p)_n z^n}{(b_1)_n \dots (b_q)_n n!}, \quad (8.6)$$

possui a vantagem sobre várias funções especiais de não apresentar qualquer ambigüidade quanto à normalização ou a condições de contorno. Na Eq. (8.6) utilizamos o símbolo de Pochhammer, definido como $(a)_n = (a + n - 1)(a)_{n-1}$, com $(a)_0 = 1$.

Integrais do seno e cosseno

As funções de integral do seno e do cosseno comumente aparecem no cálculo de transformadas de Fourier de alguns tipos de funções. Definimos ambas as

funções pelas integrais abaixo

$$\text{Si}(x) = \int_0^x dy \frac{\sin(y)}{y} \quad (8.7)$$

$$\text{Ci}(x) = \gamma + \ln(x) + \int_0^x dy \frac{\cos(y) - 1}{y}, \quad (8.8)$$

onde $\gamma = 0.577\dots$ é a constante de Euler-Mascheroni.

Apêndice II: Inferência Bayesiana

Nesta seção apresentamos uma pequena introdução ao tema da inferência Bayesiana. Ainda que do ponto de vista lógico seja interessante manter esta seção na introdução da tese — afinal trata-se apenas da revisão de conceitos estabelecidos —, o conteúdo aqui tratado não será relevante até o Cap. 7. Deste modo, o leitor pode pular esta seção para se adiantar aos primeiros capítulos e retomá-la sem o menor prejuízo antes de começar o capítulo em questão.

Muitos problemas de inferência podem ser formulados segundo o esquema geral da regra de Bayes. Denotemos por θ um conjunto de parâmetros desconhecidos, os quais se deseja estimar, e \mathbf{x} uma possível seqüência de dados experimentais. A regra de Bayes relaciona a probabilidade de obter uma certa seqüência de dados — $P(\mathbf{x}|\theta)$ que caracteriza o modelo estatístico — com a probabilidade $P(\theta|\mathbf{x})$ que codifica a ignorância sobre o valor correto de θ após considerar os dados experimentais \mathbf{x} . Escrevemos

$$P(\theta|\mathbf{x}) = P(\theta) \frac{P(\mathbf{x}|\theta)}{P(\mathbf{x})}, \quad (8.9)$$

onde $P(\theta)$ é conhecido como a probabilidade *a priori*, (ou simplesmente *priori*), codifica a ignorância sobre os parâmetros anterior à aquisição de dados. O fator $P(\mathbf{x})$ é uma constante de normalização ($P(\mathbf{x}) = \sum_{\theta} P(\theta)P(\mathbf{x}|\theta)$).

Convém ilustrar a discussão com um modelo concreto. Seja uma moeda com vício onde o parâmetro desconhecido θ representa a chance de obter uma das faces (cara) e \mathbf{x} é uma seqüência de resultados do tipo "cara,

cara, coroa, cara, ...". O modelo estatístico é

$$P(\mathbf{x}|\theta) = \frac{N!}{n!(N-n)!} \theta^n (1-\theta)^{N-n}, \quad (8.10)$$

onde N representa o número de jogadas e n o número de caras; ambos podem ser extraídos facilmente de \mathbf{x} .

A análise Bayesiana exige que se especifique o *priori* $P(\theta)$. Existe uma arbitrariedade na escolha de $P(\theta)$ pois esse não faz parte da caracterização do modelo estatístico. As próximas duas seções apresentam critérios comuns para se determinar esta probabilidade em diversos casos.

Priori conjugado

Imagine que seja possível separar os dados experimentais em duas partes. No exemplo da moeda, basta separar o vetor $\mathbf{x} = (\mathbf{y}_1, \mathbf{y}_2)$ onde \mathbf{y}_1 representa o resultado das N_1 primeiras jogadas e \mathbf{y}_2 das N_2 jogadas posteriores. Denotamos n_1 e n_2 como o número de caras obtido em cada uma destas seqüências. A probabilidade posterior para θ é

$$p(\theta|\mathbf{y}_1) \propto \pi(\theta) \theta^{n_1} (1-\theta)^{N_1-n_1}. \quad (8.11)$$

Ao considerar as N_2 jogadas seguintes, esta probabilidade muda de uma maneira bastante conveniente. Basta substituir $N_1 \mapsto N_1 + N_2$ e $n_1 \mapsto n_1 + n_2$, de maneira que

$$p(\theta|\mathbf{y}_1, \mathbf{y}_2) \propto \pi(\theta) \theta^{n_1+n_2} (1-\theta)^{N_1+N_2-n_1-n_2}. \quad (8.12)$$

É fácil verificar que cada seqüência de jogadas atualiza $p(\theta|\mathbf{x})$ apenas por uma mudança dos expoentes de θ e $1-\theta$. Desta forma, a escolha matemática-

mente mais conveniente é um prior $p(\theta)$ que seja escrito nesta mesma forma paramétrica

$$\pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1}, \quad (8.13)$$

conhecida como a distribuição beta. Os prioris que preservam a forma paramétrica do posterior são muito utilizados por sua conveniência matemática e recebem o nome de prioris conjugados [18].

Ainda que seja um pouco arbitrária, esta escolha permite montar uma regra de atualização simples. O prior $\pi(\theta)$ é caracterizado por dois coeficientes a e b denominados de hiper-parâmetros (usamos esse nome para não confundir com os *parâmetros* θ). O posterior $p(\theta|\mathbf{x})$ é calculado redefinindo os hiper-parâmetros $a \mapsto a' = a + n$ e $b \mapsto b' = b + N - n$. Lembramos algumas propriedades da distribuição:

$$\langle \theta \rangle = \frac{a}{a+b}, \quad (8.14)$$

$$\langle (\theta - \langle \theta \rangle)^2 \rangle = \frac{ab}{(a+b)^2(a+b+1)}. \quad (8.15)$$

A moda da distribuição é $\frac{a-1}{a+b-2}$ se $a > 1$ e já a mediana somente pode ser expressa como uma forma complicada de funções especiais.

Note que existe uma liberdade na escolha de hiper-parâmetros de (8.13). É razoável impor que $a = b$, forçando com que a média de $\pi(\theta)$ seja $1/2$, característica de uma moeda honesta. O prior uniforme é dado por $a = b = 1$ enquanto o prior de maior variância (que teoricamente se traduziria na maior ignorância possível para a e b) corresponde a $a = b = 0$. Se quisermos que a moda da distribuição também seja igual a $1/2$, então é necessário impor que $a = b > 1$.

Deste modo, o critério apresentado seleciona uma família de distribuições para $\pi(\theta)$. A distribuição específica dentro desta família a ser utilizada como o priori depende de considerações adicionais. Apresentamos algumas considera-

ções na próxima seção. Note que nem todo modelo possui um priori conjugado, portanto, ainda que o argumento apresentado mostre grande conveniência matemática, ele também carece de universalidade.

Priori volumétrico

Na maioria dos problemas de inferência é desejável que o priori não seja muito informativo. Isto faz parte do critério de uma falseabilidade que diz que a partir de um certo volume de dados, a crença inicial deve tornar-se irrelevante. É interessante levar esta perspectiva ao extremo: qual escolha de priori descreveria o estado de completa ignorância a respeito do modelo estatístico considerado?

Uma escolha natural é a atribuição uniforme $\pi(\theta) = \text{cte}$, algumas vezes chamada de princípio de insuficiência de Laplace. Vale notar que, em se tratando de parâmetros contínuos, esta escolha é essencialmente arbitrária. Uma distribuição uniforme segundo uma certa parametrização pode não ser uniforme em outras. No exemplo da moeda, a escolha $\pi(\theta) = 1$ corresponderia à atribuição uniforme. No entanto, é perfeitamente possível reformular o problema utilizando uma parametrização diferente, digamos, $\zeta \equiv -\ln(\theta)$. Pela regra de transformação de variáveis, $\pi(\zeta) = \pi(\theta)d\theta/d\zeta$, que fornece o priori não equiprovável $\pi(\zeta) = e^{-\zeta}$. Portanto a escolha por um priori uniforme só faz sentido se for possível argumentar em favor de uma determinada parametrização, o que nem sempre é viável.

Uma maneira de recolocar o problema consiste em reformular a idéia de equiprobabilidade a partir da geometrização do espaço de parâmetros θ . Associamos equiprobabilidade à noção de volume: porções do espaço de parâmetros com o mesmo volume possuiriam a mesma probabilidade. O volume é invariante por reparametrização, de modo que não apareceriam algumas das patologias associadas à noção tradicional de equiprobabilidade.

Em um espaço métrico com coordenadas $\theta = (\theta^1, \theta^2, \dots, \theta^N)$, o elemento de distância é dado por $dl^2 = g_{ij}d\theta^i d\theta^j$ e o elemento de volume é $dV = \sqrt{\det g} d\theta^1 \dots d\theta^N$. Para definir uma noção análoga à de equiprobabilidade em um espaço de parâmetros contínuos é necessário especificar a métrica g_{ij} , ou de maneira equivalente, as noções de distância e de volume. Ainda que isto não seja imediatamente óbvio, existe uma maneira sistemática de calcular métricas g_{ij} a partir de um modelo estatístico $p(\mathbf{x}|\theta)$.

A métrica para o espaço de parâmetros é fornecida de maneira unívoca pela matriz de informação de Fisher [19, 68, 2, 1], definida como

$$g_{ij} = - \int d\mathbf{x} p(\mathbf{x}|\theta) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln p(\mathbf{x}|\theta). \quad (8.16)$$

A partir dela, o prior volumétrico é dado por $\sqrt{\det g_{ij}(\theta)} d\theta_1 \dots d\theta_N$, que associamos à situação de plena ignorância. Note que em muitas situações é possível conceber uma reparametrização que torna o espaço de parâmetros plano, fornecendo um prior uniforme nestes parâmetros, mas esta expectativa não precisa ser sempre cumprida.

Uma dificuldade séria com o prior volumétrico aparece em espaços com volume infinito. Nestes casos $\pi(\theta)$ não pode ser normalizado. É claro que, a rigor, tais priors não representariam distribuições de probabilidades. Muitas vezes, no entanto, é possível pensar neles como o limite de uma seqüência de probabilidades bem definidas. Por exemplo, um prior equiprovável no espaço euclidiano pode ser pensado como o limite de uma distribuição uniforme sobre uma seqüência crescente de volumes V . Caso seja possível proceder desta maneira, a probabilidade não-normalizável é denominada distribuição imprópria.

Muitas vezes, a despeito da normalização de $\pi(\theta)$, a probabilidade que geralmente interessa, $p(\theta|\mathbf{x})$ ainda pode ser normalizada. Lembre-se que o teorema de Bayes pode ser escrito como $p(\theta|\mathbf{x}) \propto \pi(\theta)p(\mathbf{x}|\theta)$ e um termo

divergente em $\pi(\theta)$ eventualmente se cancela na normalização de $p(\theta|\mathbf{x})$. De qualquer maneira, priors impróprios devem ser utilizados com cautela.

O modelo da moeda representa uma das poucas situações práticas em que o prior de ignorância é normalizável. De fato, o prior calculado à partir de (8.16) é dado por uma distribuição beta como em (8.13), em que os parâmetros a e b são iguais a $1/2$. Esse resultado talvez seja surpreendente já que $\pi(\theta)$ assume o valor *mínimo* em $\theta = 1/2$.

Probabilidade preditiva

O teorema de Bayes permite utilizar uma seqüência de dados \mathbf{x} para refinar o conhecimento sobre os parâmetros de um modelo estatístico. No entanto, muitas vezes o conhecimento sobre os parâmetros é utilizado somente para prever o resultado de experimentos futuros.

No modelo da moeda, a questão seria a seguinte: qual é a probabilidade de se obter uma cara sendo que, anteriormente, obteve-se n caras em N jogadas?

Essa pergunta é respondida avaliando a probabilidade $p(x|\mathbf{x})$ de obter um valor x condicionada aos dados pregressos. A dependência em θ pode ser explicitada considerando a marginal da probabilidade conjunta de x e θ ,

$$p(x|\mathbf{x}) = \int d\theta p(x, \theta|\mathbf{x}) = \int d\theta p(\theta|\mathbf{x})p(x|\theta, \mathbf{x}). \quad (8.17)$$

Na maioria das situações, se cada dado x_i é estatisticamente independente dos outros, então $p(x|\theta, \mathbf{x}) = p(x|\theta)$. Ao realizar esta integral para o modelo da moeda, vemos que a probabilidade preditiva após uma seqüência \mathbf{x} com n caras em N jogadas é dada por

$$P(\text{"cara"}|\mathbf{x}) = \frac{a + n}{a + b + N}, \quad (8.18)$$

onde a e b são os hiper-parâmetros apresentados em (8.13). Somente o prior simétrico com $a = b$ produz $P(\text{"cara"}) = 1/2$ antes de realizar qualquer jogada ($n = N = 0$). Isto é o que normalmente se espera de uma moeda.

A probabilidade preditiva para $a = b = 1$, que corresponde a um prior uniforme em θ , implica na regra de sucessão de Laplace. Quanto mais $a = b$ seja maior 1, mais conservadora é a estimativa de $P(\text{"cara"} | \mathbf{x})$ em se tratando de desvios com relação a uma moeda honesta. Segundo a regra de Laplace, após obter uma cara na primeira jogada, $P(\text{"cara"} | 1) = 2/3$, enquanto que a estimativa de frequências na qual $a = b = 0$, forneceria $P(\text{"cara"} | 1) = 1$. Outras escolhas comuns são o prior de volume, no qual $a = b = 1/2$, implicando $P(\text{"cara"} | 1) = 3/4$ e o prior de Wigner $a = b = 3/2$, que corresponde a um semi-círculo em torno de $p = 1/2$ resultaria em $P(\text{"cara"} | 1) = 5/8$.

De qualquer maneira, independentemente da escolha do prior, quando $N \rightarrow \infty$, $P(\text{"cara"} | \mathbf{x}) \rightarrow n/N$, e o valor escolhido para a e b torna-se irrelevante. Esse é o limite em que vale a estatística freqüentista, que converge para o resultado Bayesiano independente da escolha do prior. Isto é esperado já que os métodos freqüentistas não utilizam o priori. Nesse sentido, podemos pensar que a estatística Bayesiana estende os métodos tradicionais para as situações em que a informação a priori é importante ou nos casos em que os dados não são abundantes o suficiente para que a convergência de frequência em probabilidade seja confiável.

Comparação de modelos

Uma questão importante na análise estatística consiste em dizer qual modelo é melhor suportado por um certo conjunto de dados. A resposta pode ser extraída do cálculo de probabilidades. Para isto, explicitamos que todas probabilidades que aparecem na regra de Bayes (8.9) deveriam, a rigor, serem

condicionadas a um modelo que denominamos M_1 . Deste modo, para ser explícito com relação a esse ponto, deveríamos ter escrito desde o começo

$$p(\theta_1|\mathbf{x}, M_1) = \pi(\theta_1|M_1) \frac{p(\mathbf{x}|\theta_1, M_1)}{p(\mathbf{x}|M_1)}, \quad (8.19)$$

onde algo semelhante valeria para um eventual segundo modelo M_2 .

Queremos saber a probabilidade de cada modelo M_1 e M_2 estar correto condicionada aos dados \mathbf{x} . Para determinar $P(M_i|\mathbf{x})$, utilizamos o teorema de Bayes novamente[26]

$$P(M_i|\mathbf{x}) = P(M_i) \frac{P(\mathbf{x}|M_i)}{P(\mathbf{x})}. \quad (8.20)$$

Ao comparar dois modelos distintos, a razão $P(M_1|\mathbf{x})/P(M_2|\mathbf{x})$ assume a forma

$$\frac{P(M_1|\mathbf{x})}{P(M_2|\mathbf{x})} = \frac{P(M_1)}{P(M_2)} \times \frac{P(\mathbf{x}|M_1)}{P(\mathbf{x}|M_2)}. \quad (8.21)$$

O fator $P(M_1)/P(M_2)$ mostra o quanto M_1 é preferível *a priori* com relação à M_2 . Muitas vezes tal preferência não existe e para tratar os modelos de maneira equivalente, supomos que $P(M_1) = P(M_2)$. O outro termo pode ser calculado diretamente a partir dos dados, é conhecido como *fator de Bayes* ou *evidência*. Lembramos que $P(\mathbf{x}|M_i)$ corresponde à normalização da regra de Bayes, de modo que

$$E(M_i) = P(\mathbf{x}|M_i) = \int d\theta_i \pi(\theta_i) p(\mathbf{x}|\theta_i, M_i). \quad (8.22)$$

Podemos calcular a evidência em forma fechada para o modelo M_1 da moeda com vício. Ela possui o valor

$$E(M_1) = \frac{N!}{n!(N-n)!} \frac{\beta(n+a, N-n+b)}{\beta(a, b)}, \quad (8.23)$$

onde a função beta é o binomial generalizado definido como $\beta(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$. Esse modelo pode ser comparado com o modelo M_2 de uma moeda não viciada na qual

$$P(\mathbf{x}|M_2) = \frac{N!}{2^N n!(N-n)!}. \quad (8.24)$$

A razão $E(M_1)/E(M_2)$ possui algumas propriedades interessantes. Se $a = b = 0$ então $E(M_2)/E(M_1) \rightarrow 0$ para qualquer conjunto de dados. Esse resultado ilustra o fato mais geral que qualquer modelo com priori impróprio é automaticamente rejeitado, mesmo quando contrastado com um modelo obviamente errôneo (nesse caso, selecionaríamos o modelo equiprovável mesmo quando as frequências converjam para longe de $1/2$).

Para outros valores, ocorre um comportamento interessante. Grosseiramente, se o valor $\theta = 1/2$ estiver dentro de uma região que se comporta grosseiramente como $n/M \pm 1/\sqrt{N}$, o modelo equiprovável seria escolhido. Esse é o resultado esperado do ponto de vista da estatística de frequências: a frequência converge para a probabilidade com um erro de aproximadamente $1/\sqrt{N}$ no número de amostras. A escolha de a e b influencia muito pouco o resultado final se N for grande.

Apêndice III: Teorema de Doob generalizado

Generalizamos o teorema de Doob, que fornece a condição para que um processo estacionário Gaussiano seja Markoviano [13]. Citamos o teorema original sem demonstrá-lo.

Theorem 8.1 (Doob). *Todo processo estocástico estacionário cuja probabilidade a 2-pontos é uma Gaussiana, é Markoviano se e somente se possui uma função de covariância exponencial,*

$$\langle\langle u(t)u(t') \rangle\rangle = \langle\langle u^2 \rangle\rangle e^{-\gamma|t-t'|}, \quad (8.25)$$

onde $\langle\langle \dots \rangle\rangle$ denota os momentos centrais (e.g., $\langle\langle u^n \rangle\rangle = \langle\langle (u - \langle u \rangle)^n \rangle\rangle$).

Relaxamos um pouco as condições de Gaussianidade e estacionariedade. Consideramos processos autônomos e aditivos, que são aqueles cuja probabilidade de transição $P(u_t|u_{t'}) = G_\tau(u|u')$ pode ser escrita como

$$G_\tau(u|u') = q_\tau(u - A(\tau)u' + B(\tau)), \quad (8.26)$$

onde $A(\tau)$ e $B(\tau)$ são funções arbitrárias. A interpretação física da condição anterior é que $u_{t+\tau}$ é gerada a partir de u_t segundo a seguinte expressão

$$u_{t+\tau} = A(\tau)u_t + B(\tau) + e(\tau), \quad (8.27)$$

onde $e(\tau)$ é uma variável aleatória arbitrária com a distribuição de probabilidades $P(e(\tau)) = q_\tau(e)de$. Vale a observação que o processo Gaussiano é aditivo, e assim o teorema de Doob original também será contemplado aqui.

O primeiro passo para verificar quando $G_\tau(u|u')$ é uma probabilidade Markoviana é substituí-la na equação de Chapman-Kolmogorov, que repetimos por conveniência:

$$G_{\tau+\tau'}(u|u') = \int dv G_\tau(u|v)G_{\tau'}(v|u'). \quad (8.28)$$

Multiplicamos em ambos os lados por e^{izu} e integramos em u para obter a função característica. O resultado desta operação é

$$\phi_{\tau+\tau'}(z)e^{iz[A(\tau+\tau')u'+B(\tau+\tau')]} = \int dv \phi_\tau(z)e^{iz[A(\tau)v+B(\tau)]}q_{\tau'}(v - A(\tau')u' - B(\tau')), \quad (8.29)$$

onde definimos $\phi_\tau(z)$ como a função característica de $q_\tau(e)$ e utilizamos a propriedade de deslocamento da transformada de Fourier.

A integral em v pode ser identificada essencialmente com a definição da função característica $\phi_\tau(z)$ avaliada em $A(\tau)z$ ao invés de z . Deste modo, ficamos com

$$\phi_{\tau+\tau'}(z)e^{iz[A(\tau+\tau')u'+B(\tau+\tau')]} = \phi_\tau(z)\phi_{\tau'}(A(\tau)z)e^{iz[A(\tau)A(\tau')u'+B(\tau)+A(\tau)B(\tau')]} \quad (8.30)$$

Separando o fator que depende somente de u' , fica óbvio pela equação anterior que

$$A(\tau + \tau') = A(\tau)A(\tau'), \quad (8.31)$$

e portanto $A(\tau) = e^{-\gamma\tau}$, onde γ é uma constante arbitrária.

A média da distribuição $q_\tau(e)$ pode ser escolhida arbitrariamente, já que podemos ajustá-la na definição de $B(\tau)$. Deste modo, escolhemos por conveniência $\langle e_t(\tau) \rangle = 0$. É possível calcular a média de uma distribuição avaliando derivada da sua função característica no ponto $z = 0$. Ao fazer isto na

Eq. (8.30), após considerar o resultado em (8.31), é fácil mostrar que

$$B(\tau + \tau') = e^{-\gamma\tau} B(\tau') + B(\tau). \quad (8.32)$$

Ainda que a Eq. (8.32) não seja uma relação funcional tão trivial quanto (8.31), é possível determinar $B(\tau)$ com certa facilidade. Avaliamos a derivada $\frac{\partial^2}{\partial\tau\partial\tau'}$ da Eq. (8.32) em $\tau = 0$, onde o resultado se escreve como

$$\ddot{B}(\tau') = -\gamma\dot{B}(\tau'). \quad (8.33)$$

A solução geral é

$$B(\tau) = \kappa + \eta e^{-\gamma\tau}, \quad (8.34)$$

onde κ e η são constantes de integração arbitrárias. Para que a condição inicial $G_0(x|x') = \delta(x - x')$ seja satisfeita, é necessário que $B(0) = 0$ e deste modo concluímos que $\kappa + \eta = 0$.

Obtemos, até agora, duas condições necessárias para que a probabilidade (8.26) seja Markoviana: as funções $A(\tau)$ e $B(\tau)$ devem satisfazer as relações

$$A(\tau) = e^{-\gamma\tau}; \quad B(\tau) = \kappa (1 - e^{-\gamma\tau}), \quad (8.35)$$

onde γ e κ são constantes arbitrárias. Substituindo estas relações em (8.30), mostramos que também é necessário que

$$\phi_{\tau+\tau'}(z) = \phi_\tau(z)\phi_{\tau'}(e^{-\gamma\tau}z). \quad (8.36)$$

Se $\gamma = 0$, esta equação se reduz à Eq. (3.3), cuja solução geral (3.5) define uma probabilidade infinitamente divisível.

Mostramos agora que se $\gamma \neq 0$, então $\phi_\tau(z)$ deve necessariamente constituir uma lei estável simétrica. Definimos $\phi_\tau(z) = e^{W_\tau(z)}$, que ao substituir na

Eq. (8.36) leva a

$$W_{\tau+\tau'}(z) = W_{\tau}(z) + W_{\tau'}(e^{-\gamma\tau}z). \quad (8.37)$$

Após diferenciar esta equação por $\frac{\partial^2}{\partial\tau\partial\tau'}$ e avaliá-la em $\tau = 0$, obtemos

$$\ddot{W}_{\tau'}(z) = -\gamma z \frac{\partial}{\partial z} \dot{W}_{\tau'}(z), \quad (8.38)$$

onde o ponto denota a derivada temporal. A solução geral pode ser obtida pelo método de separação das variáveis, e resulta em

$$W_{\tau}(z) = f(z) - (s_0 z)^{\lambda/\gamma} e^{-\lambda\tau}, \quad (8.39)$$

onde $f(z)$, λ e s_0 são constantes de integração arbitrárias.

Determinamos a função $f(z)$ ao impor a condição $W_0(z) = 0$, necessária para que $G_0(x|x') = \delta(x - x')$. Assim, obtemos a forma geral de $W_{\tau}(z)$

$$W_{\tau}(z) = s_0^{\lambda/\gamma} (1 - e^{-\lambda\tau}) z^{\lambda/\gamma}, \quad (8.40)$$

que claramente caracteriza uma função característica estável. Identificamos o expoente de estabilidade $\alpha = \lambda/\gamma$, onde $0 < \alpha \leq 2$, e o parâmetro de escala

$$s(\tau)^{\alpha} = s_0^{\alpha} (1 - e^{-\alpha\gamma\tau}). \quad (8.41)$$

O coeficiente de assimetria β e o parâmetro de localização ℓ desta função característica são nulos.

Cumulantes do processo estocástico

Veremos agora como as condições obtidas anteriormente se refletem na média e na covariância de $u(t)$. Para definir um processo Markoviano é necessário

especificar 2 probabilidades distintas: a probabilidade de transição $G_\tau(u|u')$ considerada anteriormente e a probabilidade marginal $P(u_t) = f_t(u)du$ que deve obedecer à equação

$$f_t(u) = \int du' f_0(u') G_t(u|u'). \quad (8.42)$$

Repetindo um procedimento semelhante ao da Eq. (8.30), relacionamos a função característica $\psi_t(z) = \int du f_t(u)e^{izu}$ com $\psi_{t=0}(z)$ que deve ser conhecida de antemão. Um cálculo trivial mostra que

$$\psi_t(z) = \psi_0(e^{-\gamma t}z)\phi_t(z) e^{iz\kappa(1-e^{-\gamma t})}. \quad (8.43)$$

Uma consequência interessante deste resultado é que se $\gamma > 0$ a distribuição de probabilidade assintótica não dependerá de $\psi_0(z)$. Essa propriedade é conhecida como condição de mistura.

A Eq. (8.43) permite relacionar facilmente os cumulantes da distribuição $f_t(u)$ com os de $f_0(u)$. Para a média, temos

$$c_1(t) = \kappa + (c_1 - \kappa) e^{-\gamma t}, \quad (8.44)$$

enquanto para os outros cumulantes a resposta é

$$c_n(t) = k_n(t) + c_n e^{-n\gamma t}, \quad (8.45)$$

onde $c_n = c_n(0)$ são os cumulantes de $\psi_0(z)$ e k_n são os cumulantes de $\phi_t(z)$.

Existem duas dependências temporais possíveis para $k_n(t)$. A primeira é acontece no caso especial em que $\gamma = 0$ e a probabilidade de transição assume a forma de um processo infinitamente divisível (a Eq. [3.5] reproduzida abaixo),

$$\phi_t(z) = e^{\lambda t(\psi(z)-1)}. \quad (8.46)$$

Nesse caso, $k_n(t) = \lambda t \mu_n$ onde λ é uma constante arbitrária e μ_n são os momentos estatísticos de $\psi(z)$.

Já no caso em que $\gamma \neq 0$, vimos que se $\phi_t(z)$ trata de uma distribuição estável. Como a única distribuição estável com média e variância finitas é a Gaussiana, aqui ficamos restritos ao caso em que o expoente de estabilidade $\alpha = 2$. Na Gaussiana, somente o segundo cumulante não se anula e é dado por $k_2(t) = 2s_0^2(1 - e^{-2\gamma t})$.

Covariância

A covariância pode ser computada diretamente como

$$\langle\langle u(t)u(t+\tau) \rangle\rangle = \int du du' f_t(u') G_\tau(u|u') [u - c_1(t+\tau)] [u' - c_1(t)]. \quad (8.47)$$

Calculamos primeiro a integral sobre u , já que ela fornece a média condicional de $u(t+\tau)$ que pode ser relacionada com a Eq. (8.27). Antes disto, observamos que segue da Eq. (8.35) que

$$c_1(t+\tau) = c_1(t)e^{-\gamma\tau} + \kappa(1 - e^{-\gamma\tau}). \quad (8.48)$$

Comparando com a Eq. (8.26), vemos que

$$\int du G_\tau(u|u') [u - c_1(t+\tau)] = [u' - c_1(t)] e^{-\gamma\tau} \quad (8.49)$$

e portanto

$$\langle\langle u(t)u(t+\tau) \rangle\rangle = c_2(t)e^{-\gamma\tau}, \quad (8.50)$$

onde $c_2(t)$ definido na equação anterior é a variância de $u(t)$.

Podemos reescrever a covariância $\langle\langle u(t)u(t') \rangle\rangle$ fazendo as identificações $\tau \mapsto$

$|t - t'|$ e $t \mapsto \min(t, t')$. Deste modo, obtemos o resultado principal

$$\langle\langle u(t)u(t') \rangle\rangle = \begin{cases} [\sigma^2 + (c_2 - \sigma^2) e^{-2\gamma \min(t, t')}] e^{-\gamma|t-t'|} & \gamma \neq 0 \\ 2D \min(t, t') & \gamma = 0 \end{cases}, \quad (8.51)$$

onde introduzimos a notação mais sugestiva $\sigma^2 = 2s_0^2$ e $2D = \lambda\mu_2$. Apresentamos novamente as condições sobre as médias por conveniência

$$\langle u(t) \rangle = \begin{cases} \kappa + (c_1 - \kappa)e^{-\gamma t} & \gamma \neq 0 \\ c_1 + Vt & \gamma = 0 \end{cases}, \quad (8.52)$$

onde o segundo termo decorre do limite $\lim_{\gamma \rightarrow 0} \kappa(1 - e^{-\gamma t}) = \kappa\gamma = V$.

Somente o caso $\gamma > 0$ pode definir um processo estacionário como aparece no teorema de Doob original. Lembramos que a única possibilidade consistente com esta escolha é que a probabilidade de transição seja uma distribuição estável. Para que possua covariância finita, somente a Gaussiana. Lembramos que, no caso de um processo estacionário, $c_2 = \sigma^2$ e $c_1 = \kappa$.

Referências Bibliográficas

- [1] Nassar H. Abdel-All, H. N. Abd-Ellah, and H. M. Moustafa. Information geometry and statistical manifold. *Chaos, Solitons and Fractals*, 15:161–172, 2003. 8
- [2] Shun-ichi Amari. Information geometry on hierarchy of probability distributions. *IEEE Transactions on Information Theory*, 47(5):1701–1711, jun 2001. 8
- [3] E. Barkai, R. Metzler, and J. Klafter. From continuous time random walks to the fractional fokker-planck equation. *Phys. Rev. E*, 61(1):132–138, Jan 2000. 3.2
- [4] N H Barton and I Wilson. Genealogies and geography. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 349(1327):49–59, 1995. 4.3
- [5] William Bialek, Curtis G. Callan, and Steven P. Strong. Field theories for learning probability distributions. *Phys. Rev. Lett.*, 77(23):4693–4697, Dec 1996. 7.4
- [6] Richard C. Bradley. Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys*, 2:107, 2005. 4.2, 4.2.2
- [7] Herbert B. Callen and Theodore A. Welton. Irreversibility and generalized noise. *Phys. Rev.*, 83(1):34–40, Jul 1951. 6

- [8] P. K. Cheah, D. A. S. Fraser, and N. Reid. Some alternatives to edgeworth. *The Canadian Journal of Statistics*, 21(2):131–138, 1993. 5.2.5
- [9] Emilio Cortés, Bruce J. West, and Katja Lindenberg. On the generalized langevin equation: Classical and quantum mechanical. *J. Chem. Phys.*, 82(6):2708–2717, 1985. 1.1, 6.2
- [10] Richard T. Cox. *The Algebra of Probable Inference*. The Johns Hopkins Press, 1961. 1.3.2
- [11] Ronald Dickman and Ronaldo Vidigal. Path integrals and perturbation theory for stochastic processes. *Brazilian Journal of Physics*, 33:73–93, May 2003. 2
- [12] M. Doi. Second quantization representation for classical many-particle system. *Journal of Physics A: Mathematical and General*, 9(9):1465–1477, 1976. 2
- [13] J. L. Doob. The brownian movement and stochastic equations. *The Annals of Mathematics*, 43(2):351–369, 1942. 8
- [14] Albert Einstein. *Investigations on the Theory of the Brownian Movement*. Number 1. Dover, 1956. 4.3, 5
- [15] B. U. Felderhof. On the derivation of the fluctuation-dissipation theorem. *Journal of Physics A: Mathematical and General*, 11(5):921–927, 1978. 6, 6.1.2, 6.4
- [16] William Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. Wiley, New York, 1970. 1.1, 3, 3, 3.2.2, 3.3
- [17] L Ferrari. Particles dispersed in a dilute gas: Limits of validity of the langevin equation. *Chemical Physics*, 336(1):27, 2007. 1.1, 5

- [18] Daniel Fink. A compendium of conjugate priors. [http://www.people.cornell.edu/pages/df36/CONJINTRnew TEX.pdf](http://www.people.cornell.edu/pages/df36/CONJINTRnew%20TEX.pdf), mar 1997. 8
- [19] B. Roy Frieden. *Science from Fisher Information, A Unification*. Cambridge University Press, 2004. 8
- [20] D Görlich and U Kutay. Transport between the cell nucleus and the cytoplasm. *Annu. Rev. Cell Dev. Biol.*, 15:607–60, 1999. 4.3
- [21] R. J. Henery. The generalized langevin equation and the fluctuation-dissipation theorems. *Journal of Physics A: General Physics*, 4(5):685–694, 1971. 6, 6.1.2
- [22] Timothy E. Holy. Analysis of data from continuous probability distributions. *Phys. Rev. Lett.*, 79(19):3545–3548, Nov 1997. 7.4
- [23] Katharine L. C. Hunt and John Ross. Path integral solutions of stochastic equations for nonlinear irreversible processes: The uniqueness of the thermodynamic lagrangian. *The Journal of Chemical Physics*, 75(2):976–984, 1981. 2, 2.3.3
- [24] E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–630, 13 may 1957. 1.3.1, 1.3.4
- [25] E. T. Jaynes. Information theory and statistical mechanics. ii. *Physical Review*, 108(2):171–191, 15 oct 1957. 1.3.4
- [26] E. T. Jaynes. *Probability Theory, the Logic of Science*. Cambridge University Press, 2003. 1.2, 8
- [27] Petr Jizba and Hagen Kleinert. Superpositions of probability distributions. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 78(3):031122, 2008. 1.1, 2, 2

- [28] W. T. Grandy Jr. Principle of maximum entropy and irreversible process. *Physics Reports*, 63(3):175–266, jun 1980. 1
- [29] M. Kac. *Probability and Related Topics in Physical Sciences*. Interscience, New York, 1959. 2
- [30] A. Kamińska and T. Srokowski. Mean first passage time for a markovian jumping process. *Acta Physica Polonica B*, 38:3119, oct 2007. 3.2.2
- [31] N. G. van Kampen. Itô versus stratonovich. *Journal of Statistical Physics*, 24(1):175–187, 8 jan 1981. 2, 5.3
- [32] N. G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland, Amsterdam, 1981. 1.2, 1.3.4, 1.4.2, 2, 5.3
- [33] A. I. Khinchin. *Mathematical Foundations of Statistical Mechanics*. Dover, New York, 1949. 1.1, 4, 4.2
- [34] K. Kitahara and H. Metiu. On the path integral representation of stochastic processes. *Journal of Statistical Physics*, 15:141–147, aug 1976. 2
- [35] Hagen Kleinert, Axel Pelster, and Mihai V. Putz. Variational perturbation theory for markov processes. *Phys. Rev. E*, 65(6):066128, Jun 2002. 2
- [36] S. C. Kou and X. Sunney Xie. Generalized langevin equation with fractional gaussian noise: Subdiffusion within a single protein molecule. *Phys. Rev. Lett.*, 93(18):180603, Oct 2004. 4.3
- [37] H. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, April 1940. 4.3, 5.2.5
- [38] R. Kubo. The fluctuation-dissipation theorem. *Reports on Progress in Physics*, 29(1):255–284, 1966. 1.1, 6, 6.4

- [39] R. Kubo, M. Toda, and N. Hashitsume. *Statistical Physics II, Nonequilibrium Statistical Mechanics*. Springer Series in Solid-State Sciences. Springer-Verlag, Berlin, 1978. 1.4.1
- [40] Luciano C. Lapas, Rafael Morgado, Mendeli H. Vainstein, J. Miguel Rubí, and Fernando A. Oliveira. Khinchin theorem and anomalous diffusion. *Physical Review Letters*, 101(23):230602, 2008. 4.2, 4.4.1
- [41] M. Howard Lee. Generalized langevin equation and recurrence relations. *Phys. Rev. E*, 62(2):1769–1772, Aug 2000. 6
- [42] M. Howard Lee. Why irreversibility is not a sufficient condition for ergodicity. *Physical Review Letters*, 98(19):190601, 2007. 1.1, 4, 4.2, 4.4.1
- [43] Don S. Lemons and Anthony Gythiel. Paul langevin’s 1908 paper “on the theory of brownian motion” [“sur la th[e-acute]orie du mouvement brownien,” c. r. acad. sci. (paris) [bold 146], 530–533 (1908)]. *American Journal of Physics*, 65(11):1079–1081, 1997. 5
- [44] P. Lévy. *Théorie de L’Addition des Variables Aléatoires*. Gauthier-Villars, Paris, 1954. 2.2.1
- [45] Eugene Lukacs. *Characteristic Functions*. Griffin, London, 2 edition, 1970. 1.1, 2, 4.1.2, 4.2.3
- [46] Daniel C. Mattis and M. Lawrence Glasser. The uses of quantum field theory in diffusion-limited reactions. *Rev. Mod. Phys.*, 70(3):979, Jul 1998. 2
- [47] Ralf Metzler and Joseph Klafter. The random walk’s guide to anomalous diffusion: a fractional dynamics approach. *Physics Reports*, 339(1):1–77, December 2000. 4.3

- [48] Elliott W. Montroll. Random walks on lattices. ii. *Journal of Mathematical Physics*, 6(2):167, 1965. 4.3, 4.3
- [49] Rafael Morgado, Fernando A. Oliveira, G. George Batrouni, and Alex Hansen. Relation between anomalous and normal diffusion in systems with memory. *Phys. Rev. Lett.*, 89(10):100601, Aug 2002. 1.1, 4, 4.3.1, 4.4.2
- [50] Kent E. Morrison. Random walks with decreasing steps. July 1998. 4.2.3
- [51] Roger B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer, Berlin, 2 edition, 2006. 1.4.1
- [52] H. Nyquist. Thermal agitation of electric charge in conductors. *Phys. Rev.*, 32(1):110–113, Jul 1928. 6
- [53] Ben O’Shaughnessy and Itamar Procaccia. Analytical solutions for diffusion on fractal objects. *Phys. Rev. Lett.*, 54(5):455–458, Feb 1985. 3.2.2, 4.3
- [54] Abraham Pais. *"Sutil é o Senhor...": a ciência e vida de Albert Einstein*. Nova Fronteira, Rio de Janeiro, 1995. 13
- [55] L. Peliti. Path integral approach to birth-death processes on a lattice. *J. Phys. (Paris)*, 46:1469, mar 1985. 2
- [56] Vasiliki Plerou, Parameswaran Gopikrishnan, and H Eugene Stanley. Econophysics: Two-phase behaviour of financial markets. *Nature*, 421(6919):130, 2003. 4.3
- [57] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, Massachusetts, 2006. 1.1, 4.1.2, 7.2.3, 7.1, 7.3

- [58] L. S. Schulman. *Techniques and Applications of Path Integration*. Wiley, New York. 1981. 2.2
- [59] Cosma Shalizi. *Almost none of the theory of stochastic processes*. 2007. 10
- [60] J. Skilling. American institute of physics conference series. In A. Mohammad-Djafari, editor, *Bayesian Inference and Maximum Entropy Methods In Science and Engineering*, volume 872, pages 321–330, nov 2006. 1.1, 7.3.2
- [61] T. Srokowski and A. Kamińska. Stochastic equation for a jumping process with long-time correlations. *Phys. Rev. E*, 70(5):051102, Nov 2004. 3.2.2
- [62] H. E. Stanley, L. A. N. Amaral, D. Canning, P. Gopikrishnan, Y. Lee, and Y. Liu. Econophysics: Can physicists contribute to the science of economics? *Physica A: Statistical Mechanics and its Applications*, 269(1):156–169, 1999. 4.3
- [63] John S. Toll. Causality and the dispersion relation: Logical foundations. *Phys. Rev.*, 104(6):1760–1770, Dec 1956. 6.3
- [64] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the brownian motion. *Phys. Rev.*, 36(5):823–841, Sep 1930. 5
- [65] A Uvarov. A semi-phenomenological approach to the transport and diffusion of small spherical macromolecules in solution. *Chemical Physics Letters*, 401(1-3):296, 2005. 5, 5.1.1
- [66] Eric W Weisstein. Devil’s staircase. MathWorld. 4.2.3
- [67] V. V. Yanovsky, A. V. Chechkin, D. Schertzer, and A. V. Tur. Lévy anomalous diffusion and fractional fokker-planck equation. *Physica A: Statistical Mechanics and its Applications*, 282(1-2):13–34, 2000. 2

- [68] Huaiyu Zhu and Richard Rohwer. Information geometry, bayesian inference, ideal estimates and error decomposition. 8

Índice Remissivo

- álgebra sigma, 18
- ação, 40
- ação, processo estocástico, 184
- coeficiente de difusão, 44, 93
- coeficiente de drift, 44
- complemento de Schur, 159
- Condição de compatibilidade de Chapman, 27
- condição de mistura, 85
- covariância, 72
- De Finetti, teorema de, 42
- densidade espectral, 79
- diagramas de Venn, 17
- difusão, 92
- difusão anômala, 94
- distribuição contínua, 90
- distribuição descontínua, 90
- distribuição singular, 90
- E. T. Jaynes, 15
- Einstein, 104
- Ensemble, 24
- equação de Chapman-Kolmogorov, 33
- equação de Langevin generalizada, 135
- equação de Schrödinger, 47
- equação de Smoluchowski, 33
- Equação Mestra, 35
- ergodicidade, 85
- evidência, 210
- fórmula de Einstein, 93, 105
- fator de Bayes, 210
- fatores de Bayes, 168
- Fermat, 14
- flutuação-dissipação, 116
- Fokker-Planck, 48
- força de arrasto inercial, 107
- força de Stokes, 107
- função de Bessel, 200
- função de sobrevivência, 37, 59, 62
- função hipergeométrica, 201
- gás de esferas rígidas, 117
- Gaussiana multidimensional, 155
- Gaussiana, função característica, 156
- Girolamo Cardano, 14
- Hamiltoniana, 40, 42, 47

infinitamente divisível, 42
 Integrais de trajetória, 38
 integrais do seno e cosseno, 201
 integral de Itô, 41
 integral de Riemann, 38
 Jakob Bernoulli, 14
 Lagrangiana, 50
 Langevin, 104
 lei de equipartição, 123
 lei de van't Hoff, 110
 métrica de Fisher, 194, 207
 mínima ação, 51
 malha temporal, 26
 marginalização, 22
 matriz positivo-definida, 75
 momentos centralizados, 24
 número de Reynolds, 107
 normalidade assintótica, 166
 normalização, 22
 operador positivo-definido, 76
 operadores de momentum e posição,
 47
 partícula Browniana, 106
 partícula de Rayleigh, 106
 partícula Stokes, 106
 Pascal, 14
 priori conjugado, 204
 priori geométrico, 194
 priori volumétrico, 206
 probabilidade a N -pontos, 27
 probabilidade *a priori*, 165
 probabilidade preditiva, 167, 208
 processo autônomo, 55
 processo autônomo, operadores, 66
 processo de saltos, 58
 processo de saltos aditivo, 119
 processo estocástico, 26
 processo estocástico, definição de Kol-
 mogorov, 28
 processo homogêneo, 56
 regra de Bayes, 20, 165, 203
 regressão não-paramétrica, 170
 regressão paramétrica, 162
 relações de Kramers-Kronig, 150
 representação de Lévy-Khinchin, 43
 ruído branco, 113
 tempo estocástico, 192
 teorema de Bochner, 77
 teorema de Cox, 19
 Teorema de Doob, 212
 teorema de Mercer, 78
 teorema de Pawula, 50
 teorema espectral, 78
 teoria de resposta linear, 134

variância, 24

variação quadrática, 191