

Teoria: Sistemas de Processamento de Informação

Nestor Caticha

23 de novembro de 2015

Sumário

1	Teoremas de Regraduação de Cox	5
1.1	Probabilidades e Informação	5
1.2	Teoremas de Cox	6
1.2.1	Axiomas de Cox	8
1.2.2	A regra da soma	11
1.2.3	Regra do produto: quais as variáveis relevantes?	13
1.2.4	Regra do produto: qual é a função G ?	15
1.2.5	O teorema de Bayes e Informação Incompleta	18
1.2.6	Jaynes e o bom senso	21
1.2.7	Exemplo do Teorema de Bayes e Ajuste de funções	22
1.2.8	Obtendo a posterior	24
2	Probabilidades e o Teorema do Limite Central	27
2.1	Introdução: Kolmogorov e as probabilidades	27
2.2	Momentos...	28
2.2.1	Marginais, Independência	30
2.3	Convoluções e Cumulantes	30
2.4	O Teorema do Limite Central I	36
2.4.1	Aquecimento	36
2.5	A média e a concentração em torno da média	37
2.6	O Teorema do Limite Central II	39
2.7	O Teorema do Limite Central III	40
2.7.1	A distribuição uniforme	41
2.7.2	A distribuição exponencial	41
2.7.3	A distribuição binomial revisitada	42
2.7.4	Caminho Aleatório	45
2.8	Apêndice: Stirling	46
3	Entropia	49
3.1	Simetria	50
3.2	Incerteza, ignorância e Informação	51
3.3	Exemplos	57
3.3.1	O problema variacional: variáveis discretas	57
3.3.2	Exemplo: Variável de dois estados	58

3.3.3	3 estados	59
3.3.4	Variáveis reais: a gaussiana	59
3.4	Entropia e convexidade	62
3.4.1	Independência	63
3.4.2	Exercício sobre independência: Variáveis Gaussianas cor- relacionadas	64
3.5	Caso Geral para informação sobre valores esperados	65
3.6	Conexão com a Termodinâmica: estrutura matemática	66
3.6.1	Transformações de Legendre	66
3.6.2	Relações de Maxwell	67
3.7	Conexão com a Termodinâmica: Micro e Macroestados	68
3.8	Conexão com a Termodinâmica: Física	71
3.8.1	O Ensemble Microcanônico	71
3.9	Conexão com a Termodinâmica: T, P, μ	73
3.9.1	Temperatura	73
3.9.2	Pressão	75
3.9.3	Potencial químico	76
3.10	Conexão com a Termodinâmica	76
3.11	Ensemble Canônico	79
3.11.1	Comparação da Entropia nos diferentes Ensembles	80
3.12	Apêndice: Multiplicadores de Lagrange	81
4	Aplicações Simples	83
4.1	Sistemas Paramagnéticos	83
4.1.1	Ensemble Microcanônico: Energia fixa	84
4.1.2	Ensemble Canônico: valor esperado da Energia fixo	85
4.2	Gás Ideal: Microcanônico	86
4.3	Gás ideal: Canônico	89
4.4	Mistura de gases ideais clássicos	91
4.4.1	Reações e potenciais químicos	92
4.5	Sólido Clássico	93
4.6	Sólido de Einstein	94
4.6.1	Microcanônico	94
4.6.2	Canônico	96
4.6.3	Gases ideais clássicos com estrutura interna quântica	96
4.7	Ensemble Grande Canônico	97
4.8	Gases Quânticos	99
4.9	Densidade de Estados	104
4.10	Gás de Fermi-Dirac	106
4.10.1	Gás ideal de Férmions a temperatura finita	108
4.11	Bósons	112
4.11.1	Radiação de corpo negro	112
4.11.2	Condensação de Bose-Einstein	114
4.11.3	Calor Específico	117

5	Aplicações: Modelo de Ising e Campo Médio	119
5.1	O modelo de Ising	121
5.2	Modelo de Ising em uma dimensão	122
5.2.1	Matriz de Transferência	122
5.3	Campo Médio: uma interpretação geométrica	126
5.3.1	Campo médio: Ising spin $\frac{1}{2}$	127
5.3.2	Campo médio: Blume-Capel spin 1	131
5.4	O Modelo de alcance infinito: Ising spin $\frac{1}{2}$	134
5.4.1	Método de Ponto de sela	134
5.4.2	Cálculo rigoroso para o Modelo de Ising com interações de alcance infinito	137
5.4.3	Método de ponto de sela para o modelo de Blume Capel de spin 1	140
5.5	Modelo de Ising 1d revisitado: O Grupo de Renormalização	141
5.6	Argumentos de Energia- Entropia	145
5.7	Escalas de Tempo	145
5.8	Vidro de Spin	146
5.8.1	Máxima Entropia	147
5.8.2	Método das Réplicas	149
5.8.3	Integração pelo ponto de sela	151
5.8.4	Simetria de réplicas	153
5.8.5	Diagrama de fases	155
5.8.6	Entropia negativa	156
6	Modelo de Memória associativa: Hopfield	159
6.1	Memória associativa	159
6.1.1	Hebb	161
6.1.2	Modelo de Hopfield com número finito de padrões: $\alpha = 0$	161
6.1.3	Estrutura das soluções	163
6.2	O Modelo de Hopfield: Memória Associativa: Número de padrões extensivo	164
7	Monte Carlo	169
7.1	Integração Numérica em espaços de alta dimensão	169
7.2	Monte Carlo	170
7.2.1	Teorema Central do Limite: revisitado	170
7.2.2	Monte Carlo	171
7.2.3	Exemplos analíticos.	173
7.2.4	Métodos Estáticos: rejeição	176
7.2.5	Círculo	176
7.2.6	Métodos Dinâmicos	177
7.2.7	Algoritmo de Metropolis	180

Capítulo 1

Teoremas de Regradação de Cox

1.1 Probabilidades e Informação

Estas notas para o curso “Mecânica Estatística de Sistemas de Processamento de Informação” estão e parece sempre estarão em estado embrionário e preliminar. Eventuais críticas, correções ou sugestões serão bemvindas. Algumas partes são baseadas nos livros de E. Jaynes ¹ e de A. Caticha ² que estas notas não substituem. Outras não, grande parte são baseadas em conjunto de notas de uma versão anterior deste curso.

A primeira parte discute o uso de probabilidades. O leitor, aluno deste curso não deverá ler isto como uma validação das suas crenças. Afinal, se está estudando Mecânica Estatística já parece natural o uso de probabilidades. Deve procurar falhas no raciocínio. Procurar exceções.

A idéia de apresentar uma forma de pensar que tem aplicações em uma vasta gama de assuntos, pode levar o leitor a pensar que está na presença de alguém que com um martelo, pensa que todos os problemas são pregos. Ou que estamos apresentado dogmas, dos quais não abriremos mão. No fim talvez não saiba como me defender de tais acusações, exceto alegando que o único ponto sobre o qual serei inflexível será que só podemos acreditar naquilo que a informação e evidência permitem, e só enquanto não surgir informação contraditória. Não faz sentido acreditar em algo que não é respaldado por informação. ³

Estas notas servem dois propósitos. Um como material de suporte para um curso de Mecânica Estatística, outro para Física de processamento de informação.

¹E. T. Jaynes, Probability Theory: the Logic of Science

²Entropic Inference and the Foundations of Physics

³Há outras formas de pensar, por exemplo acreditar em algo porque isso me deixa mais feliz. Mas eu não saberia dar um curso sobre isso. "I have a lot of beliefs, and I live by none of them - that's just the way I am. . . they make me feel good about who I am." –Louis CK

1.2 Teoremas de Cox

Há muitas definições matemáticas possíveis que poderiam ser usadas na tentativa de formalizar o conceito coloquial de informação. Uma forma de avançar, que é bastante comum em ciência, começa por definir matematicamente algo e depois tentar interpretar as fórmulas matemáticas para mostrar que esta interpretação esta de acordo com algumas das características que podemos atribuir ao conceito coloquial de informação que temos.

Em lugar de começar por uma estrutura matemática pré-escolhida para servir de ferramenta de análise, começamos por uma interpretação e depois encontramos a estrutura matemática que se adapte à interpretação. A interpretação passa por estabelecer em alguns casos particulares suficientemente simples, tais que haja algum tipo de consenso, o que deveria resultar da teoria. É possível que este procedimento pareça novo ao leitor e será surpreendente quantos resultados serão extraídos deste método e do rigor matemático que a teoria se vestirá. Como este procedimento permite saber mais claramente do que estamos falando e do que não estamos, achamos que esta é atualmente a melhor maneira de introduzir a teoria de informação.⁴

Pode parecer estranho para o estudante Física que o elemento principal a seguir seja a idéia de asserção, isto é, uma frase que em princípio é uma proposição que se apresenta como verdadeira. Denotaremos asserções por letras $A, B, C \dots a, b, c \dots$; Uma frase pode ser julgada correta ou não de várias maneiras. Podemos pensar se é correta do ponto de vista da sua estrutura gramatical ou sintática. Não é isto que queremos fazer e consideraremos as asserções a seguir suficientemente bem formadas. Queremos analisar seu conteúdo informacional, se realmente a podemos creer verdadeira. Mas quando se diz " a massa de Saturno está entre m_1 e m_2 " ou "... entre m_3 e m_4 " estamos usando asserções diferentes e a tarefa é determinar quanto acreditamos que uma ou a outra sejam verdade e aqui o estudante reconhece a linguagem científica.

Consideremos a asserção "Existem zumbies". Isto é verdade? Se o contexto for o de filmes gravados em Pittsburgh na década de oitenta, a resposta será uma. Se for no mundo real, outra. Nenhuma asserção sozinha pode ser analisada, no que diz respeito a se é verdadeira ou não, de forma independente do resto do universo conceitual. Ela será julgada verdadeira ou não quando analisada dentro de um contexto. A informação trazida por uma asserção C , será usada para atribuir um grau de verdade à asserção A , ou seja dentro do contexto C . Poderíamos chamar esse grau de, por exemplo, probabilidade de que A seja verdade se C for dada. Mas fazendo isto estaríamos definindo de antemão que a ferramenta matemática apropriada para descrever informação é a teoria de probabilidades. Isto parece bem razoável mas não escapa às críticas acima e permite que outra ferramenta matemática seja usada por simplesmente expressar o gosto de outras pessoas ou a facilidade de uso em determinados problemas práticos com a mesma justificativa: *parece razoável, eu gosto, funciona,*

⁴Também ocorrerá que os resultados não serão universalmente satisfatórios, pois há lugar a discussões sobre o tipo de interpretação *a priori* que será imposta. Ver [?] possíveis extensões e críticas leves que talvez não sejam tão relevantes

é *prático*. Não descartamos o uso de outras ferramentas matemáticas, mas queremos deixar claro que estas poderão ser vistas como aproximações mais ou menos adequadas de uma estrutura que unifica e tem um posição diferente. O **objetivo** deste capítulo é mostrar que a escolha da teoria de probabilidades como a ferramenta matemática adequada para tratar informação é muito mais do que simplesmente conveniente. Isto nos levará à teoria de inferência, baseada na teoria de probabilidades, que tem **exatamente** a estrutura da Mecânica Estatística dos pioneiros Boltzmann e Gibbs. O “exatamente” não é coincidência. Somos levados a repensar a Mecânica Estatística como uma teoria de inferência, mas muito mais sobre isto será dito adiante. Antes disso há muito o que fazer.

Se a informação em C não permite a certeza sobre a verdade de A então diremos que a crença que temos sobre A esta baseada em informação incompleta. Em casos particulares poderá ocorrer que dado C possa ser concluído, com certeza que a asserção A é verdadeira ou ainda em outros casos que é falsa. Quando não há alternativa para a conclusão, quando ela segue por força da informação disponível, dizemos que a conclusão é racional ou lógica. Dizemos que estamos frente a casos de raciocínio dedutivo. Nestes casos a informação disponível é *completa* pois nada falta para ter certeza. A análise destes casos remonta a Aristóteles.

Exemplos de informação completa são dados pelos silogismos Aristotélicos: suponha que recebemos a informação contida em $C = “A \rightarrow B”$, isto é, A implica B . Traduzindo, isto significa “se souber que A é certamente verdade, segue que a proposição B também o é.” Dado isso, o que podemos dizer sobre B ? Nada com certeza, mas se também recebemos a informação adicional A , isto é, que “ A é Verdade”, então segue B , ou seja “ B é Verdade”.

Outro caso de informação completa é \bar{B} ou seja “ B é Falso”, então segue \bar{A} , isto é, que “ A é Falso”.

Nas condições que $C = “A \rightarrow B”$ e “ A é Falso”, o quê pode ser concluído? Do ponto de vista lógico clássico nada podemos concluir sobre B . Da mesma forma se for dada a informação “ B é Verdade”, nada podemos concluir sobre A . Estamos frente a casos de informação incompleta e a lógica clássica não serve para chegar a uma conclusão. Não é possível deduzir nada. A indução,⁵ o que quer que isto seja, e que será discutido mais à frente, será necessária para avançar. A forma dedutiva da lógica permite somente tres tipos de respostas, *sim*, *não* e *não segue*. A indução nos força ou permite dividir esta última em várias possibilidades e os casos extremos nesse espectro são aqueles onde

⁵Segundo Harold Jeffreys em seu livro *Theory of Probability*, Bertrand Russell disse que “induction is either disguised deduction or a mere method of making plausible guesses”. Jeffreys diz que “é muito melhor trocar a ordem dos dois termos e que muito do que normalmente passa por dedução é indução disfarçada, e que até alguns dos postulados de *Principia Mathematica* foram adotados por motivações indutivas” (e adiciona, são falsos). Com o tempo o próprio Russell mudou de posição, dobrado pela evidência (?) e diz no fim da sua autobiografia: “I was troubled by scepticism and unwillingly forced to the conclusion that most of what passes for knowledge is open to reasonable doubt”. Sobre indução disse ainda: “The general principles of science, such as the belief of the reign of law, and the belief that every event must have a cause, are as completely dependent on the inductive principle as are the beliefs of daily life.” (On Induction)

havendo certeza absoluta, haverá portanto a força da dedução. Podemos falar então sobre quais das alternativas intermediárias é mais razoável acreditar com base no que sabemos. Nota-se então a necessidade de estender a lógica para poder tratar de forma racional casos de informação incompleta. Richard T. Cox, ao se defrontar com este problema por volta da década de 1940, decidiu, como dito acima, estabelecer um conjunto de desejos (*desiderata* [?]) que a teoria deveria satisfazer, e estes serão então os axiomas da extensão da lógica. Aqui podemos discordar, propor outros axiomas, mas uma vez aceitos serão provados os teoremas de reparametrização de Cox que mostram que a teoria de probabilidade é a ferramenta para o tratamento de forma racional de situações de informação incompleta. O surpreendente disto é que surge a teoria das probabilidades como a forma para lidar de forma *racional*⁶ com a informação e que corremos riscos de ser inconsistentes caso a regras de manipulação de probabilidades não sejam seguidas. Segue que não há probabilidades que não sejam condicionais embora às vezes simplesmente a linguagem esqueça de deixar explícitas as relações de condicionalidade. A amplidão da aplicabilidade da teoria que emerge é impressionante e por exemplo, quando o tipo de asserção for limitado àqueles entendidos em teoria de conjuntos as regras de manipulação serão não mais nem menos que aquelas dadas pelos axiomas de Kolmogorov. Veremos que emerge uma relação natural entre probabilidade e frequência e ficará claro de que forma estes conceitos estão ligados e mais importante, de que forma são distintos.

1.2.1 Axiomas de Cox

É interessante notar que os axiomas de Cox descritos por Jaynes não são exatamente iguais aos que Cox apresenta no seu livro *The algebra of probable inference*. A exposição de Jaynes é muito mais simples. Cox, por sua vez, esclarece sua dívida com J. M. Keynes e seu livro *A treatise on Probability*, que deve muito a Laplace e Bernuolli. A exposição de Jaynes teve uma grande influencia, mas ainda recebeu críticas e complementos [?]. Eu seguirei a apresentação de A. Caticha, que é mais completa.⁷

A maneira de construir a teoria está baseada na seguinte forma de pensar bastante simples. Queremos construir uma teoria geral para a extensão da lógica nos casos de informação incompleta. Se ela for suficientemente geral, deverá ser válida em casos particulares. Se o caso for suficientemente simples, então podemos saber qual é o resultado esperado que não viole expectativas razoáveis. Poderia ocorrer que ao analisar um número de casos particulares sejam reveladas as inconsistências entre eles, nesse caso não poderemos chegar a uma teoria geral. Mas pode ser que os casos particulares sirvam para restringir e determinar a teoria geral.⁸ Isto é o que mostraremos a seguir.

⁶Ao leitor que demande uma definição de racional, podemos dizer que pelo menos não queremos ser manifestamente irracionais e que terá que esperar por uma resposta.

⁷Notem que há lugar ainda para avanços nestes primeiros passos. Tentem encontrar defeitos, generalizações, melhorias nos argumentos.

⁸Este comentário parece trivial, mas o uso que será dado a seguir é totalmente não trivial.

Em primeiro lugar queremos falar sobre uma asserção A no caso de informação incompleta. Nos referimos então à crença ou plausibilidade de A ser verdade dado B e a denotamos pelo símbolo $A|B$ que lemos "a plausibilidade A dado B " ou ainda de "... de A condicionada a B ". Por que não à probabilidade de A dado B ? Porque já existe uma teoria matemática de probabilidade e não sabemos se será a estrutura matemática que emergirá desta análise. Poderíamos usar outras palavras, mas crença ou plausibilidade são conhecidas o suficiente para serem úteis neste contexto e não tem por agora o problema de ser definidas formalmente .

Queremos analisar o primeiro caso simples que lida com o conceito de *mais plausível*. Se A dado B é mais plausível do que A dado C escrevemos $A|B \succ A|C$. Suponha ainda que é ainda que $A|C \succ A|D$. Queremos, para seguir o uso cotidiano da linguagem impor que A dado B seja mais plausível que A dado D

Temos assim nosso primeiro desejo, a plausibilidade deverá satisfazer alguma forma de transitividade:

- D_1 : Se $A|B \succ A|C$ e $A|C \succ A|D$ então deve ser o caso que $A|B \succ A|D$

Isto é fácil se impusermos:

- A plausibilidade $A|B$ deverá ser representada por um número real.

Dados

$$A|B > A|C$$

e

$$A|C > A|D,$$

segue imediatamente, uma vez que são números reais, que

$$A|B > A|D,$$

de acordo com o axioma 1. Note que dizer que alguma coisa é um número real nos dá imediatamente a transitividade, mas não diz nada sobre que número deve ser atribuído, nem sobre como mudá-lo se a informação passa de B para C .

Através de certas operações e de diferentes asserções podemos criar asserções compostas. Exemplos de operadores são a negação, o produto e a soma lógicos.

- A **negação** de A é denotada por \bar{A} .
- O **produto** ou conjunção de duas asserções é uma terceira asserção, há diferentes notações equivalentes possíveis: $C = AB$, $C = A \wedge B$ ou ainda $C = A \text{ e } B$.
- A **soma** ou disjunção de duas asserções é uma terceira asserção, que costuma ser denotada por $D = A + B$ ou $D = A \vee B$, ou ainda $D = A \text{ ou } B$.

Neste contexto de probabilidades foi colocado primeiro por J.Skilling, mas não de forma explícita. O destaque a este procedimento apareceu por primeira vez no livro de A. Caticha. Usaremos novamente este estilo de fazer teoria ao introduzir o conceito de entropia.

A tabela 1.1 mostra a tabela verdade para as operações de soma e produto lógico, onde $V = \text{Verdade}$ e $F = \text{Falso}$. Note que as últimas duas colunas, colocadas aqui para futura referência, mostram que $A + B$ e $\overline{A\overline{B}}$ são iguais.

A	B	$A + B$	AB	$\overline{A + B}$	$\overline{A\overline{B}}$
V	V	V	V	F	F
V	F	V	F	F	F
F	V	V	F	F	F
F	F	F	F	V	V

Tabela 1.1

Suponha que haja um método, usando a teoria geral que procuramos e ainda não temos, de analisar a plausibilidade de uma asserção composta por várias asserções através de conjunções ou disjunções. Esperamos que a plausibilidade possa ser expressa em termos da plausibilidade de asserções mais simples. Talvez haja mais de uma forma de realizar essa análise. Queremos então que:

- D_2 : Se a plausibilidade de uma asserção puder ser representada de mais de uma maneira, pela plausibilidade de outras asserções, todas as formas deverão dar o mesmo resultado.

Há várias formas de usar a palavra *consistência*. Aqui a usamos da seguinte forma. Impor que duas formas de análise devam dar o mesmo resultado não garante a consistência da teoria geral, no entanto uma teoria onde isso não ocorra será inconsistente. Usamos consistência no sentido de não manifestamente inconsistente, que é o que D_2 acima declara.

Agora olhamos para o caso simples em que a e b são mutuamente exclusivos na condição c e em qualquer outra condição. Então $a|b$ e $b|a$ representam a plausibilidade de algo que sabemos ser falso. Assim como $a|a$ e $b|b$ são a plausibilidade de algo que sabemos ser verdade. Poderia ser que hajam falsidades mais falsas que outras, ou verdades mais verdadeiras que outras, mas achamos razoável impor

- D_3 : Para todo A , $A|A = v_v$ e para A e B mutuamente exclusivos $A|B = v_f$.

Não sabemos que valores dar para v_v ou v_f , mas supomos o mesmo valor em todos os casos que tenhamos certeza de verdade ou falsidade

Todo operador na álgebra Booleana pode ser representado pelas operações conjunção (e) e negação (\neg)⁹, isto é, o produto e a negação lógicas. A soma lógica pode ser obtida usando $A + B = \overline{\overline{A}\overline{B}}$. Precisamos então analisar a

⁹Este conjunto não é mínimo, mas é útil e claro.

plausibilidade de asserções compostas usando esses operadores em termos das plausibilidade de asserções mais simples. Já que este conjunto de operadores é completo, esperamos que só tenhamos que analisar estes dois operadores.

Agora olhamos para a soma lógica. Novamente c se refere à informação subjacente e estamos interessados na plausibilidade $y = a \vee b|c$. Há 4 plausibilidades que serão interessantes para esta análise:

$$x_1 = a|c, x_2 = b|c, x_3 = a|bc, x_4 = b|ac$$

. Notamos que deve haver uma dependência entre $a \vee b|c$ e algum subconjunto de $\{x_i\}$, então

- D_4 : Deve existir uma função F que relaciona $a \vee b|c$ e algum subconjunto de $\{x_i\}$.

É claro que trocando soma por produto parece razoável desejar:

- D_5 : Deve existir uma função G que relaciona $ab|c$ e algum subconjunto de $\{x_i\}$.

Não impomos nada além da existência dessas funções, além de que dependam em algumas, se não todas, as variáveis $\{x_i\}$.

Porque um subconjunto? Qual subconjunto? Todos? Como decidir? Há 11 subconjuntos de dois ou mais membros: Seis $(\frac{4!}{2!2!})$ pares (x_i, x_j) , quatro $(\frac{4!}{3!1!})$ triplas (x_i, x_j, x_k) e o conjunto inteiro (x_1, x_2, x_3, x_4) . Parece difícil que desta lista $D_1 \dots D_5$ surja uma estrutura matemática única.¹⁰ Talvez o que será surpreendente para o leitor, é que seja a teoria de probabilidades.

1.2.2 A regra da soma

Começamos com a função F e consideramos a e b mutuamente exclusivos

$$a \vee b|c = F(a|c, b|c, a|bc, b|ac) = F(a|c, b|c, v_f, v_f)$$

mas esta é uma função de apenas duas variáveis, e da constante desconhecida v_f :

$$a \vee b|c = f(a|c, b|c)$$

Agora consideremos tres asserções a, b e c mutuamente excludentes nas condições d . Duas maneiras equivalentes de escrever a disjunção das tres são $(a \vee b) \vee c|d = a \vee (b \vee c)|d$ o que permite usar a função f

$$\begin{aligned} a \vee b \vee c|d &= f(f(a|d, b|d), c|d) \\ &= f(a|d, f(b|d, c|d)) \end{aligned}$$

ou em notação óbvia

$$f(f(x, y), z) = f(x, f(y, z)) \quad (1.1)$$

¹⁰Como veremos, essencialmente única a menos de regradações montônicas.

chamada equação da associatividade, primeiramente estudada por Abel no contexto de teoria de grupos. Pode se provar ¹¹ que existe um bijeção ϕ , dos reais nos reais, que tomaremos como monotonicamente crescente, tal que

$$f(x, y) = \phi^{-1}(\phi(x) + \phi(y)). \quad (1.2)$$

O leitor deverá neste ponto mostrar que a expressão 1.2 é uma solução da equação 1.1. Agora um ponto importante: podemos *regraduar* as atribuições de plausibilidade e não mais falar dos números do tipo $a|d$ mas de números $\phi(a|d)$. Note que o fato de ser uma bijeção resulta que a ordem de preferências não se altera, se antes as crenças sobre as asserções tinham uma certa ordem, depois da regradação, a representação numérica das crenças não se altera. ¹² Continuamos sem saber que números são esses, mas avançamos a ponto de poder dizer que para quaisquer eventos mutuamente exclusivos

$$\phi(a \vee b|d) = \phi(a|d) + \phi(b|d). \quad (1.3)$$

No caso particular que $d = \bar{a}$, isto significa

$$\phi(a \vee b|\bar{a}) = \phi(a|\bar{a}) + \phi(b|\bar{a}) \quad (1.4)$$

$$\phi(b|\bar{a}) = \phi(a|\bar{a}) + \phi(b|\bar{a}) \quad (1.5)$$

$$(1.6)$$

pois a crença $\phi(a \vee b|\bar{a})$ é equivalente à crença $\phi(b|\bar{a})$. Segue que

$$\phi(a|\bar{a}) = \phi(v_f) = \phi_f = 0 \quad (1.7)$$

Embora modesto, eis o primeiro resultado numérico. **O valor regradado da certeza da falsidade é zero.**

Mas e se não forem mutuamente exclusivos? O interessante é que o resultado anterior serve para o caso geral, mas precisamos usar o truque de escrever

$$a = (a \wedge b) \vee (a \wedge \bar{b}) \quad e \quad b = (b \wedge a) \vee (b \wedge \bar{a})$$

Podemos escrever $a \vee b$ como uma disjunção de asserções mutuamente exclusivas:

$$\begin{aligned} a \vee b &= (a \wedge b) \vee (a \wedge \bar{b}) \vee (b \wedge a) \vee (b \wedge \bar{a}) \\ &= (a \wedge b) \vee (a \wedge \bar{b}) \vee (b \wedge \bar{a}) \end{aligned}$$

assim a equação 1.3 pode ser usada, levando a

$$\begin{aligned} \phi(a \vee b|d) &= \phi(a \wedge b|d) + \phi(a \wedge \bar{b}|d) + \phi(b \wedge \bar{a}|d) \\ &= \phi(a \wedge b|d) + \phi(a \wedge \bar{b}|d) + \phi(b \wedge \bar{a}|d) + \phi(a \wedge b|d) - \phi(a \wedge b|d) \end{aligned}$$

¹¹Aequationes mathematicae 1989, Volume 37, Issue 2-3, pp 306-312 The associativity equation revisited R. Craigen, Z. Páles, ou o livro do Axcel

¹²Note que tomaremos a função ϕ como estritamente monotônica: se $x > y$ segue que $\phi(x) > \phi(y)$, sem poder haver igualdade. Isto é uma decisão pragmática, pois não queremos que asserções com crenças diferentes sejam mapeadas em valores ϕ iguais, pois antes da regradação tínhamos uma separação de preferências e depois da regradação poderíamos ter confusão entre asserções mapeadas no mesmo valor de ϕ . Mas de qualquer forma veja A. Patriota referencia

onde, na última linha adicionamos e subtraímos o mesmo número. Usando novamente a equação 1.3 para asserções mutuamente exclusivas

$$\begin{aligned}\phi(a \vee b|d) &= \phi(a \wedge b \vee a \wedge \bar{b}|d) + \phi(b \wedge \bar{a} \vee a \wedge b|d) - \phi(a \wedge b|d) \\ &= \phi(a|d) + \phi(b|d) - \phi(a \wedge b|d),\end{aligned}\quad (1.8)$$

o que será óbvio uma vez que o leitor provar que $a \wedge b \vee a \wedge \bar{b} = a$ e $b \wedge \bar{a} \vee a \wedge b = b$.

Exercício Desenhe o diagrama de Venn adequado a esta situação.

1.2.3 Regra do produto: quais as variáveis relevantes?

Queremos expressar $y = \phi(ab|c)$ em termos da função ainda por determinar G e de algum dos subconjuntos de $\{x_i\}$ Tribus sugeriu a análise das 11 possibilidades para verificar que só há duas que sobrevivem a casos extremos. Os dois conjuntos são (x_1, x_3) e (x_2, x_4) . Note que se o primeiro deles fosse um dos sobreviventes, o segundo também deveria ser pela simetria trazida pela comutatividade do produto lógico.

Vejamos como chegar a esta conclusão (novamente seguimos AC)

1. $y = G(\phi(a|c), \phi(b|c))$ (1 possibilidade)
2. $y = G(\phi(a|c), \phi(a|bc))$ (2 possibilidades $a \leftrightarrow b$)
3. $y = G(\phi(a|c), \phi(b|ac))$ (2 possibilidades $a \leftrightarrow b$)
4. $y = G(\phi(a|bc), \phi(b|ac))$ (1 possibilidade)
5. $y = G(\phi(a|c), \phi(b|c), \phi(a|bc))$ (2 possibilidades $a \leftrightarrow b$)
6. $y = G(\phi(a|c), \phi(a|bc), \phi(b|ac))$ (2 possibilidades $a \leftrightarrow b$)
7. $y = G(\phi(a|c), \phi(b|c), \phi(a|bc), \phi(b|ac))$ (1 possibilidade)

Caso 1 Mostraremos que $y = a \wedge b|c = G(\phi(a|c), \phi(b|c)) = G(x_1, x_2)$ não funciona pois não satisfaz o esperado em um caso simples. Porque não serve o subconjunto mais óbvio (x_1, x_2) ? Seja $a =$ 'Helena usa um tenis esquerdo vermelho' enquanto que $b =$ 'Helena usa um tenis direito preto'. A plausibilidade dessas duas asserções será julgada dada a seguinte informação $c =$ 'Helena gosta de tenis pretos e de tenis vermelhos', e talvez seja possível concluir que as duas asserções são bastante plausíveis. Mas se tivéssemos $y = G(x_1, x_2)$ poderíamos ser levados a pensar que 'Helena usa um tenis esquerdo vermelho e um tenis direito preto' é bastante plausível. Posso acreditar bastante nas duas asserções, mas não que use um tenis de cada cor. Devemos rejeitar esta forma para G .

Para convencer os incrédulos no exposto acima, um argumento mais formal: Suponha que $a|d = a'|d$ e $b|d = b'|d$, mas que embora a e b sejam mutuamente exclusivos, a' e b' não o sejam. Neste caso teríamos que

$$\phi(a'b'|d) = G(\phi(a'|d), \phi(b'|d)) = G(\phi(a|d), \phi(b|d)) = \phi(ab|d) = \phi_F = 0.$$

E isto ocorreria para qualquer par de asserções não mutuamente exclusivas (a', b') , pois sempre poderíamos supor um caso auxiliar (a, b) adequado.

Caso 2 Se $y = G(\phi(a|c), \phi(a|bc))$ em geral, consideramos o caso particular em que $b = ad$ para qualquer d não seja mutuamente exclusivo a a . Logo

$$\begin{aligned} G(\phi(a|c), \phi(a|bc)) &= G(\phi(a|c), \phi(a|adc)) \\ &= G(\phi(a|c), \phi_v) = g(\phi(a|c)) \end{aligned}$$

segue que

$$\begin{aligned} y &= \phi(ab|c) = \phi(b|c) \\ \phi(b|c) &= g(\phi(a|c)) \end{aligned}$$

onde o lado esquerdo depende de d mas o lado esquerdo não. Esperamos que isso não ocorra em geral, e portanto eliminamos este candidato.

Caso 3 Para o caso $y = G(a|c, b|ac)$ e a alternativa $G(b|c, a|bc)$ ninguém tem encontrado casos que se oponham ao bom senso. Este será o único candidato a sobreviver e será a pedra de sustentação a toda a teoria que segue.

Caso 4 Se $y = G(\phi(a|bc), \phi(b|ac))$ somos levados a algo inaceitável considerando $a = b$, pois seguiria que

$$\phi(ab|c) = \phi(a|c) = G(\phi(a|ac), \phi(a|ac)) = G(\phi_v, \phi_v)$$

e $\phi(a|c)$ seria constante independente de a .

Caso 5 $y = G(\phi(a|c), \phi(b|c), \phi(a|bc))$. Este caso é mais complicado de analisar. Mostraremos, no entanto que se reduz a algum dos casos anteriores, sob a hipótese razoável de diferenciabilidade de G com respeito a qualquer um dos seus argumentos. Ainda consideraremos a conjunção de mais de duas asserções, $abc|d$, que pode ser escrito de duas formas diferentes $(ab)c|d = a(bc)|d$, portanto, considerando a primeira forma obtemos

$$\begin{aligned} \phi((ab)c|d) &= G(\phi(ab|d), \phi(c|d), \phi(ab|cd)) \\ &= G(G(\phi(a|d), \phi(b|d), \phi(a|bd)), \phi(c|d), G(\phi(a|cd), \phi(b|cd), \phi(a|bcd))) \\ &= G(G(x, y, z), u, G(v, w, s)) \end{aligned} \tag{1.9}$$

$$\begin{aligned} \phi(a(bc)|d) &= G(\phi(a|d), \phi(bc|d), \phi(a|bcd)) \\ &= G(\phi(a|d), G(\phi(b|d), \phi(c|d), \phi(b|cd), \phi(a|bcd))) \\ &= G(x, G(y, u, w), s) \end{aligned} \tag{1.10}$$

Notamos duas maneiras de escrever a mesma coisa, por D_2 que declarava que não queremos ser manifestamente inconsistentes, devemos ter

$$G(G(x, y, z), u, G(v, w, s)) = G(x, G(y, u, w), s).$$

Ainda notamos que embora estas variáveis possam ter quaisquer valores, não ocorre o mesmo conjunto dos dois lados: Lado esquerdo $\{x, y, z, u, v, w, s\}$, lado

direito $\{x, y, u, w, s\}$. Portanto o lado esquerdo não deve depender de $z = \phi(a|bd)$ nem de $v = \phi(a|cd)$ explicitamente. As derivadas parciais com respeito a z ou v devem dar zero:

$$\begin{aligned} 0 &= \frac{\partial}{\partial z} G(G(x, y, z), u, G(v, w, s)) \\ &= \frac{\partial}{\partial r} G(r, u, G(v, w, s))_{r=G(x, y, z)} \frac{\partial}{\partial z} G(x, y, z) \end{aligned} \quad (1.11)$$

Se um produto é zero, pelo menos um dos fatores é zero, de onde concluímos que ou G não depende do primeiro argumento ou não depende do terceiro. Se não depende do primeiro

$$y = G(\phi(a|c), \phi(b|c), \phi(a|bc)) = G(\phi(b|c), \phi(a|bc)),$$

voltamos ao **Caso 3**. Se não depende do terceiro

$$y = G(\phi(a|c), \phi(b|c), \phi(a|bc)) = G(\phi(a|c), \phi(b|c))$$

e voltamos ao **Caso 1**.

Fica como

Exercício mostrar que o **Caso 6** pode ser reduzido ao **Caso 3** ou ao **Caso 4** e que o **Caso 7** aos **Caso 5** ou **Caso 6**

Concluímos portanto que

$$\begin{aligned} \phi(ab|c) &= G(a|c, b|ac) \\ &= G(b|c, a|bc) \end{aligned} \quad (1.12)$$

Cox coloca isto como um axioma, mas não precisamos fazer isto, basta dizer que existe uma função G mas que não sabemos *a priori* quais seus argumentos. A eliminação dos casos que contradizem o bom senso em casos suficientemente simples, mostra de forma satisfatória (o leitor pode pular e reclamar, mas terá que encontrar argumentos) que as equações 1.12 refletem a única opção. Uma das queixas pode ser sobre a diferenciabilidade de G . Mas estamos interessados em situações onde a informação pode mudar e não alterar significativamente as crenças e esperamos ao menos continuidade de G .

Note que agora será possível concluir que ‘Helena usa um tenis esquerdo vermelho e um tenis direito preto’ pode ser pouco plausível por que precisamos saber a plausibilidade de ‘Helena usa um tenis esquerdo vermelho dado que Helena usa um tenis direito preto’ e isto pode ser pouco plausível.

Mas ainda não acabamos. Precisamos determinar a função específica G .

1.2.4 Regra do produto: qual é a função G ?

Novamente olhamos para um caso simples, onde podemos escrever o resultado de duas maneiras. Considere a, b, c e d com $b|d$ e $c|d$ mutuamente exclusivos, e a asserção $a(b \vee c)$ uma conjunção que pode ser escrita como uma disjunção:

$$a(b \vee c) = (ab) \vee (ac). \quad (1.13)$$

Podemos usar o resultado para a soma para estudar o produto $\phi(a(b \vee c)|d)$:

$$\begin{aligned}\phi(a(b \vee c)|d) &= G(\phi(a|d), \phi(b \vee c|ad)) \\ &= G(\phi(a|d), \phi(b|ad) + \phi(c|ad))\end{aligned}\quad (1.14)$$

$$\begin{aligned}\phi((ab) \vee (ac)|d) &= \phi(ab|d) + \phi(ac|d) \\ &= G(\phi(a|d), \phi(b|ad)) + G(\phi(a|d), \phi(c|ad))\end{aligned}\quad (1.15)$$

onde a equação 1.14 usa primeiro que $a(b \vee c)$ é um produto e em segundo lugar a regra da soma para asserções mutuamente exclusivas $b|d$ e $c|d$. A equação 2.1 mostra o resultado de considerar a soma $(ab) \vee (ac)$. Mas devido à equação 1.13 e D_2 , estas duas formas devem dar o mesmo resultado:

$$G(x, y + z) = G(x, y) + G(x, z).\quad (1.16)$$

Novamente requerindo a diferenciabilidade, desta vez duas vezes, e definindo $w = y + z$ obtemos a equação diferencial

$$\frac{\partial^2 G(x, w)}{\partial w^2} = 0\quad (1.17)$$

que tem solução geral $G(x, w) = A(x)w + B(x)$ em termos de duas funções desconhecidas, mas fáceis de determinar. Substituindo esta forma em 1.16 obtemos

$$A(x)(y + z) + B(x) = A(x)(y + z) + 2B(x),\quad (1.18)$$

portanto $B(x) = 0$, ou seja $G(x, w) = A(x)w$. Agora olhamos para $a|d$ e usamos $a|d = ad|d$ para a e d quaisquer.

$$\begin{aligned}\phi(a|d) &= \phi(ad|d) = G(\phi(a|d), \phi(d|ad)) \\ &= G(\phi(a|d), \phi_v) = A(\phi(a|d))\phi_v\end{aligned}\quad (1.19)$$

onde $\phi(d|ad) = \phi_v$ pois, obviamente d é informação completa para d . Ou seja $x = A(x)\phi_v$, logo

$$G(x, w) = \frac{xw}{\phi_v}\quad (1.20)$$

isto significa que, para $e = b \vee c$, b e c mutuamente exclusivos

$$\phi(ae|d) = \frac{\phi(a|d)\phi(e|ad)}{\phi_v}\quad (1.21)$$

o que permite regraduar mais uma vez os números associados as crenças sem mudar a ordem.

Mas resta um problema: e se retirarmos a restrição de b e c mutuamente exclusivos? Precisamos usar a equação 1.8 para obter:

$$\phi(a \vee b|d) = \phi(a|d) + \phi(b|d) - \phi(ab|d)\quad (1.22)$$

$$\begin{aligned}
\phi(a(b \vee c)|d) &= G(\phi(a|d), \phi(b \vee c|ad)) \\
&= G(\phi(a|d), \phi(b|ad) + \phi(c|ad) - \phi(bc|ad)) \\
\phi((ab) \vee (ac)|d) &= \phi(ab|d) + \phi(ac|d) - \phi(abc|d) \\
&= G(\phi(a|d), \phi(b|ad)) + G(\phi(a|d), \phi(c|ad)) - G(\phi(a|d), \phi(bc|ad)) \\
&= G(\phi(a|d), \phi(b|ad)) + G(\phi(a|d), \phi(c|ad)) - G(\phi(a|d), G(\phi(b|ad), \phi(c|abd)))
\end{aligned} \tag{1.23}$$

$$\tag{1.24}$$

igualando os lados direitos das equações 1.23 e 1.24, obtemos uma nova equação funcional. Substituindo a forma para G da equação 1.20, vemos que o produto também funciona neste caso geral.

Exercício Mostre que a forma produto (eq. 1.20) é solução da equação funcional. Mostre que esta é a única forma se G for diferenciável duas vezes em cada argumento.

Da equação 1.21 obtemos

$$\frac{\phi(ae|d)}{\phi_v} = \frac{\phi(a|d)}{\phi_v} \frac{\phi(e|ad)}{\phi_v} \tag{1.25}$$

o que permite regradar mais uma vez os números associados as crenças sem mudar sua ordem. Crenças regradas, de forma bijetora representam o mesmo ordenamento e portanto podem ser ainda chamados de crenças. Definimos os novos números

$$p(a|b) = \frac{\phi(a|b)}{\phi_v} \tag{1.26}$$

e reescrevemos os resultados

$$\begin{aligned}
p(a|a) &= p_v = 1 \\
p(a|\bar{a}) &= p_f = 0 \\
p(a \vee b|c) &= p(a|c) + p(b|c) - p(ab|c) \\
p(ab|c) &= p(a|c)p(b|ac) \\
&= p(b|c)p(a|bc)
\end{aligned} \tag{1.27}$$

Começamos a reconhecer as fórmulas que descrevem as probabilidades da soma e do produto. Mas ainda não acabamos. Precisamos determinar o que acontece com a negação. Começamos com $a \vee \bar{a}|d$ que deve ser sempre verdade e com $a\bar{a}|d$ que deve ser sempre falso:

$$\begin{aligned}
1 &= p(a \vee \bar{a}|d) \\
&= p(a|d) + p(\bar{a}|d) - p(a\bar{a}|d) \\
&= p(a|d) + p(\bar{a}|d),
\end{aligned} \tag{1.28}$$

ou a soma das crenças regradas de uma asserção e da sua negação é um.

Isso completa a identificação das crenças ou plausibilidade regradas em números que satisfazem as regras da probabilidade. Concluimos que a estrutura

matemática adequada, e que usaremos nestas notas, para descrever situações de informação incompleta é a teoria de probabilidades.

O que foi obtido pode ser comparado com os axiomas de Kolmogorov ¹³. Vemos uma diferença importante. Na formulação da teoria de probabilidade como um capítulo da teoria da medida, as probabilidades são medidas e não há menção a condicionais. Rao adicionou mais tarde a complementação introduzindo, como uma idéia tardia, a probabilidade condicional definida a partir do teorema de Bayes, que Cox obteve como uma consequência direta da consistência. A partir de

$$\begin{aligned} p(ab|c) &= p(a|c)p(b|ac) \\ &= p(b|c)p(a|bc) \end{aligned} \quad (1.29)$$

obtemos o teorema de Bayes ¹⁴

$$p(a|bc) = \frac{p(a|c)p(b|ac)}{p(b|c)} \quad (1.30)$$

que esconde, atrás de sua grande simplicidade, uma importância enorme, que deriva em parte do número ilimitado que tem encontrada em várias áreas da ciência.

Este é o conteúdo dos teoremas de Cox: uma atribuição de números para descrever as crenças em asserções, dada a informação, que satisfaça os casos particulares, pode ser mudada de forma a não alterar o ordenamento das crenças e preferências e a satisfazer as regras da probabilidade. Tem cheiro e cor de probabilidade e tem todas as propriedades das probabilidades. Não falaremos mais sobre plausibilidade. Não sabemos o que era, e a abandonamos como a um andaime, após ter construído o edifício da teoria de probabilidades. Obviamente este exercício não forneceu os valores das probabilidades. Que bom, senão fechariam os institutos dedicados ao estudo e às aplicações das probabilidades. Mais sérios, podemos dizer que a nossa grande preocupação agora será dirigida à busca de técnicas que baseadas na informação disponível permitam atribuições ou talvez o problema associado mas diferente, de atualização dos números associados a probabilidades dos eventos ou asserções de interesse quando recebemos nova informação. Esta é a preocupação central da inferência e da teoria de aprendizado e nos levará à introdução da idéia de entropia. A entropia no sentido de teoria de informação está intimamente ligada à idéia de entropia termodinâmica e mais ainda à de Mecânica Estatística como veremos mais tarde. Poderemos afirmar que a Mecânica Estatística foi a primeira teoria de informação, embora não seja costumeiro colocá-la nessa luz.

1.2.5 O teorema de Bayes e Informação Incompleta

Vejamos agora alguns exemplos da utilização destes resultados em casos simples onde há informação incompleta.

¹³Kolmogorov

¹⁴T. Bayes formulou a parte da inversão: $p(a|bc) \propto p(b|ac)$, Laplace o escreveu pela primeira vez e deu-lhe a devida importância

Voltemos agora aos silogismos iniciais. Suponha que

- $A = \text{“Está chovendo”}$
- $B = \text{“Há nuvens”}$
- $C = \text{“}A \rightarrow B\text{”}$

Note que a implicação lógica não segue da causalidade física. Chove porque há nuvens do ponto de vista de causalidade, mas do ponto de vista lógico saber que chove obriga à conclusão que deve haver nuvens. Suponha que seja dada a informação B , ou seja é dado que há nuvens. Dentro da lógica aristotélica nada podemos dizer. Devemos com base nisso desprezar por ilógicos quem nos aconselha a levar um guarda-chuva porque há nuvens? Vejamos o que nos diz a teoria das probabilidades. Neste caso o teorema de Bayes começa a mostrar a sua força. A probabilidade $P(A|CI)$ representa a crença que esteja chovendo, sob a informação C , mas não levando em conta se há ou não nuvens. Também leva em conta I , tudo o que é sabido sobre o clima nesta estação do ano, podendo ser muita informação ou nenhuma. Não importa efetivamente que número $P(A|CI)$ seja, estará entre zero e um. Esta probabilidade é dita *a priori* em relação a B . Uma vez que se recebe e incorpora a informação que efetivamente há nuvens, ou seja B , então passaremos a $P(A|BCI)$, outro número, que é chamada a probabilidade *a posteriori* ou simplesmente posterior. Aplicando Bayes

$$P(A|BCI) = \frac{P(A|CI)P(B|ACI)}{P(B|CI)}, \quad (1.31)$$

que relaciona a probabilidade *a priori* e a posterior. Cortando e deixando para depois uma discussão longa sobre inferência, podemos dizer que é razoável que usemos a posterior para decidir se levaremos ou não o guarda-chuvas. A probabilidade $P(B|ACI)$ recebe o nome de verossimilhança (*likelihood* e poderia ser calculada se tivéssemos um modelo sobre a influência de A em B , mas é isso o que temos, este é um caso de informação completa! Temos certeza da veracidade de B se AC for dado. Assim

$$P(B|ACI) = 1. \quad (1.32)$$

O quê pode ser dito sobre o denominador $P(B|CI)$? O mínimo que pode ser dito é que

$$P(B|CI) \leq 1. \quad (1.33)$$

Substituindo estes resultados obtemos

$$P(A|BCI) \geq P(A|CI), \quad (1.34)$$

a probabilidade que atribuiremos a que A seja verdade é maior ou igual se levarmos em conta o fato que há nuvens, que aquela que atribuímos sem saber se há nuvens ou não. Finalmente nos diz que a pessoa que percebe que há nuvens e leva o guarda-chuvas está agindo de forma lógica, não dentro da lógica

aristotélica, mas segunda a extensão da lógica para casos de informação incompleta, representada pela teoria das probabilidades. Vemos que o bom senso diário desta situação pode ser deduzido dos desejos impostos por Cox.

Suponha outro caso de informação incompleta. Agora A é dado como falso. Continuaremos a insistir que não podemos dizer nada sobre B do ponto de vista da lógica? O teorema de Bayes, nos diz

$$P(B|\bar{A}CI) = \frac{P(B|CI)P(\bar{A}|BCI)}{P(\bar{A}|CI)}, \quad (1.35)$$

e também sabemos que $P(A|BCI) \geq P(A|CI)$ da análise anterior. Ainda mais, temos que $P(A|BCI) = 1 - P(\bar{A}|BCI)$ e $P(A|CI) = 1 - P(\bar{A}|CI)$, portanto

$$P(B|\bar{A}CI) \leq P(B|CI) \quad (1.36)$$

levando à conclusão que se não está chovendo, devemos atribuir uma probabilidade menor a que haja nuvens. Quem está mais disposto a carregar um chapéu de sol porque recebeu informação que não está chovendo, age de forma lógica.

Exemplo

Consideremos um exemplo clássico de testes médicos. Um teste médico serve para ajudar a determinar se um paciente está doente, mas ele não é perfeito e há evidência, baseado na história que há falsos positivos e falsos negativos. O que significa um resultado positivo? Para proceder, o mais importante é esclarecer quais são as asserções relevantes.

Consideremos as asserções

- D = "paciente está doente"
- A = "resultado do teste é positivo"

junto com os dados sobre

- especificidade: $P(A|D) = .90$, a probabilidade de dar positivo no teste na condição de estar doente
- sensibilidade: $1 - P(A|\bar{D}) = 1 - .2 = .8$, a probabilidade de teste dar positivo no caso em que o paciente não está doente,

Vemos que o teste é bastante específico (90%) e bastante sensível ((80 = 100 - 20)%).

Suponha que seu resultado no teste deu positivo, A é verdade. Isto significa que está doente? Há possibilidade de erros portanto não temos informação completa. Qual é a pergunta que devemos fazer? Pode não ser o mais óbvio a se fazer quando se recebe uma notícia ruim, mas em geral devemos aplicar o teorema de Bayes. Assim poderemos calcular $P(D|AI)$ que é o que realmente interessa, a probabilidade de ter a doença,

$$P(D|AI) = \frac{P(D|I)P(A|DI)}{P(A|I)}, \quad (1.37)$$

e também

$$P(\bar{D}|AI) = \frac{P(\bar{D}|I)P(A|\bar{D}I)}{P(A|I)}, \quad (1.38)$$

os denominadores são inconvenientes e os eliminamos olhando para a razão

$$\frac{P(D|AI)}{P(\bar{D}|AI)} = \frac{P(D|I)P(A|DI)}{P(\bar{D}|I)P(A|\bar{D}I)}. \quad (1.39)$$

Após considerar a equação acima percebemos que não temos dados suficientes para entrar em pânico. A razão entre as probabilidades que nos interessa é $P(D|AI)/P(\bar{D}|AI)$ depende de dados que temos, sobre a especificidade e sensibilidade do teste e de dados que não temos sobre a distribuição da doença na população. A teoria que não pode nesta altura nos dar a resposta que buscamos, faz a segunda melhor coisa, indicando que informação adicional devemos procurar. Após esta análise voltamos ao médico e perguntamos se ele tem informação sobre a distribuição *a priori* da doença na população caracterizada por I . Suponha que recebamos informação que $\frac{P(D|I)}{P(\bar{D}|I)} = .99/.01$, só 1% da população tem a doença. Segue que

$$\frac{P(D|AI)}{P(\bar{D}|AI)} = \frac{P(D|I)P(A|DI)}{P(\bar{D}|I)P(A|\bar{D}I)} = \frac{.01 \times .90}{.99 \times .20} = 0.045. \quad (1.40)$$

ou seja a probabilidade de não ter a doença é aproximadamente .95. Não que isto seja uma boa notícia, afinal a probabilidade que era de 1% de ter a doença passou para 4.5% : aumentou quase cinco vezes. Mas não devemos ainda entrar em pânico nem jogar fora a informação que ganhamos com o teste.

1.2.6 Jaynes e o bom senso

O próximo caso simples lida com informação neutra. Suponha que

$$A|C \geq A|C',$$

ou seja a plausibilidade de A diminui quando a informação disponível passa de C para C' . Suponha que para B isso não aconteça. Pensemos no caso que B é indiferente ante a mudança de C para C' . Isto é

$$B|C = B|C'.$$

Parece razoável que se a asserção conjunta AB for considerada, esta seria mais plausível nas condições C que C' ; isto é seria desejável que a teoria satisfizesse

- $A|C \geq A|C'$ e $B|DC = B|DC'$, para qualquer D , implicam que $AB|C \geq AB|C'$

Jaynes defende que este desejo está de acordo com o *bom senso*. Talvez seja difícil definir o que é bom senso, mas seria mais difícil negar que isto seja

razoável. Jaynes coloca isto como um dos axiomas para chegar à teoria de probabilidades.

O leitor talvez possa se convencer através de um simples exemplo. Seja A ='Há vida em Marte', C = 'Há água em Marte', $C' = \bar{C}$, a negação de C . Suponhamos óbvio que $A|C \geq A|C'$. Suponha que B ='Hoje é segunda feira'. Certamente $B|C = B|C'$. e também é razoável que a plausibilidade de que haja vida em Marte e hoje seja segunda feira' dado que 'há água em Marte' é maior ou igual a plausibilidade de que 'haja vida em Marte e hoje seja segunda' dado que 'não há água em Marte'.

Pelo regra do produto, podemos provar isto

$$P(AB|C) = P(A|C)P(B|AC) = P(A|C)P(B|AC')$$

$$P(AB|C) \geq P(A|C')P(B|AC') = P(AB|C').$$

1.2.7 Exemplo do Teorema de Bayes e Ajuste de funções

Uma das primeiras lições que os estudantes de física tem ao entrarem num laboratório é sobre ajuste de curvas usando conjuntos de medidas empíricas.

Um objeto cai e medimos as posições ou velocidades como função do tempo. Estão de acordo com o que se espera de um objeto que cai na presença de um campo gravitacional? Qual é o valor de g , a aceleração da gravidade? Só para deixar isto claro, não faltarão exemplos complicados mais adiante nestas notas, olharemos para o caso em que obtemos um conjunto de dados

$$D = \{v_1, v_2, \dots, v_N\} \quad (1.41)$$

para as velocidades medidas em

$$T = \{t_1, t_2, \dots, t_N\}. \quad (1.42)$$

O modelo que queremos avaliar, refutar ou aceitar (pelo menos até ter mais dados) é

$$\mathcal{M} : v = v_0 + gt \quad (1.43)$$

A pergunta que quer ser respondida diz respeito a asserções do tipo $H(g)$: "O valor da aceleração da gravidade é g ". Para cada valor de g que for inserido nessa frase, teremos uma asserção diferente. O que queremos é comparar o mérito de cada asserção, qual é a probabilidade de cada uma delas, para todos os valores que possam ser inseridos.

O teorema de Bayes nos permite escrever

$$P(H|DI) = \frac{P(H|I)P(D|HI)}{P(D|I)}. \quad (1.44)$$

O que será discutido a seguir é fundamental para este curso. Será discutido em contextos mais complicados e portanto vale a pena o esforço de entender cada passo. É tão importante que cada termo recebe um nome.

Em primeiro lugar temos que definir as asserções relevantes ao problema. A parte que parece menos importante, mas que na realidade é fundamental é I , que define várias coisas que de tão importantes são consideradas desnecessárias pois, para que falar o óbvio? I denota toda a informação sobre a experiência. Qual é a teoria que queremos confrontar com os dados? Quais são as características do aparelho de medida? Em que instantes de tempo t_i fazemos as medidas, quais as incertezas que estas medidas têm? Em que planeta estamos? e muito mais que ficará tácitamente escondida, mas ainda relevante.

D é o conjunto de dados, H é a hipótese que quer ser testada.

Agora o significado das probabilidades que aparecem na equação 1.44. Começamos pelo conhecimento que temos sobre o contexto experimental mas sem levar em consideração os dados. A distribuição de probabilidades *a priori* $P(H|I)$ codifica tudo o que sabemos sobre a gravitação antes de entrar no laboratório. Se não soubermos o planeta onde a experiência é realizada, fica difícil esperar um valor e não outro. Todas as gerações de estudantes que fizeram esta experiência, dos quais temos notícia, o fizeram na terra. O resultado deu algo que se parece com 9.8 ms^{-2} . Se o resultado final fosse 9.8 kms^{-2} o aluno ficaria tentado a mudar seu resultado, mudaria de forma ad hoc seu valor no relatório, o que seria desonesto, ou faria novamente as contas. Se ainda persistir o problema, jogaria fora os dados. Isto é desonesto? Não se estiver de acordo com a sua probabilidade a priori. Qual é a probabilidade que a aceleração da gravidade seja 9.8 kms^{-2} em São Paulo? qual é a probabilidade que você atribuiria antes de entrar no laboratório? Quanto voce estaria disposto a apostar contra a veracidade dessa asserção? A priori, o estudante sabe que o valor estará por volta de dez, e pode ser constante entre 7 e 15. Muito mais que isso ou muito menos, deve ser erro, e é melhor jogar o que o estudante chama de ponto fora da curva. Isso é perfeitamente lógico e deve ser feito a não ser que em I haja a possibilidade de que algo possa mudar o valor esperado. Por exemplo a experiência esta sendo feita em cima de uma cratera aberta por um meteorito composto do elemento X. Então podemos permitir a suposição que novos valores sejam encontrados. Seriamos cegos se considerassemos a probabilidade a priori de encontrar valores muito diferentes, nula e se assim for feito, certamente não os encontraremos.

A probabilidade $P(D|HI)$ descreve quão verossímil seria encontrar esse conjunto de dados se além de I , o valor particular de g representado por H fosse o correto. Esta é a famosa contribuição do reverendo Thomas Bayes¹⁵: a inversão. Queríamos saber a probabilidade de g ter um certo valor nas condições que os dados foram observados, mas estamos olhando para a probabilidade dos dados no caso que a teoria (contida em I) e um valor particular do parâmetro g sejam verdade. Este termo recebe o nome verossimilhança (likelihood em inglês).

O denominador $P(D|I)$ será interessante em outros contextos. Em geral é chamado de evidência. Pode ser obtido usando o fato que g não pode ter dois valores diferentes. As asserções para valores de g diferentes são mutuamente exclusivas. Portando a soma sobre todas as possibilidades é um. Neste caso em que g toma valores reais, é interessante considerar que as asserções tem o

¹⁵referencia de bayes

significado que o valor da aceleração da gravidade está entre g e $g + dg$, somas são substituídas por integrais.

O resultado de toda a análise será a obtenção de $P(H|DI)$ que se chama a distribuição de (densidade de) probabilidade posterior, ou simplesmente a posterior.

A crítica mais comum é que a *realidade objetiva* é única e portanto não é possível que haja uma probabilidade para o valor de g . Mas não é isso o que esta probabilidade significa. g pode ter um valor único objetivo ¹⁶. O que a posterior, ou a a priori significam é que não temos informação completa e que só podemos atribuir probabilidades às diferentes asserções sobre o valor de g . Mais dados, ou seja mais informação, permitirão novas estimativas. O que estas probabilidades codificam não é o valor de g , mas a crença que esse seja o valor correto.

1.2.8 Obtendo a posterior

Há vários exemplos que mostram a importância de determinar a distribuição a priori com muito cuidado. Podemos dizer que a probabilidade que $g < 0$ deve ser zero. Os objetos mais densos que o ar não caem para cima. Também podemos limitar os valores superiores. Poderíamos dizer que $P(H|I) = c$ se $g_{min} < g < g_{max}$ e zero fora desse intervalo. A constante c é tal que $\int_{g_{min}}^{g_{max}} P(H|I)dg = 1$ ou $c^{-1} = g_{max} - g_{min}$.

A verossimilhança $P(D|HI)$ leva em conta que as medidas são sujeitas a erros. Poderíamos dizer, por exemplo, que o modelo teórico e o modelo sobre o aparelho de medidas, juntos nos levam a esperar, que para os valor de tempo t_i , onde é feita a medida,

$$v_i = v_0 + gt_i + \eta_i. \quad (1.45)$$

O resultado esperado puramente pelo modelo teórico (eq. 1.43) é corrompido por algo que chamamos ruído. Isto esconde uma grande quantidade de ignorância sobre o processo de medida. Se pudessemos aumentar o controle sobre o aparelho de medida (e.g. temperatura, vento, correntes elétricas, valores das resistências, ...etc.) a amplitude de η_i poderia ser menor. Mas sempre há uma incerteza sobre o valor medido. Temos que fazer algumas hipóteses sobre η_i . Estas, supostas verdadeiras, serão incluídas na asserção I . Como não temos informação completa, devemos descrever o conjunto de η s por uma distribuição de probabilidade $P(\eta_1 \dots \eta_N | I_{exp})$. É razoável supor que as diferentes medidas

¹⁶Sabemos que g uniforme, constante é só uma aproximação válida para quedas em distâncias pequenas em comparação ao raio da terra dentro da teoria de Newton. Mas também sabemos que essa teoria não é final, tendo sido substituída pela de Einstein, e certamente não sabemos por qual teoria vai ser substituída em anos futuros. Não sobra muito do conceito de um g que descreve uma realidade objetiva

são independentes, e usando a regra do produto lógico

$$\begin{aligned}
P(\eta_1\eta_2\dots\eta_N|I_{exp}) &= P(\eta_1|I_{exp})P(\eta_2\dots\eta_N|\eta_1I_{exp}) \\
&= P(\eta_1|I_{exp})P(\eta_2\eta_3\dots\eta_N|I_{exp}) \\
&= P(\eta_1|I_{exp})P(\eta_2|I_{exp})P(\eta_3\dots\eta_N|\eta_2I_{exp}) \\
&\dots \\
&= \prod_i^N P(\eta_i|I_{exp}), \tag{1.46}
\end{aligned}$$

onde usamos na primeira e terceira linha a regra do produto e na segunda a independência dos valores de $\eta_2, \eta_3, \dots, \eta_N$ e o de η_1 . Temos que a distribuição conjunta é o produto das distribuições individuais.

Qual é a distribuição $P(\eta|I_{exp})$ a ser usada. Ainda devemos supor algo mais, por exemplo média nula e variância finita σ^2 . No capítulo sobre entropia justificaremos porque isto nos leva a uma distribuição gaussiana

$$P(\eta_1, \eta_2, \dots, \eta_N | I_{exp}) = \frac{e^{-\sum_{i=1}^N \frac{\eta_i^2}{2\sigma^2}}}{(2\pi\sigma^2)^{N/2}}$$

mas $\eta_i = v_i - v_0 - gt_i$, portanto

$$P(\eta_1, \eta_2, \dots, \eta_N | I_{exp}) = \frac{e^{-\sum_{i=1}^N \frac{(v_i - v_0 - gt_i)^2}{2\sigma^2}}}{(2\pi\sigma^2)^{N/2}}$$

e a posterior

$$P(H|DI) = \frac{P(H|I) e^{-\sum_{i=1}^N \frac{(v_i - v_0 - gt_i)^2}{2\sigma^2}}}{PD|I} \tag{1.47}$$

O problema de inferência está pronto. Mas qual é a resposta a ser dada? Há várias quantidades que podem ser extraídas da posterior. Por simplicidade nos contentamos com o valor de g que é mais provável. Se a a priori é constante na região que a gaussiana é relevante, podemos esquecer o prefator. Teremos a estimativa conhecida como máxima verossimilhança. A resposta é simplesmente o valor que torna o argumento da exponencial máximo,

$$g_{MV} = \arg \min_g \sum_{i=1}^N \frac{(v_i - v_0 - gt_i)^2}{2\sigma^2} \tag{1.48}$$

que é o velho método de mínimos quadrados. Se $P(H|I)$ fosse relevante, teríamos o máximo a posteriori g_{MAP} .

Mas escolher um valor sobre os outros esconde que não temos certeza absoluta. A largura da gaussiana σ/\sqrt{N} nos dá uma medida da incerteza.

Ainda podemos levar em conta que valores vizinhos de g_{MAP} tem probabilidade não desprezível e apresentar o valor esperado

$$g^* = \int gP(H|DI)dg. \tag{1.49}$$

O que ganhamos em apresentar assim o método dos mínimos quadrados que os estudantes devem ter visto há muito tempo? Suponha por exemplo, que voce colha mais informação sobre o aparelho de medida e chegue à conclusão que a distribuição dos η não é gaussiana. Ainda assim usaria o método dos mínimos quadrados? Podemos ver quais as suposições necessárias e tentar verificar se cada uma delas é razoável ou não. Isto não é pouco, a apresentação cuidadosa pode evitar suposições que não gostaríamos de fazer ao analisar os dados de uma experiência. Tão importante quanto usar a informação disponível é não usar a que não o é. O próximo capítulo levará esta idéia adiante.

Capítulo 2

Probabilidades e o Teorema do Limite Central

2.1 Introdução: Kolmogorov e as probabilidades

Kolmogorov introduziu na década dos trinta ¹ os seus famosos axiomas para a teoria das probabilidades. No seu livro ele declara que não vai entrar no debate filosófico sobre o significado de probabilidades e depois dá uma pequena justificativa dos axiomas com base na interpretação freqüentista de von Mises. No capítulo anterior descrevemos os motivos que nos levam a achar tal posição, isto é freqüentista, incompleta e até, como mostraremos abaixo, errada. Pelo contrário, os axiomas de Kolmogorov, que codificam o bom senso da área já existente no trabalho de Laplace, podem ser vistos equivalentes aos resultados obtidos no capítulo 1.

Kolmogorov começa por considerar E uma coleção de elementos A, B, C, \dots que são eventos elementares e em nossa discussão anterior chamamos de asserções. \mathcal{F} é o conjunto de subconjuntos de E . Um sistema de conjuntos é chamado um campo se a soma, produto, interseção de dois elementos quaisquer pertencem ao sistema. Os axiomas são

- AK1) \mathcal{F} é um campo de conjuntos fechado ante um número de uniões e interseções enumeráveis e se $A \in \mathcal{F}$ e $\bar{A} = E - A$, então $\bar{A} \in \mathcal{F}$

ou seja \mathcal{F} é um σ -campo,

- AK2) \mathcal{F} contém o conjunto E .
- AK3) A cada conjunto $A \in \mathcal{F}$ é atribuído um número real não negativo, chamado de probabilidade do evento A .

¹e.g. ver <http://www.mathematik.com/Kolmogorov/index.html>

- AK4) $P(E) = 1$
- AK5) Se $A \vee B = \emptyset$, então $P(A + B) = P(A) + P(B)$

Vemos que estes axiomas estão de acordo com os resultados do capítulo anterior. A probabilidade da *certeza* é 1 por AK4; a probabilidade está entre zero e um e a probabilidade da disjunção de asserções que não tem elementos em comum é a soma das probabilidades. Notamos porém a falta de uma regra para o produto. Kolmogorov também e na página 6 ele introduz as probabilidades condicionais através de

$$p(A|B) = \frac{P(AB)}{P(B)} \quad (2.1)$$

de onde segue para a prova do teorema de Bayes.

Exercício

O que falta para poder obter o teorema de Bayes?

Se uma vez estabelecidos os axiomas, partirmos para as aplicações matemáticas, não haverá nenhuma diferença de resultados. Enfatizamos que as diferenças que temos são sobre a motivação dos axiomas e com a interpretação da idéia de probabilidades.

2.2 Momentos...

A partir de agora introduziremos alguns resultados matemáticos que serão úteis no desenrolar do curso. Em particular estamos interessados em grandezas físicas descritas por variáveis que tomam valores em intervalos dos reais, que chamaremos I .

No que segue lidaremos com asserções do tipo “a variável X toma valores entre x e $x + dx$ ”. Não interessa ainda como, mas suponha que atribuímos um número a esta probabilidade. Introduziremos a densidade $P(x)$ tal que a probabilidade de que “a variável X toma valores entre x e $x + dx$ ” é dada por $P(x)dx$. A função $P(x)$ não é uma probabilidade mas é chamada de densidade de probabilidade ². Teremos então que

- $P(x) \geq 0$
- $\int_I P(x)dx = 1$

Suponha que queremos ao falar de X dar um número que nos dê alguma informação sobre os seu valor. A informação disponível será equivalente à densidade para todo x . Isto talvez seja muito. Queremos poder comunicar o valor de x com um número, isto é um estimador ou estimativa de X . Há várias possibilidades e cada uma tem utilidade

²Usamos a letra P por motivos históricos e eventualmente a chamaremos de probabilidade, por preguiça.

- (1) $x_M = \text{maxarg}P(x)$
- (2) $\langle x \rangle = \mathbb{E}[x] = \int_I xP(x)dx$
- (3) x_m tal que $\int_{x \leq x_m} P(x)dx = \int_{x \geq x_m} P(x)dx$

estes números recebem os nomes de (1) moda, (2) valor esperado ou média e (3) mediana. Podemos pensar em outros.

Podemos ser muito útil caracterizar a distribuição pelas *flutuações* em torno da média: quanto se afasta x da sua média, $\Delta x = x - \langle x \rangle$. Novamente podemos olhar para a média, só que agora das flutuações e vemos que $\langle \Delta x \rangle = 0$, isto não significa que a ideia flutuação não seja útil, só que por construção a sua média é nula. A média do seu quadrado é muito útil:

$$\sigma_x^2 = \langle (x - \langle x \rangle)^2 \rangle. \quad (2.2)$$

σ_X recebe o nome de variância. Mostre que $\sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2$.

Exercício

Pense no significado de cada um dos estimadores e da variância σ_X e proponha outros estimadores.

O valor esperado será muito usado no que segue. Podemos generalizar a sua definição e introduzir os momentos de uma distribuição:

- $\langle x^n \rangle = \mathbb{E}[x^n] = \int_I x^n P(x)dx$

para valores inteiros de n (claro que caso a integral exista). A notação que usamos de alguma forma deixa esquecida a ideia que a probabilidade depende da informação disponível portanto usaremos ³ a notação

- $\langle x^n \rangle_{|C} = \mathbb{E}[x^n | C] = \int_I x^n P(x|C)dx$

para identificar claramente que estes são os momentos de X sob a informação C .

Os momentos centrais são definidos da mesma forma mas para a variável deslocada para que a média seja nula:

- $\langle (x - \langle x \rangle)^n \rangle_{|C} = \int_I (x - \langle x \rangle)^n P(x|C)dx$

As grandezas de interesse em Mecânica Estatística serão tipicamente originadas por somas de grande número de outras variáveis, por exemplo a energia de um gás terá contribuições das energias cinéticas de cada molécula mais as interações entre elas. Suponha que $Y = X_1 + X_2$. O que Y significa? Do ponto de vista de aritmética não insultaremos o leitor. Significa o óbvio. Do ponto de vista de asserções, temos um conjunto de asserções simples do tipo “a variável X_i toma valores entre x_i e $x_i + dx_i$ ” para $i = 1, 2$ e suponha que de alguma

³usaremos esta notação às vezes, pois usaremos o direito de ser inconsistentes na notação, esperando que isso não confunda o leitor, mas torne imune às várias notações na literatura.

forma atribuímos números a suas probabilidades. Queremos analisar, sob essa informação a asserção “a variável Y toma valores entre y e $y + dy$ ”. Notemos que a asserções compostas $A_1 = “x_1 = .17$ e $x_2 = .25”$ e $A_2 = “x_1 = .42$ e $x_2 = 0.”$ levam à mesma conclusão sobre o valor de Y . Mas elas são disjuntas no sentido que $A_1 A_2$ como produto lógico não pode ser verdade. As duas não podem ser simultaneamente verdadeiras. A probabilidade da soma lógica $A_1 + A_2$ é então a soma das probabilidades. Mas há outros casos de conjunções que dão o mesmo resultado para Y e devem ser levadas em conta: devemos somar sobre todas elas. Olharemos para somas deste tipo, $Y = X_1 + X_2 + X_3 + \dots X_N$, quando o número de termos na soma é muito grande. Lembrem que o número de átomos em alguns poucos gramas é da ordem de 10^{23} .

2.2.1 Marginais, Independência

Considere a distribuição conjunta de duas variáveis X , e Y . Para cada valor x que X pode tomar, podemos *marginalizar*, isto é calcular a soma sobre todos os possíveis valores de Y

$$p(x) = \sum_y p(x, y) \quad (2.3)$$

é chamada de distribuição de probabilidades marginal, que não é nada além da distribuição de probabilidades de X . O processo de soma (ou integração no caso de valores no contínuo). É claro que também podemos marginalizar somando sobre os valores que X pode tomar:

$$p(y) = \sum_x p(x, y). \quad (2.4)$$

Também podemos usar a regra do produto $p(x, y) = p(x)p(y|x) = p(y)p(x|y)$. No caso especial em que conhecimento de X não diz nada sobre Y , ou seja quando $p(y|x)$ não depende de x segue que $p(y|x) = p(y)$ e assim

$$p(x, y) = p(y)p(x|y) = p(x)p(y), \quad (2.5)$$

e as variáveis X e Y são ditas *independentes*.

Exercício

Prove que se $p(x|y)$ não depende de y , então $p(y|x)$ não depende de x .

2.3 Convoluções e Cumulantes

Considere variáveis idênticas X_i que tomam valores reais $\{x_1, x_2, \dots\}$ tal que $P(x_i)$ é o mesmo para todo i . Consideraremos o caso em que para qualquer $i \neq j$, os X_i são independentes entre si e são igualmente distribuídos⁴. Estamos

⁴Independentes e igualmente distribuídos: usualmente abreviado por i.i.d.

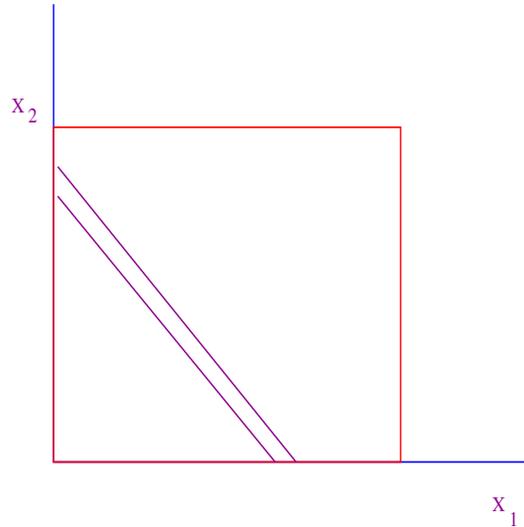


Figura 2.1: No plano X_1X_2 temos a região onde o valor de Y está entre y e $y + dy$. Todos os pares x_1 e x_2 nela contribuem para a probabilidade de Y

interessados na variável Y que toma valores em $y = \sum_{i=1..n} x_i$. Em particular, qual é a distribuição de $P(y|N = n)$? Começemos com $N = 2$, a probabilidade que Y tenha um valor entre y e $y + dy$ é obtida a partir de todas as formas que $y \leq x_1 + x_2 \leq y + dy$, com pesos iguais à probabilidade de ocorrência de x_1 e x_2 . Ver a figura 2.1. Para ser específicos chamaremos $P(x_i)$ a distribuição de valores de x_i , embora estejamos considerando que independe de i . A asserção que o “valor de Y esta entre y e $y + dy$ ” é a soma lógica de todas as asserções do tipo “ X_1 tem valor x_1 e X_2 tem valor x_2 ”, restritas ao caso em que $y \leq x_1 + x_2 \leq y + dy$ e portanto tem probabilidade

$$P(y|N = 2)dy = \int_{y \leq x_1 + x_2 \leq y + dy} dx_1 dx_2 P(x_1)P(x_2), \quad (2.6)$$

pois cada par de valores temos uma asserção disjunta. O vínculo $y \leq x_1 + x_2 \leq y + dy$ pode ser removido introduzindo a função χ_A que é 1 se a condição A for satisfeita e zero se não ⁵.

$$P(y|N = 2)dy = \int \chi_{y \leq x_1 + x_2 \leq y + dy} dx_1 dx_2 P(x_1)P(x_2), \quad (2.7)$$

⁵ χ_A é chamada a função característica do intervalo ou conjunto A , não confunda com a função característica da distribuição de probabilidades definida abaixo.

onde agora a integração é sobre todo o domínio de (x_1, x_2) . Introduzimos uma representação para χ em termos da integral de uma seqüência de funções $\delta_n(A)$:

$$\begin{aligned}\delta_n(y \leq x_1 + x_2 \leq y + \Delta y_n) &= \frac{1}{\Delta y_n}, \quad \text{se } y \leq x_1 + x_2 \leq y + \Delta y_n \\ &= 0, \quad \text{se não}\end{aligned}\tag{2.8}$$

e obtemos, tomando o limite para $n \rightarrow \infty$, tal que Δy_n va para zero,

$$P(y|N = 2) = \int dx_1 dx_2 P(x_1)P(x_2)\delta(y - x_1 + x_2),\tag{2.9}$$

$$P(y|N = 2) = \int dx P(x)P(y - x),\tag{2.10}$$

isto é, a convolução de $P(x_1)$ e $P(x_2)$ denotada por $(P * P)(y)$.

Outra forma: marginalização

Podemos ver como o resultado acima decorre das regras da probabilidade de outra forma: marginalizando. Começamos com a distribuição conjunta das variáveis Y, X_1 e X_2 e integramos sobre todos os valores de X_1 e X_2 :

$$P(y) = \int dx_1 dx_2 P(y, x_1, x_2).\tag{2.11}$$

Da regra do produto

$$P(y) = \int dx_1 dx_2 P(y|x_1, x_2)P(x_1, x_2).\tag{2.12}$$

Mas Y esta totalmente determinado se X_1 e X_2 forem conhecidos, portanto $P(y|x_1, x_2) = \delta(y - x_1 - x_2)$. Se novamente considerarmos X_1 e X_2 independentes: $P(x_1, x_2) = P(x_1)P(x_2)$, obtemos novamente a equação 2.10. A vantagem disto é que podemos facilmente obter expressões para funções gerais. Se $y = f(x_1, x_2)$, então

$$P(y) = \int dx_1 dx_2 \delta(y - f(x_1, x_2))P(x_1, x_2)\tag{2.13}$$

Exercício

Discuta a diferença entre $P(Y)$ a distribuição da soma e $P(X_1 X_2)$, para o produto lógico, que denotaremos por $P(x_1, x_2)$ a chamamos de distribuição conjunta de X_1 e X_2 . Considere também a variável Z que toma valores iguais ao produto dos valores de X_1 e X_2 : obtenha uma expressão para $P(z)$ quando $z = x_1 x_2$.

Distribuição da soma de variáveis e a função característica

Suponha que $P(x)$ satisfaz as seguintes condições:

- $\int P(x)dx = 1$, $P(x) \geq 0$ para todo x
- $\langle x \rangle = \int_{-\infty}^{\infty} xP(x)dx < \infty$,
- $\langle x^2 \rangle = \int_{-\infty}^{\infty} x^2P(x)dx < \infty$,

podemos introduzir a transformada de Fourier (TF) ⁶ e a inversa

$$\hat{P}(k) = \int_{-\infty}^{\infty} e^{-ikx} P(x)dx \quad (2.14)$$

$$P(x) = \int_{-\infty}^{\infty} e^{ikx} \hat{P}(k) \frac{dk}{2\pi} \quad (2.15)$$

A TF de uma distribuição de probabilidades é chamada de função característica. Ela também é chamada de função geradora dos momentos, pois uma expansão formal em série de potências da exponencial na equação 2.14 e a troca da ordem de integração e somatória nos mostra que os coeficientes estão relacionados aos momentos:

$$\begin{aligned} \hat{P}(k) &= \int_{-\infty}^{\infty} \sum_{s=0}^{\infty} \frac{(-ikx)^s}{s!} P(x) \\ &= \sum_{s=0}^{\infty} \frac{(-ik)^s}{s!} \langle x^s \rangle \end{aligned} \quad (2.16)$$

Tomemos a TF dos termos da equação 2.10, e usando :

$$\delta(k) = \int \frac{dx}{2\pi} e^{ikx} \quad (2.17)$$

obtemos

$$\begin{aligned} \hat{P}(k|N=2) &= \int dy dx e^{-iky} P(x)P(y-x), \\ &= \int \frac{dx dy dk_1 dk_2}{(2\pi)^2} \hat{P}(k_1|1) \hat{P}(k_2|1) e^{-iky + ik_1x + ik_2(y-x)}. \end{aligned} \quad (2.18)$$

Integrando sobre x e usando a representação da delta:

$$\begin{aligned} &= \int \frac{dy dk_1 dk_2}{2\pi} \hat{P}(k_1|1) \hat{P}(k_2|1) e^{-iky + ik_2y} \delta(k_1 - k_2), \\ &= \hat{P}(k|1) \hat{P}(k|1) = \hat{P}^2(k|N=1) \end{aligned} \quad (2.19)$$

⁶Para que exista é suficiente ainda que P seja seccionalmente contínua em cada intervalo $[-M, N]$ e definir $\hat{P} = \lim_{N, M \rightarrow \infty} \int_{-M}^N e^{-ikx} P(x)dx$

Para a soma de $N = n$ variáveis x_i

$$P(y|N = n) = \int \prod_{i=1 \dots n} dx_i P(x_1)P(x_2) \dots P(y - \sum_{i=1}^{n-1} x_i), \quad (2.20)$$

ou, introduzindo uma integral mais

$$P(y|N = n) = \int \prod_{i=1 \dots n} dx P(x_1)P(x_2) \dots P(x_n) \delta(y - \sum_{i=1}^n x_i), \quad (2.21)$$

obtemos

$$\hat{P}(k|N = n) = \hat{P}^n(k|N = 1) \quad (2.22)$$

e a inversão da transformada nos dá a distribuição de $P(y|N = n)$. No espaço de Fourier a convolução é simples produto, ou seja vamos para o espaço de Fourier, multiplicamos e depois voltamos ao espaço original fazendo a transformação inversa.

Caso as funções características sejam positivas podemos tomar o logaritmo de cada lado da equação 2.22 e dado que produtos, ao tomar logaritmos, viram somas, temos

$$\log \hat{P}(k|N = n) = \sum_i^n \log \hat{P}(k|N = 1) = n \log \hat{P}(k|N = 1). \quad (2.23)$$

Isto nos leva a discutir os cumulantes $\{C_s(n)\}$ de uma distribuição, definidos através da expansão em série de potências de ik da sua função característica:

$$\log \hat{P}(k|N = n) = \sum_{s=0}^{\infty} C_s(n) \frac{(-ik)^s}{s!}. \quad (2.24)$$

A equação 2.23 nos indica o motivo do nome dos cumulantes: a aditividade (ou acúmulo) ante convoluções

$$\begin{aligned} C_s(\sum_i x_i) &= \sum_i C_s(x_i), \\ C_s(n) &= n C_s(1), \end{aligned} \quad (2.25)$$

onde a equação 2.25 segue porque as $\{x_i\}$ são identicamente distribuídas. Concluímos que quando variáveis aleatórias independentes se somam, os cumulantes da distribuição da soma são a soma dos cumulantes das distribuições.

Mas qual é a interpretação dos cumulantes? Pela definição através da série de potências, vemos que em termos da função característica

$$C_s = \frac{1}{(-i)^s} \left. \frac{d^s \log \hat{P}}{dk^s} \right|_{k=0} \quad (2.26)$$

Podemos calcular alguns dos primeiros,

$$\begin{aligned}
\log \hat{P}(k|N=1) &= \log \int e^{-ikx} P(x) dx \\
&= \log \int \sum_{s=0}^{\infty} \frac{(-ikx)^s}{s!} P(x) dx \\
&= \log \left(1 + \sum_{s=1}^{\infty} \frac{(-ik)^s}{s!} \langle x^s \rangle \right) \\
&= \sum_{s_1=1}^{\infty} \frac{(-ik)^{s_1}}{s_1!} \langle x^{s_1} \rangle - \frac{1}{2} \sum_{s_1, s_2=1}^{\infty} \frac{(-ik)^{s_1+s_2}}{s_1! s_2!} \langle x^{s_1} \rangle \langle x^{s_2} \rangle \\
&\quad + \frac{1}{3} \sum_{s_1, s_2, s_3=1}^{\infty} \frac{(-ik)^{s_1+s_2+s_3}}{s_1! s_2! s_3!} \langle x^{s_1} \rangle \langle x^{s_2} \rangle \langle x^{s_3} \rangle + \dots \quad (2.27)
\end{aligned}$$

onde usamos $\log(1+u) = -\sum_{l=1}^{\infty} (-u)^l/l$. Juntando os termos com a mesma potência de k obtemos os cumulantes em função dos momentos $\langle x^s \rangle$:

$$\begin{aligned}
C_0 &= 0, \\
C_1 &= \langle x \rangle, \\
C_2 &= \langle x^2 \rangle - \langle x \rangle^2, \\
C_3 &= \langle x^3 \rangle - 3 \langle x^2 \rangle \langle x \rangle + 2 \langle x \rangle^3, \\
C_4 &= \langle x^4 \rangle - 4 \langle x^3 \rangle \langle x \rangle - 3 \langle x^2 \rangle^2 + 12 \langle x^2 \rangle \langle x \rangle^2 \\
&\quad - 6 \langle x \rangle^4, \quad (2.28)
\end{aligned}$$

O cumulante para $s=0$ é nulo, devido à normalização da distribuição. Para $s=1$ é a média e para $s=2$ é a variância, ficando mais complicados para valores maiores de s .

Fica mais interessante se olharmos além da soma Y , para $Z = \frac{Y}{\sqrt{n}}$ e para $W = \frac{Y}{n}$. Colocamos um índice para indicar a que variável se refere o cumulante e obtemos a propriedade que é chamada de homogeneidade:

$$nC_1^x = C_1^Y(n) = \sqrt{n}C_1^Z(n) = nC_1^W(n), \quad (2.29)$$

Portanto $C_1^W(n) = C_1^x$ independe de n , o que é óbvio. Mas para valores de s maiores

$$nC_s^x = C_s^Y(n) = n^{s/2}C_s^Z(n) = n^s C_s^W(n), \quad (2.30)$$

Portanto

$$\begin{aligned}
C_s^Y(n) &= nC_s^x \\
C_s^Z(n) &= \frac{1}{n^{s/2-1}} C_s^x \\
C_s^W(n) &= \frac{1}{n^{s-1}} C_s^x, \quad (2.31)
\end{aligned}$$

que mostram o decaimento dos cumulantes como função de n . O expoente de n tem duas contribuições; o 1, que vem do acúmulo, e o $s/2$ ou s que vem do fator de escala de Z ou Y respectivamente. É mais interessante olhar para quantidades adimensionais para poder entender o significado relativo desses decaimentos. Podemos olhar para $(C_2^x)^{1/2}$ como a escala típica das flutuações de x em torno da média. A razão $u_s^x = C_s^x / (C_2^x)^{s/2}$ é adimensional e

$$u_s^Y(n) = \frac{C_s^Y(n)}{(C_2^Y(n))^{s/2}} = n^{1-\frac{s}{2}} u_s^x, \quad (2.32)$$

Este decaimento mostra que para s fixo, $s \geq 3$ a contribuição relativa dos cumulantes superiores fica cada vez menor com o aumento de n . Já que independe da escala, isso vale para Z e W também (verifique).

Exercício

Calcule os cumulantes para a distribuição normal $\mathcal{N}(\mu, \sigma)$, ou seja $P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Calcule a função característica. É óbvio que $C_1 = \mu$ e $C_2 = \sigma^2$. Mostre que $C_s = 0$ para $s \geq 3$. Segue que as quantidades adimensionais u_s são nulas para $s \geq 3$.

O que significa, frente a este resultado para a gaussiana, o decaimento de $u_s^Y(n) = n^{1-\frac{s}{2}} u_s^x$? De forma pedestre isto mostra que a distribuições de Y , Z e W estão ficando mais perto de uma gaussiana para n grande. E de forma não pedestre? Este é o tema da próxima secção.

2.4 O Teorema do Limite Central I

2.4.1 Aquecimento

Este teorema tem um papel central em qualquer discussão de probabilidades. Isto no entanto não é o que *central* quer dizer aqui. O termo se refere a que, nas condições da secção anterior, as distribuições de Y , Z ou W são aproximadamente gaussianas na parte *central* ou seja perto do valor máximo. Para aquecer vemos que pela equação 2.24, a função característica de Y é

$$\hat{P}(k|n) = \exp\left(\sum_{s=0}^{\infty} C_s^Y(n) \frac{(-ik)^s}{s!}\right) \quad (2.33)$$

e a distribuição e'

$$P(Y|N = n) = \int_{-\infty}^{\infty} \exp(iky - \sum_{s=0}^{\infty} C_s^Y(n) \frac{(-ik)^s}{s!}) \frac{dk}{\sqrt{2\pi}}. \quad (2.34)$$

Até aqui é exato. Desprezando os cumulantes de ordem superior a segunda, pela eq. 2.32

$$P(Y|N = n) = \int_{-\infty}^{\infty} \exp(iky + ikC_1^Y(n) - k^2 C_2^Y(n)/2) \frac{dk}{\sqrt{2\pi}} \quad (2.35)$$

Chamemos $\langle x \rangle = \mu$ e $\sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2$. Realizando a transformada de Fourier

$$P(Y|N = n) = \frac{1}{\sqrt{2\pi n\sigma_x^2}} e^{-\frac{(y-n\mu)^2}{2n\sigma_x^2}} \quad (2.36)$$

Da mesma forma, e com o mesmo grau de rigor ou falta dele:

$$\begin{aligned} P(Z|N = n) &= \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(z-\sqrt{n}\mu)^2}{2\sigma_x^2}} \\ P(W|N = n) &= \frac{1}{\sqrt{2\pi\frac{\sigma_x^2}{n}}} e^{-\frac{(w-\mu)^2}{2\frac{\sigma_x^2}{n}}} \end{aligned} \quad (2.37)$$

Vemos que as distribuições são gaussianas e escrevemos as tres para mostrar que as diferentes formas de ajustar a escala da soma leva a que diferentes quantidades tenham um valor limite fixo ou que mude com alguma potência de n .

O leitor, neste ponto, deve se perguntar qual é a operação matemática de desprezar que nunca antes viu definida. Analisaremos isto nas próximas secções. Mas antes um exemplo onde o cálculo é exato.

Exercício

Mostre que os resultados acima (eqs. 2.36 e 2.37) são exatos no caso particular que a distribuição $P(x)$ é gaussiana:

$$P(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-\mu)^2}{2\sigma_x^2}}$$

Solução: O exercício anterior mostra que os cumulantes com $s \geq 3$ são nulos. Logo, não é necessário *desprezá-los*. Temos o resultado importante que somas de variáveis gaussianas tem distribuição gaussiana. Este é um exemplo de uma distribuição dita estável sob adições. Somas de variáveis gaussianas são gaussianas.

Exercício

Mostre que a distribuição de Cauchy $P(x) = \frac{1}{\pi} \frac{b}{x^2+b^2}$ é estável. Note que portanto a soma de variáveis de Cauchy não é gaussiana. Discuta primeiro a variância de x para ver onde os argumentos acima falham.

2.5 A média e a concentração em torno da média

A distribuição de W dada pela eq. 2.36 mostra que a média de W é igual à média de x . Isto não deve causar nenhuma surpresa, devido à linearidade da integral. Se os diferentes x_i forem considerados como diferentes medidas de X , então W pode ser entendido como a média empírica de X . Isto é o conteúdo da

lei fraca dos grandes números. Quanto se afasta a média empírica da média? Ou de outra forma, diferentes experiências levam a diferentes médias empíricas, qual é a probabilidade de que hajam flutuações grandes? Usemos a desigualdade de Chebyshev que pode ser obtida desta forma:

Considere $\epsilon > 0$ e pela equação 2.31, temos que $C_2^W(n) = \sigma_x^2/n$, satisfaz

$$\begin{aligned}
 C_2^W(n) &= \int_{-\infty}^{\infty} dw(w^2 - \langle w \rangle^2)P(W = w|N = n) \\
 &= \int_{-\infty}^{\infty} dw(w - \langle w \rangle)^2 P(W = w|N = n) \\
 &\geq \int_{|w - \langle w \rangle| \geq \epsilon} dw(w - \langle w \rangle)^2 P(W = w|N = n) \\
 &\geq \int_{|w - \langle w \rangle| \geq \epsilon} dw \epsilon^2 P(W = w|N = n) \\
 &\geq \epsilon^2 \text{Prob}(|w - \langle w \rangle| \geq \epsilon).
 \end{aligned} \tag{2.38}$$

onde usamos $\text{Prob}(|w - \langle w \rangle| \geq \epsilon) = \int_{|w - \langle w \rangle| \geq \epsilon} dw P(W = w|N = n)$ e chegamos à desigualdade de Chebyshev, que dá uma cota do decaimento com ϵ da probabilidade de ter flutuações maiores que ϵ :

$$\text{Prob}(|w - \langle w \rangle| \geq \epsilon) \leq \frac{C_2^W(n)}{\epsilon^2} \tag{2.39}$$

Mas $C_2^W(n)$ depende de maneira simples de n . Extraíndo esta dependência temos que a “probabilidade de que uma amostra de n valores $\{x_i\}$ que tenha uma média empírica $\langle w \rangle$ e que este valor se afaste do valor médio por mais que ϵ ”, isto é, $\text{Prob}(|w - \langle w \rangle| \geq \epsilon)$ está limitada por:

$$\text{Prob}(|w - \langle w \rangle| \geq \epsilon) \leq \frac{C_2^x}{n\epsilon^2} \tag{2.40}$$

As flutuações de w de tamanho maior que ϵ fixo, ficam mais improváveis quando n cresce.

O próximo exercício mostra de que forma a frequência de um evento esta relacionada com a probabilidade.

Exercício: frequência e probabilidade

Considere a seguinte informação $I =$ “Uma moeda é jogada para cima, bate no teto, no ventilador do teto, e cai no chão plano”. Há vários motivos para atribuir $p = 1/2$ à probabilidade que caia a cara para cima, isto é $p = P(s = 1|I) = 1/2$ e $q = P(s = -1|I) = 1/2$. Poderíamos considerar outra experiência I'^7 onde p, q tem outros valores (entre zero e um). Consideremos as jogadas independentes,

⁷por exemplo $I' =$ “Deixe a moeda, inicialmente de cara para cima e num plano horizontal, cair até a mesa, a partir de uma altura h , sem girar”. Considere $h = 1$ mm, $h = 1$ cm e $h = 1$ m.

para duas jogadas i e j quaisquer $P(s_i|s_j I') = P(s_i|I')$. Chame m o número de caras para cima, quando a moeda é jogada n vezes. A frequência de caras é definida por $f = m/n$

- (A) Mostre que a distribuição de m , é a distribuição binomial:

$$P(m|N = nI') = \frac{n!}{m!(n-m)!} p^m q^{n-m} \quad (2.41)$$

- (B) Calcule $\langle m \rangle$, $\langle m^2 \rangle$. [Dica: Use a expansão binomial de (i) $(p+q)^n$, (ii) $p \frac{\partial}{\partial p} p^m = mp^m$ e (iii) a normalização $p+q=1$; resposta: $\langle m \rangle = np$, $\langle m^2 \rangle = n^2 p^2 + np(1-p)$]
- (C) Refaça a dedução da desigualdade de Chebyshev para distribuições de variáveis que tomam valores discretos e mostre que para ϵ fixo, a probabilidade que a frequência f se afaste do *valor esperado* $\langle f \rangle = p$ por mais que ϵ , cai com $1/n$.
- (D) Discuta e pense: Então de que forma a frequência está ligada à probabilidade? A frequência *converge*, quando n cresce, para a probabilidade p . Toda convergência precisa ser definida em termos de uma distância, que vai para zero quando se toma algum limite. É fundamental entender que a distância aqui não é ϵ , mas é a **probabilidade** que f se afaste de p por mais de ϵ . Assim, a frequência f converge **em probabilidade** à probabilidade p .

A conclusão do exercício acima é fundamental. Como poderíamos definir probabilidades em termos de frequência, se para mostrar que a frequência está associada à probabilidade usamos o conceito de convergência em probabilidade? Discuta se é errado ou não definir um conceito usando esse conceito na definição.

Mas o exercício acima mostra porque pode parecer sedutor usar a frequência em lugar da probabilidade. Se tivermos informação I' sobre uma experiência e dados sobre uma sequência de experimentos nas condições I' podemos atribuir valor à probabilidade de forma mais segura. A frequência é informação que pode ser usado para atribuir um número à probabilidade, mas não é o único tipo de informação para fazer isso.

2.6 O Teorema do Limite Central II

Não há uma prova só, mais muitas, que refletem os objetivos em estudar este problema. Podemos olhar para diferentes condições sobre $P(x)$ e com isso mudar os resultados sobre a região central que é gaussiana e sobre quão grandes são os erros nas caudas das distribuições. Dependendo das condições, a região central vai depender de forma diferente do valor de n .

Esperamos pela eq. que a variável $Z = \frac{Y - N\mu}{\sqrt{N}\sigma}$ tenha distribuição normal de média nula e variância 1, pelo menos na região *central*.

Podemos transladar a origem de x e tornar $\mu = 0$.

Teorema LC

(Kinchin) Suponhamos que existam A, a, b, c e d constantes positivas tal que

- $dP(x)/dx$ é contínua
- $\int |dP(x)/dx| dx < A$
- $a < \langle x^2 \rangle = \sigma^2 < b$
- $\langle |x^3| \rangle < b$
- $\langle x^4 \rangle < b$
- $\langle |x^5| \rangle < b$
- $|\hat{P}(k)| > d$ para $|k| < c$
- Para cada intervalo (k_1, k_2) , com $k_1 k_2 > 0$, existe um número $\rho(k_1, k_2) < 1$, tal que para $k_1 < k < k_2$ temos

$$|\hat{P}(k)| < \rho.$$

Então

- Na região central, definida por $|x| < 2 \log^2 n$

$$P(Y|N = n) = \frac{1}{\sqrt{2\pi n\sigma^2}} e^{-\frac{y^2}{2n\sigma^2}} + \frac{S_n + yT_n}{(n\sigma^2)^{5/2}} + O\left(\frac{1 + |x|^3}{n^2}\right)$$

onde S_n e T_n são independentes de y e não crescem mais rápido que n .

- Para y arbitrário

$$P(Y|N = n) = \frac{1}{\sqrt{2\pi n\sigma^2}} e^{-\frac{y^2}{2n\sigma^2}} + O\left(\frac{1}{n}\right)$$

A prova é razoavelmente simples e pode ser encontrada no Apêndice de [Kinchin]. O leitor poderá ver que a essência da prova está no controle dos termos superiores da expansão de Taylor da equação 2.34 que foram *desprezados* anteriormente para chegar até a equação 2.35. Pense na diferença entre *desprezar* e *controlar*. Muitas vezes, em Física *desprezamos sem controlar*, pois tentar *controlar* pode ser tão difícil que evitaria a possibilidade de avanço. Uma vez que se encontra algo interessante, sempre podemos voltar atrás e tentar *controlar* termos antes *desprezados* usando a nova maneira de enxergar um problema, que o avanço menos cuidadoso permitiu.

2.7 O Teorema do Limite Central III

Apresentamos alguns exemplos para distribuições $P(x)$ simples.

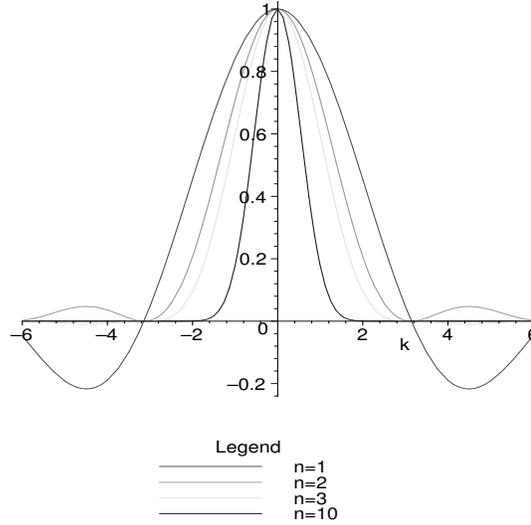


Figura 2.2: A função característica $\hat{P}(k|N = n)$ para a soma de n variáveis uniformemente distribuídas.

2.7.1 A distribuição uniforme

$P(x) = 1/L$ para $-L/2 < x < L/2$ e 0 para outros valores de x . A função característica

$$\hat{P}(k|1) = \frac{1}{L} \int_{-L/2}^{L/2} e^{-ikx} dx = \frac{2}{kL} \sin\left(\frac{kL}{2}\right) \quad (2.42)$$

$$P(Y = y|N = n) = \int_{-\infty}^{\infty} [\hat{P}(k|1)]^n e^{iky} \frac{dk}{2\pi} = \int_{-\infty}^{\infty} \left[\frac{2}{kL} \sin\left(\frac{kL}{2}\right)\right]^n e^{iky} \frac{dk}{2\pi} \quad (2.43)$$

A figura 2.2 mostra que a função característica fica mais parecida com uma gaussiana e na figura 2.3 vemos que efetivamente o $\log(|\hat{P}(\sqrt{|u|}|N = n)|)$ com $u = \pm k^2$ fica cada vez mais perto de $-c|u|$ (gaussiana).

2.7.2 A distribuição exponencial

A distribuição $P(x) = \Theta(x)ae^{-xa}$ é chamada de exponencial. Mostre que $\mu = \sigma^2 = a^{-1}$. A função característica

$$\hat{P}(k|1) = \int_0^{\infty} ae^{-xa} e^{-ikx} dx = \frac{a}{a + ik} \quad (2.44)$$

$$P(Y = y|N = n) = \int_{-\infty}^{\infty} \left(\frac{a}{a + ik}\right)^n e^{iky} \frac{dk}{2\pi} \quad (2.45)$$

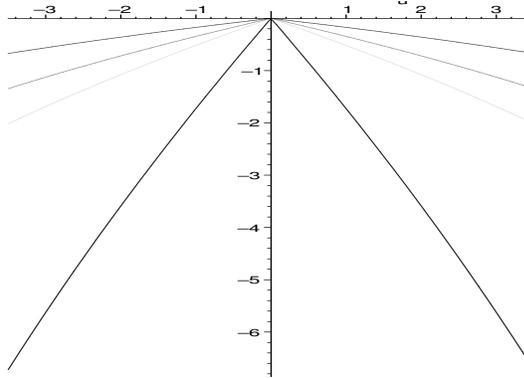


Figura 2.3: A função característica $\log(|\hat{P}(\sqrt{|u||N=n|})|)$ contra $u = \pm k^2$ para a soma de n variáveis uniformemente distribuídas. Nesta representação gaussianas aparecem como retas. A legenda é a mesma da figura 2.2

Integrando por partes ($u = e^{iky}$, $dv = (\frac{a}{a+ik})^l dk$, para $l = n, n-1, \dots, 1$) obtemos:

$$P(Y = y|N = n) = \Theta(y) a^n \frac{y^{n-1} e^{-ay}}{(n-1)!}. \quad (2.46)$$

Faça a conta. Obviamente não é uma gaussiana, mas uma distribuição gamma. No entanto, a região central sim, se parece com uma gaussiana.

A figura 2.4 mostra que a distribuição para n baixo não se parece em nada com uma gaussiana, mas à medida que n aumenta fica mais parecida com uma gaussiana, figura 2.5. Note que as distribuições, nessa figura são claramente assimétricas. Pense no que significa que a distribuição resultante seja gaussiana se as variáveis somadas são sempre positivas e portanto $Y > 0$ sempre. Esse é o significado de *central*, nas caudas não dizemos nada.

2.7.3 A distribuição binomial revisitada

A distribuição de Bernoulli é dada por $P(x) = p\delta(x-1) + q\delta(x+1)$. O número de aplicações que usaram esta distribuição é enorme. Só para ter uma ilustração em mente, podemos pensar em jogadas de uma moeda, ou um passo dado por um bêbado numa caminhada unidimensional. Se há N repetições ($i = 1 \dots N$) e $P(x_i)$ é a mesma para todo i e $P(x_i|x_j) = P(x_i)$ para qualquer $i \neq j$, e queremos $P(Y|N)$ para $Y = \sum_{i=1..n} x_i$. Este é exatamente nosso exemplo acima sobre a distribuição binomial onde estudamos a relação entre frequência e probabilidade. Aqui há um pequeno problema. A distribuição de probabilidades binomial deve ser comparada com a densidade de probabilidade gaussiana. Note que se N é par a probabilidade de que ocorra um valor de Y ímpar é zero, ou seja $\Delta Y = 2$. Ao apresentar os gráficos da figura 2.7 a binomial foi dividida por ΔY . De outra

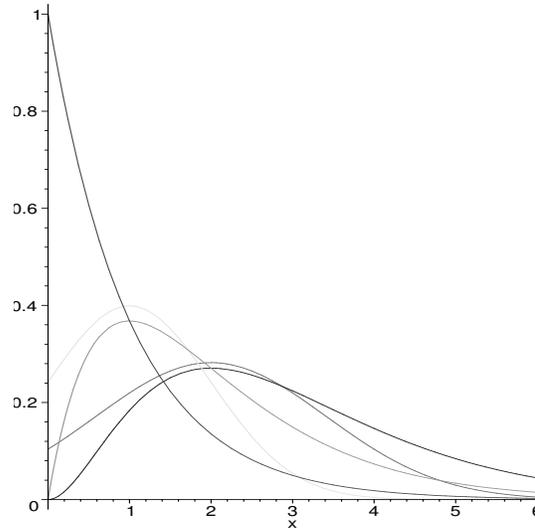


Figura 2.4: A densidade de probabilidade $P(y|N = n, a = 1)$ para a soma de n variáveis exponencialmente distribuídas, $n = 1, 2, 3$. Para as duas últimas mostramos as gaussianas com $\mu = \sigma^2 = n - 1$

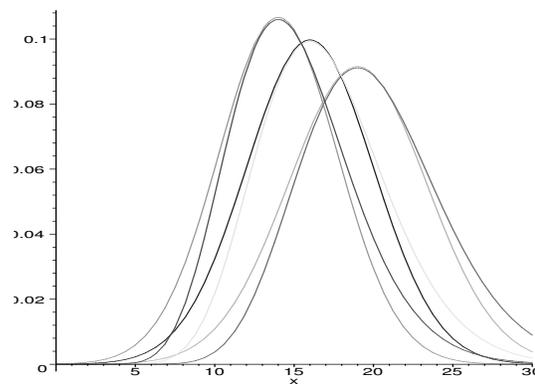


Figura 2.5: A densidade de probabilidade $P(y|N = n, a = 1)$ para a soma de n variáveis exponencialmente distribuídas, $n = 15, 17, 20$. Junto estão mostradas as gaussianas com $\mu = \sigma^2 = n - 1$.

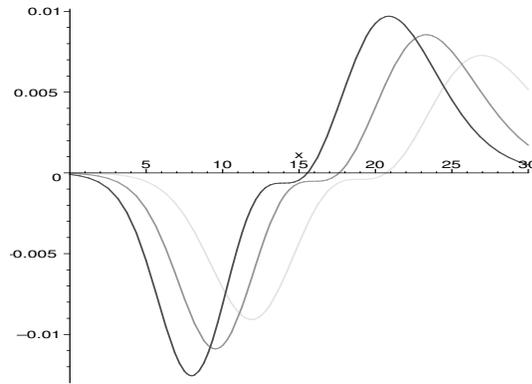


Figura 2.6: A diferença entre a densidade de probabilidade $P(y|N = n, a = 1)$ para a soma de n variáveis exponencialmente distribuídas, $n = 15, 17, 20$ e as gaussianas com $\mu = \sigma^2 = n - 1$. Os mesmos parâmetros da figura 2.5. Note que a região central é bem aproximada. Há uma região de transição, ao afastar-se para as caudas, e finalmente as caudas vão rapidamente para zero, assim como a sua diferença

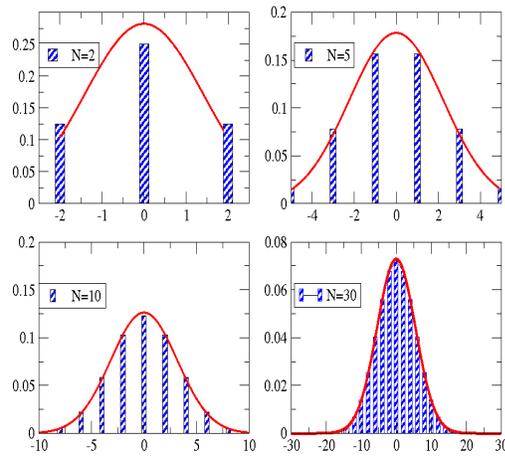


Figura 2.7: A binomial (dividida por $\Delta Y = 2$, barras) e a densidade gaussiana correspondente (linha contínua), para $N = 2, 5, 10$ e 30

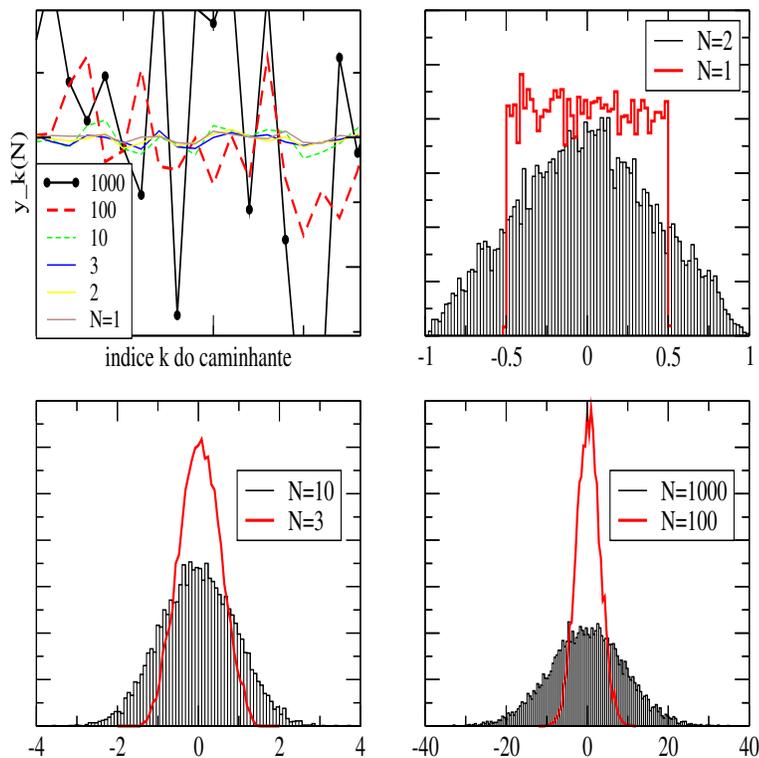


Figura 2.8:

forma: a probabilidade da binomial que Y tenha um dado valor num intervalo $(y, y + 2)$ é aproximado pela integral da gaussiana entre y e $y + 2$.

2.7.4 Caminho Aleatório

Novamente olhamos para a distribuição binomial. Olhe para a figura 2.8. Definimos o caminho aleatório através de

Difusão: $K(= 10000$ na figura 2.8) seqüências de N passos de um processo binomial, definidos por

$$y_n = y_{n-1} + x_n \quad (2.47)$$

onde $P(x = 1) = p$ e $P(x = -1) = q = 1 - p$. O índice n pode ser interpretado como tempo numa dinâmica discreta, a cada intervalo de tempo Δt uma partícula se desloca uma quantidade x . O deslocamento total tem probabilidade dada pela binomial. Mostramos agora que para valores altos de n a binomial se parece com uma gaussiana.

O resultado necessário é a aproximação de Stirling para o fatorial,

$$\log N! = \left(N + \frac{1}{2}\right) \log N - N + \frac{1}{2} \log 2\pi + \log \left(1 + \frac{1}{12N} + \frac{1}{288N^2} + \mathcal{O}\left(\frac{1}{N^3}\right)\right) \quad (2.48)$$

que é uma expansão assintótica, isto é melhora quando N aumenta ⁸. Não precisamos todos esses termos, basta $\log N \approx N \log N - N$ onde desprezamos $\mathcal{O}(\log N)$

$$\begin{aligned} \log P(m|N = nI') &= \log n! - \log m! - \log(n-m)! + m \log p + (n-m) \log q \\ &= n \log n - n - m \log m + m - (n-m) \log(n-m) \\ &\quad + n - m + m \log p + (n-m) \log q \end{aligned} \quad (2.49)$$

Podemos tratar m como uma variável real e encontrar onde a probabilidade $P(m|N = nI')$ atinge o valor máximo. Tomamos a primeira e segunda derivadas

$$\begin{aligned} \frac{\partial \log P(m|N = nI')}{\partial m} &= -\log m + \log(n-m) + \log p - \log q \\ \frac{\partial^2 \log P(m|N = nI')}{\partial m^2} &= -\frac{1}{m} - \frac{1}{n-m} \end{aligned}$$

Temos que em $m^* = np$ a probabilidade máxima e nesse ponto a segunda derivada vale $= \frac{1}{npq}$. A expansão de Taylor até segunda ordem do $\log P(m|N = nI')$ nos leva a uma gaussiana para $P(m|N = nI')$ de forma óbvia (qualquer expansão até segunda ordem é quadrática). O que deve ser verificado é se essa expansão faz sentido. Esperamos, pela equação 2.32, que os termos superiores decaiam com n . Verifiquemos isso explicitamente para o termo cúbico (proporcional ao terceiro cumulante) e superiores:

$$\frac{\partial^3 \log P(m|N = nI')}{\partial m^3} = \frac{1}{m^2} - \frac{1}{(n-m)^2}$$

e notamos que as derivadas de ordem superior aumentarão o decaimento dos cumulantes.

Vemos que a distribuição binomial é bem aproximada por

$$P(m|N = nI) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2\sigma^2}(m - m^*)^2 \quad (2.50)$$

onde repetindo $m^* = np$ e $\sigma^2 = npq$. Repetimos que um dos pontos importantes é a dependencia $\sigma \propto n^{\frac{1}{2}}$.

2.8 Apêndice: Stirling

Considere a função Gama, definida para nossos propósitos imediatos sobre os números reais positivos:

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt, \quad (2.51)$$

⁸Jeffreys, note quão boa é a aproximação $1! = 0.9221$ sem o termo $1/12N$ e $1! = 1.002$ com esse termo. Para $2! = 1.9190$ e 2.0006 respectivamente.

que podemos estender por continuação analítica para o plano complexo ($x \neq$ inteiro menor que 1). Integre por partes e mostre que $\Gamma(x+1) = x\Gamma(x)$. Mostre que $\Gamma(1) = 1$ e portanto $\Gamma(n+1) = n!$. Mostre também para uso futuro que $\Gamma(1/2) = \sqrt{\pi}$ fazendo a mudança de variável $t = u^2/2$.

Podemos escrever

$$\Gamma(x+1) = \int_0^{\infty} e^{-t} t^x dt = \int_0^{\infty} e^{f(t;x)} dt \quad (2.52)$$

onde $f(t;x) = x \log t - t$. Para x fixo, o valor máximo de f ocorre em $t_{max} = x$. A derivada segunda com respeito a t , nesse ponto é $f''(t;x) = -1/x$ e as derivadas superiores, de ordem k caem com x^{1-k} . Para valores de $t \gg x$ o integrando morre rapidamente e para valores grandes de x , f é bem aproximado por

$$f(t;x) = x \log x - x - \frac{1}{2x}(x-t)^2. \quad (2.53)$$

Supondo que não é uma aproximação muito ruim desprezar as derivadas superiores,

$$\begin{aligned} \Gamma(x+1) &\approx e^{x \log x - x} \int_{-x}^{\infty} e^{-\frac{1}{2x}(x-t)^2} dt \\ &\approx e^{x \log x - x} \int_{-\infty}^{\infty} e^{-\frac{1}{2x}(x-t)^2} dt \\ &= \sqrt{2\pi x} e^{x \log x - x} \end{aligned} \quad (2.54)$$

onde estendemos o limite de integração até $-\infty$ porque devido ao decaimento gaussiano não há contribuições importantes nessa região. Assim temos

$$\log n! = \log \Gamma(n+1) = \left(n + \frac{1}{2}\right) \log n - n + \frac{1}{2} \log 2\pi, \quad (2.55)$$

mostrando de forma intuitiva os primeiros termos da equação 2.48. Também é intuitivamente razoável que a aproximação deva ser melhor para valores maiores de x , dado que as derivadas superiores desprezadas são menores: $f'''(t_{max})/f''(t_{max}) = 2/t_{max} = 2/x$. Estas considerações podem ser provadas, assim como as correções da equação 2.9, mas o leitor deverá procurar outros textos (e.g. [?])

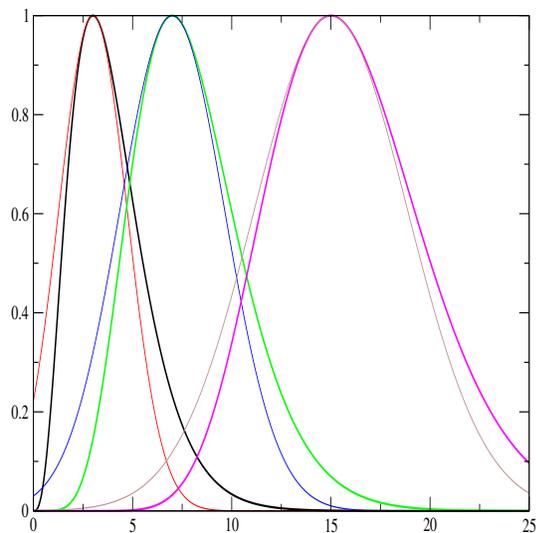


Figura 2.9: A figura mostra três pares de gráficos, para $x = 3, 7$ e 15 respectivamente. Cada par é formado pelas funções $e^{-t^x}/(x^x e^{-x})$ e $\exp(-(x-t)^2/(2x))$

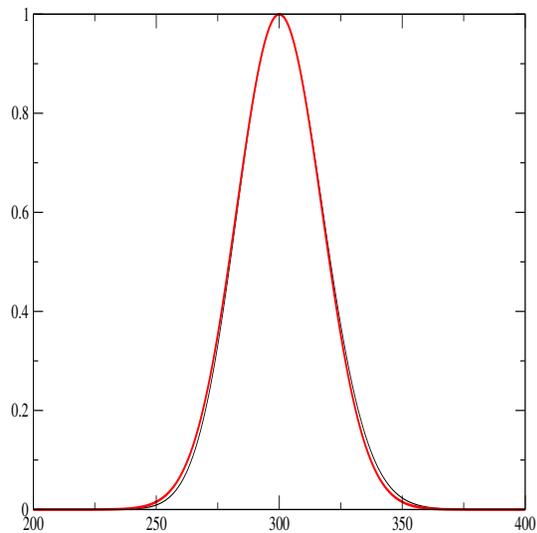


Figura 2.10: O mesmo que a figura 2.9 para $x=300$. Note aqui, assim como na figura anterior, que ao usar a gaussiana que é simétrica em relação ao máximo, o erro na integral cometido à esquerda do máximo tem sinal oposto ao cometido à direita.

Capítulo 3

Entropia

I wish to propose for the reader's favourable consideration a doctrine which may, I fear, appear wildly paradoxical and subversive. The doctrine in question is this: that it is undesirable to believe in a proposition when there is no ground whatever for supposing it true.

Bertrand Russell, in *Sceptical Essays*

É inquestionável a capacidade de Russell de criar frases de efeito para começar um livro. Como discordar da frase acima? “É indesejável acreditar numa proposição quando não há nenhuma base para supô-la verdadeira”. Parece óbvia, mas ele usa algumas centenas de páginas para mostrar que a doutrina é efetivamente alheia à forma de pensar e agir das pessoas em geral e por isso ele está justificado em chamá-la paradoxal e subversiva. Ainda assim, parece óbvia para quem quer uma descrição científica de alguma área de conhecimento. É interessante notar as conseqüências que ela tem. Russell as analisa do ponto de vista das relações humanas. Shannon, Jaynes e toda a escola de teoria de informação o fazem do ponto de vista das conseqüências matemáticas que essa doutrina tem quando devemos raciocinar sob a luz de informação incompleta. Esse é o tema deste capítulo.

Especificamente queremos encontrar um método para atribuir números às probabilidades satisfazendo vínculos impostos por informação dada. Este problema está relacionado à atualização das probabilidades para novos números devido à inclusão de nova informação. Novamente vamos pensar em casos particulares, que sendo suficientemente simples permitam expressar desejos de como deveria ser a teoria. Há várias formas de expressar informação e isso terá conseqüências sobre os métodos para atualizar probabilidades. Começamos descrevendo algo aparentemente simples, quando a informação não faz diferença entre as diferentes possibilidades. A partir dessa primeira atribuição continuamos em direção a um método geral. O resultado será que a cada distribuição de probabilidades será atribuído um número que indica quanta informação a mais seria necessário coletar para ter informação completa. Dentre aquelas distribuições que satisfazem os vínculos da informação conhecida como verdadeira, escolhe-

mos a que é mais ignorante e portanto faz menos suposições não necessárias. A determinação desta distribuição requer o uso de algumas técnicas matemáticas. Nesse ponto o estudante deverá consultar algum livro de cálculo ou o Apêndice sobre o uso dos multiplicadores de Lagrange no cálculo de extremos sujeitos a vínculos. Nas últimas seções do capítulo começaremos a estudar sistemas físicos e fazer a conexão com a Termodinâmica.

3.1 Simetria

Um experimento é descrito pela informação contida em $I_1 =$ “Suponha que temos uma moeda com duas faces, que descrevemos pela variável $\sigma = \{\pm 1\}$. O valor $\sigma = 1$ está associado à cara e $\sigma = -1$ à coroa. Jogo a moeda para cima, bate no ventilador do teto, e cai num lugar onde não podemos no momento ver o resultado.”

Consideremos o seguinte jogo J_1 . Se $\sigma = 1$ você ganha e eu perco. Do contrário, eu ganho. Eu aposto um feijão. Quanto você estaria disposto a apostar?¹ A resposta tem relação, para pessoas racionais, que não dependem do feijão para sobreviver², com as probabilidades $P(\sigma = 1|IJ_1)$ e $P(\sigma = -1|IJ_1)$ que você atribui com base na informação I que inclui todo o que se sabe sobre a moeda e a forma como foi jogada.

Suponhamos que o jogo fosse totalmente diferente, J_2 : Introduzimos a nova variável $\sigma' = -\sigma$, você ganha se $\sigma' = -1$. Quanto você estaria disposto a apostar contra o meu feijão? Note que temos uma informação que nos permite associar a este jogo probabilidades

$$\begin{aligned} P(\sigma' = 1|IJ_2) &= P(\sigma = -1|IJ_1) \\ P(\sigma' = -1|IJ_2) &= P(\sigma = 1|IJ_1). \end{aligned} \quad (3.1)$$

Isto é, podemos usar as equações acima para transformar o que se sabe de um jogo no outro e estas são válidas para qualquer I incluindo I_2 descrita abaixo.

Suponha agora que a informação que temos I_2 , que difere de I por muito pouco, nos leve a ser **indiferentes** sobre que jogo está sendo jogado. Suponha que não faça diferença dado I_2 se é $\sigma = 1$ que está associada à cara ou se é $\sigma' = 1$ que o está. Isto é o “+1” esta pintado numa dos lados da moeda e “-1” está pintado no outro, mas não sabemos qual. É claro que as equações 3.1 ainda valem com I_2 no lugar I . Mas agora há uma simetria entre os jogos J_1 e J_2 , que nos leva a concluir por consistência

$$\begin{aligned} P(\sigma' = 1|I_2J_2) &= P(\sigma = 1|I_2J_1) \\ P(\sigma' = -1|I_2J_2) &= P(\sigma = -1|I_2J_1). \end{aligned} \quad (3.2)$$

¹Jaynes não gosta de basear os fundamentos da teoria em algo tão vulgar como apostas por dinheiro. No entanto esperamos que qualquer noção *a priori* sobre apostas tenha evoluído por seleção natural onde as apostas amiúde não são por dinheiro mas sim pela própria vida.

²Este problema é talvez muito mais complicado pois não sabemos o que seja uma pessoa racional, mas simplesmente consideremos alguém que quer jogar e quer ganhar.

Note a sutil diferença entre as equações 3.1 (escrita com I_2 em vez de I) e 3.2. Dado que $P(\sigma = 1|IJ_1) + P(\sigma = -1|IJ_1) = 1$, devemos concluir que $P(\sigma = 1|IJ_1) = 1/2$ e $P(\sigma = -1|IJ_1) = 1/2$.

Porque tantas voltas para chegar ao óbvio? Por vários motivos. Em primeiro lugar notamos que este não é o único exemplo onde usaremos simetria. A história da Física mostra muitas generalizações do uso de simetria para atribuir probabilidades ou definir a dinâmica, o que não é totalmente diferente, pois dinâmica vem das interações e as interações estão relacionadas, como veremos adiante, com probabilidades condicionais e dependência. A idéia de analisar este caso simples deve-se a que as coisas vão ficar mais difíceis e é interessante se apoiar em casos simples.

Se tivéssemos um dado de n faces, σ toma valores de 1 a n , teríamos chegado a $P(\sigma = i|I) = 1/n$, a distribuição uniforme. Note que esta atribuição tem a ver com a simetria da nossa informação sobre o experimento do dado e não é postulada *a priori*. Generalize o problema para n estados.

Este método de atribuição de probabilidades parece ter sido usado pela primeira vez por J. Bernuolli e posteriormente por Laplace. Recebe nomes como princípio da razão insuficiente ou da indiferença.

3.2 Incerteza, ignorância e Informação

Leia novamente a citação inicial deste capítulo. Suponha que sob dada informação I_A tenhamos que escolher entre diferentes distribuições de probabilidade que são igualmente compatíveis com os vínculos impostos pela informação I_A . Qual escolher? Para começar consideremos que temos só duas candidatas, $P_1(x_i)$ e $P_2(x_i)$. Suponha que $P_2(x_i)$, além de satisfazer os vínculos impostos por I_A também satisfaz àqueles impostos por informação adicional I_B , mas esta informação não é dada. Repetimos: só sabemos I_A . Qual das duas escolhemos?

Um critério poderia ser: “escolha P_2 , pois pode ser que I_B seja verdadeiro.” E se não for? Se você souber alguma coisa sobre a validade da asserção I_B claro que é bem vindo a usá-la. Mas se não souber, porque supor que é verdade? A escolha de P_2 é equivalente a assumir que I_B é verdadeira, mas isso não queremos fazer. Portanto parece natural que a escolha caia sobre P_1 . Esta é, das distribuições compatíveis com os vínculos, aquela que faz menos suposições não garantidas pela informação disponível. Ao fazer esta escolha fazemos a escolha da distribuição que representa a maior ignorância possível. A outra escolha assume como certo informação que pode não ser verdadeira. É melhor não saber do que achar que se sabe algo que está errado ³.

Se você ainda insiste em escolher a outra distribuição, mesmo sem poder oferecer a veracidade de I_B como argumento, será por motivos não racionais e esse método, por mais que você o ache interessante, estará fora do alcance de nosso análise. Talvez a escolha seja por que você gostaria que I_B fosse verdade. Isso é ideologia, capricho ou dogma e não nos interessa discutir agora.

³He who knows best knows how little he knows. Thomas Jefferson

O método que procuramos - seguindo Shannon - procura atribuir a cada distribuição de probabilidade $P(x_i)$ uma *medida da ignorância* que essa distribuição representa sobre a variável x .

Se soubermos, por exemplo, que $x = 1$ não temos nenhuma ignorância. Se não soubermos nada sobre x , a não ser que está é uma variável que toma valores num conjunto discreto $\{x_i\}$ de n membros, então por simetria, $P(x_i) = 1/n$. Note que não é por simetria do sistema físico, mas porque a informação que temos não distingue as diferentes possibilidades i , que são simétricas quanto às preferências. Este é o princípio da razão insuficiente de Laplace, que Boltzmann seguiu, e que Jaynes discute de forma detalhada.

Máxima Entropia: Uma vez atribuída essa medida, que é chamada de *entropia*, tomaremos a distribuição com o maior valor possível dentre aquelas que satisfazem as restrições da informação conhecida. Esta é a distribuição mais incerta ou a que faz menos hipóteses e portanto está menos arriscada a fazer hipóteses incorretas.

Suponha que temos informação, na forma de um vínculo, no seguinte problema. $I_A =$ "Uma moeda indestrutível de três lados ($s = -1, 0$ ou 1) foi jogada muitas vezes, batendo no ventilador que gira no teto, e o valor médio \bar{s} desse experimento é compatível com o valor 0 ."

É razoável atribuir valores $P(-1) = P(1) = 0$ e $P(0) = 1$? Se eu insistir que isso é razoável, a pergunta que você fará é: "porque foram eliminadas as possibilidades ± 1 ?" e ainda "o quê é que ele sabe que eu não sei?". Simplesmente dizer que eu não gosto de ± 1 não é argumento razoável. A mesma coisa pode ser dito para a escolha de outra distribuição. Porque esta e não aquela?

Como proceder? Procuramos um critério que dê a cada distribuição $P(x)$, uma medida da falta de informação para determinar o valor de x . Queremos um método geral. Em princípio não sabemos se é possível tal método. Novamente usaremos a idéia que se um método geral existe, então deve ser aplicável a casos particulares. Se tivermos um número suficientemente grande de casos pode ser que o método geral se mostre incompatível com esse grande número de casos particulares. Se não fizermos a tentativa não saberemos.

Quais são os casos particulares que devemos adotar para este programa? A medida $H[P]$, a entropia, deve satisfazer, em primeiro lugar a transitividade, pois queremos escolher uma distribuição como sendo mais preferível que outra e se P_1 é preferida ante P_2 que por sua vez é preferida ante P_3 , então P_1 deverá ser preferida ante P_3 . Satisfazemos isto impondo que

- (1) A cada distribuição de probabilidades p associamos um número real $H[p]$

É tão razoável que é até difícil imaginar ante que críticas deveríamos defender este 'caso particular'. $H[p]$ é uma função de todo o conjunto de valores $p = \{p_i\}$. Para variáveis X que tomam valores reais, a entropia será um funcional da densidade $p(x)$. Isto é, a cada função $p(x)$ será atribuído um número.

Pequenas mudanças na distribuição $\{p_i\}$ não devem levar a grandes mudanças em $H[P]$.

- (2) $H[p]$ deve ser contínua nos p_i .

No caso particular em que a informação é simétrica ante troca dos rótulos i , teremos $P(x_i) = 1/n$. Neste caso, por simplicidade notacional denotamos $H[1/n, 1/n, \dots, 1/n] = F(n)$. Suponha que temos dois problemas. No primeiro $n = 2$ e no segundo $n = 10000$. Quanta informação falta em cada problema para determinar o valor de x ? Não sabemos, mas é razoável supor que no segundo caso falta mais. Logo, a medida que buscamos deve satisfazer

- (3) $F(n)$ é uma função crescente de n .

Outro caso particular onde sabemos algo pois achamos que se não fosse assim, algo estranho estaria acontecendo, é nosso velho amigo: Não queremos ser manifestamente inconsistentes, portanto o mais consistente possível é impor:

- (4) Se há mais de uma forma de analisar uma situação, todas devem dar o mesmo resultado.

Isto ainda não é suficiente e é preciso colocar uma condição que é menos óbvia e que diz respeito às possíveis maneiras de analisar a incerteza frente a diferentes agrupamentos dos estados. Suponha um dado cúbico. Os estados possíveis do sistema, para simplificar, são rotulados pelo número de pontos na face que aponta para cima. Podemos agrupar em dois grupos e.g. (de pares ou ímpares), ou de (maiores ou menores que 3.5), ou de (primos ou não-primos). Outra forma poderia ser menos uniforme, e.g. de (maiores ou menores que 2.5) ou de muitas outras maneiras, e.g. em mais de dois grupos.

Suponha que X possa ter n valores possíveis $\{x_i\}_{i=1\dots n}$. Estes *microestados* são mutuamente exclusivos. Queremos atribuir as probabilidades p_i . Seja $\{m_g\}$ um conjunto de números inteiros positivos tal que $\sum_{g=1}^N m_g = n$ e que denotam o tamanho de agrupamentos de estados de X . Escolha o conjunto de valores $\{m_g\}$ e mantenha-o fixo durante a análise. Depois escolheremos outro conjunto e o manteremos fixo novamente. A probabilidade de que X tenha um valor dentro de um agrupamento g é $P_g = \sum_{i \in g} p_i$. Se for dado que está no grupo g , a probabilidade que esteja no estado i é $p(i|g) = \frac{p_i}{P_g}$. A incerteza associada à variável X pode ser medida em dois passos, o primeiro passo mede a incerteza de estar em um dos agrupamentos g , chamemos H_G esta entropia e o segundo, uma vez que está em g , sobre o estado i em particular esta entropia será chamada H_g .

A última imposição, que chamaremos de

- (5) aditividade sob agrupamento,

diz que a entropia dos dois passos é aditiva e portanto será $H_G + \sum_{g=1}^N P_g H_g$ que é a entropia do primeiro passo mais a média das entropias da incerteza associada a cada agrupamento g .

Isto, junto com a consistência do item (3) nos dá

$$H[p_i] = H_G[P_g] + \sum_{g=1}^N P_g H_g[p(i|g)] \quad (3.3)$$

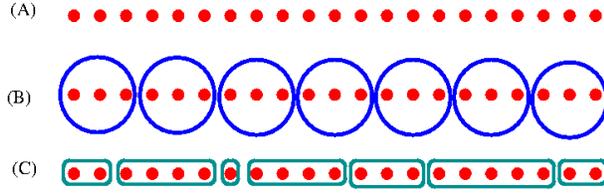


Figura 3.1: (A) Cada ponto (vermelho) representa um possível estado (mutuamente exclusivos) de uma variável. (B) Agrupamos em grupos com o mesmo tamanho. (C) Os agrupamentos são arbitrários.

Que isto é suficiente para determinar a forma funcional da entropia, será provado a seguir e será feito em dois passos, para cada um fazemos um dos dois tipos de agrupamento mencionados acima.

(I) Começamos analisando o caso particular (figura 3.1.B) em que a informação é simétrica para todos os estados i e os agrupamentos são do mesmo tamanho $m_g = m = n/N$ para todos os N agrupamentos g . Então: $p_i = 1/n$, $P_g = 1/N$ e $p(i|g) = N/n = 1/m$

$$H[1/n \dots 1/n] = H_G[1/N, \dots, 1/N] + \sum_{g=1}^N \frac{1}{N} H_g[1/m, \dots, 1/m] \quad (3.4)$$

como todos os termos lidam com entropias de distribuições uniformes, podemos introduzir a função monotônica desconhecida F :

$$F(n) = F(N) + \sum_{g=1}^N \frac{1}{N} F(m) \quad (3.5)$$

e, dado que $n = mN$, obtemos a equação funcional:

$$F(mN) = F(N) + F(m) \quad (3.6)$$

A solução óbvia é dada por

$$F(n) = k \log n. \quad (3.7)$$

Mas esta solução não é única a não ser que usemos o fato que $F(n)$ é monotônica pela imposição (2). A constante k não é muito importante neste estágio pois mede a escala das unidades, mudanças de k equivalem a mudanças na base do logaritmo que não alteram em nada a ordem de preferências de uma distribuição sobre outra. Mais para a frente veremos dentro da Mecânica Estatística, que tais mudanças equivalem a mudanças na escolha da escala de temperatura e nesse contexto poderá ser interpretada como k_B a constante de Boltzmann.

(II) Ainda não obtivemos a forma de H no caso geral. Passamos a analisar o caso em que os tamanhos dos agrupamentos m_g são arbitrários (figura 3.1.C), salvo o fato $\sum_{g=1}^N m_g = n$, mas ainda temos que $p_i = 1/n$ é uniforme.

Temos então que a probabilidade $P_g = m_g/n$ é arbitrária, mas a probabilidade condicional dentro de cada grupo é uniforme: $p(i|g) = 1/m_g$

$$H[1/n, \dots, 1/n] = H_G[P_g] + \sum_{g=1}^N P_g H_g[1/m_g, \dots, 1/m_g] \quad (3.8)$$

e substituímos a entropia da distribuição uniforme:

$$F(n) = H_G[P_g] + \sum_{g=1}^N P_g F(p(i|g)), \quad (3.9)$$

então a entropia da distribuição de probabilidades arbitrária P_g é dada por

$$H_G[P_g] = F(n) - \sum_{g=1}^N P_g F(m_g). \quad (3.10)$$

Introduzimos um 1 através de $1 = \sum_{g=1}^N P_g$, substituímos $F(n) = k \log n$

$$H_G[P_g] = \left(\sum_{g=1}^N P_g \right) k \log n - k \sum_{g=1}^N P_g \log(m_g) \quad (3.11)$$

$$H_G[P_g] = -k \sum_{g=1}^N P_g \log(m_g/n) \quad (3.12)$$

e finalmente

$$H_G[P_g] = -k \sum_{g=1}^N P_g \log(P_g) \quad (3.13)$$

que é a forma da entropia de Shannon, mas pode ser chamada de entropia de Y, onde Y é um subconjunto de nomes extraídos de { Shannon, Wiener, Boltzmann, Gibbs, Jaynes }, o que torna a vida dos historiadores da ciência mais interessante.

Deve ficar claro que não há nenhuma controvérsia quanto à utilidade deste formalismo. No entanto há muita controvérsia sobre a interpretação desta fórmula em geral, da necessidade dos *axiomas*, da suficiência, de se o nome de Boltzmann e Gibbs, associado à entropia termodinâmica deveria ser associado neste ponto a esta forma. Há também discussões atuais sobre o efeito acarretado pela mudança de um ou outro dos casos particulares que usamos. Sobre se estes deveriam ser chamados de axiomas. Sobre como generalizar isto para o caso em que as variáveis x tomam valores em intervalos reais. Nesse caso a idéia de uniformidade fica relativa à escolha do sistema de coordenadas. A teoria fica muito mais interessante e a invariância antes transformações gerais de coordenadas leva as discussões para o contexto da geometria diferencial.

A contribuição de Boltzmann e Gibbs será discutida no próximo capítulo e não entraria aqui em um par de parágrafos. De Shannon e Wiener ⁴ vem idéia de discutir informação do ponto de vista de volume, quantidade com o objetivo de entender limitações para a transmissão em canais de comunicação. Isto inclui como codificar uma mensagem, comprimindo-a e como, no outro lado do canal de comunicação, reobté-la. Não foi discutida a qualidade da informação ou a utilidade da informação: se fosse, toda a teoria não teria utilidade para descrever por exemplo meios como a televisão.... Isto não é totalmente piada. Se olharmos uma máquina de processamento de informação, como por exemplo um sistema nervoso de um animal, o conceito do valor da informação toma uma posição muito mais central. Há vantagens evolutivas em realizar inferência de uma forma frente a outra. Mais sobre tudo isso em aulas posteriores ⁵.

Uma contribuição importante de Jaynes foi a percepção que a maximização da entropia sujeita aos vínculos da informação leva a um método fundamental de inferência, chamado por ele de MaxEnt. Outros nomes estão associados a extensão de idéias de entropia como método de inferência fundamental ⁶. Em particular tem se conseguido nos últimos anos uma formulação muito mais satisfatória e menos ad hoc dos princípios por trás deste método.

A generalização do método de máxima entropia para inferência em problemas com variáveis em variedades contínuas leva a que o método não é tanto para a atribuição de números mas de uma atualização frente à chegada de nova informação. De novo aparece uma densidade de probabilidades *a priori* $p_0(x)$ que era o melhor que podia ser feito antes de levar em conta a nova informação. A nova densidade de probabilidades $P(x)$ será determinada pelo máximo do funcional

$$H_G[P|p_0] = -k \int dx P(x) \log\left(\frac{P(x)}{p_0(x)}\right) \quad (3.14)$$

sujeito aos vínculos impostos pela nova informação. A história desta entropia também é extensa. Uma mostra disto é que recebe vários nomes: entropia cruzada, (menos) distância (ou número ou divergência) de Kullback-Leibler. Não é uma distância, pois por exemplo não é simétrica. No entanto se P e p_0 estiverem “muito perto” então sim poderemos introduzir uma distância entre densidades de probabilidade e a partir de ai um tensor métrico e uma geometria. A posição de Jaynes era que MaxEnt serviria para *atribuir* probabilidades, pois chega a números sem menção de nenhuma probabilidade *a priori*. No entanto o papel da distribuição uniforme, que veio através da função $F(n)$, é fundamental. A entropia de Shannon é na realidade um método de *atualizar* probabilidades a partir da distribuição uniforme, mesmo que não o diga explicitamente. No resto do curso de Mecânica Estatística faremos uso da forma simplificada.

Exercício Mostre que $H_{SJ} = -\sum_i p_i \log p_i$ difere por uma constante de

⁴Aparentemente Alan Turing também teve estas idéias durante a 2da Guerra Mundial enquanto tentava decifrar códigos, mas esse trabalho teve menos influência porque era confidencial. Ver livro de Good.

⁵Não cabe no curso introdutório de Mecânica Estatística

⁶ver A.C

$H[p|q] = -\sum_i p_i \log \frac{p_i}{q_i}$, se $q_i = 1/N$ é constante e portanto o valor das distribuições que as extremizam coincidem.

3.3 Exemplos

3.3.1 O problema variacional: variáveis discretas

Seja dado que a variável X toma um entre n valores e temos a informação que o valor médio de X é conhecido e dado por $\langle X \rangle = \mu$. Qual é a distribuição de probabilidades de X que reflete a informação que temos e faz o menor número de hipóteses adicionais?

Queremos encontrar $\{P_i\}$ que maximize $H[P]$ sujeito a

$$\sum P_i = 1 \quad (3.15)$$

$$\sum x_i P_i = \mu \quad (3.16)$$

Introduzimos os vínculos através dos multiplicadores de Lagrange. Procuramos o máximo de

$$H[P] - \lambda_1 \left\{ \sum x_i P_i - \mu \right\} - \lambda_0 \left\{ \sum P_i - 1 \right\} \quad (3.17)$$

então olhamos para o seguinte problema variacional (tomamos a liberdade de fazer $k = 1$):

$$\delta \left[-\sum P_i \log P_i - \lambda_1 \left\{ \sum x_i P_i - \mu \right\} - \lambda_0 \left\{ \sum P_i - 1 \right\} \right] = 0 \quad (3.18)$$

O que significa o símbolo δ ? Fazemos mudanças $P_i \rightarrow P_i + \delta P_i$, tratamos as variações δP_i como independentes, e finalmente escolhemos os multiplicadores de Lagrange para que os vínculos sejam satisfeitos. Temos três termos que olharemos separadamente:

$$(1) \quad \delta H[P] = H[P + \delta P] - H[P] = -\sum_i \delta P_i [\log P_i + 1] \quad (3.19)$$

$$(2) \quad \delta(-\lambda_1 \left\{ \sum_i x_i P_i - \mu \right\}) = -\lambda_1 \sum_i x_i \delta P_i \quad (3.20)$$

$$(3) \quad \delta(-\lambda_0 \left\{ \sum_i P_i - 1 \right\}) = -\lambda_0 \sum_i \delta P_i \quad (3.21)$$

Juntando os três termos:

$$-\sum_i \delta P_i \{ \log P_i + 1 + \lambda_1 x_i + \lambda_0 \} = 0 \quad (3.22)$$

Mas fazemos as variações δP_i independentes entre si, de tal forma que se a soma na equação acima for zero, o será termo a termo:

$$\log P_i + 1 + \lambda_1 x_i + \lambda_0 = 0 \quad (3.23)$$

e segue que

$$P_i = e^{-1-\lambda_0-\lambda_1 x_i} \quad (3.24)$$

Os vínculos determinam os multiplicadores de Lagrange. Definimos Z , que será uma grandeza central em todo desenvolvimento futuro, $Z = e^{1+\lambda_0}$, que fica determinado pelo vínculo da normalização:

$$Z(\lambda_1) = \sum_i e^{-\lambda_1 x_i} \quad (3.25)$$

Para ilustrar uma utilidade de Z , chamada *função de partição*, notamos que:

$$\frac{d \log Z}{d \lambda_1} = -\frac{1}{Z} \sum_i x_i e^{-\lambda_1 x_i} \quad (3.26)$$

de onde segue que o valor esperado de X é dado por

$$\langle x \rangle = -\frac{d \log Z}{d \lambda_1} \quad (3.27)$$

Esta é uma equação para λ_1 , dado que o resto é conhecido: os valores x_i que a variável X pode tomar, e o seu valor médio, que é o vínculo imposto nas distribuições $P(x)$.

3.3.2 Exemplo: Variável de dois estados

Uma variável binária, chamada de Bernoulli, tem dois valores possíveis. A informação disponível não permite distinguir entre os dois estados. Qual é a distribuição de probabilidade $P(x)$? Já sabemos a resposta pelo princípio de razão insuficiente de Laplace. E pelo método de máxima entropia? Fazemos a variação da entropia, sujeita ao único vínculo de normalização:

$$\delta \left[-\sum P_i \log P_i - \lambda_0 \left\{ \sum P_i - 1 \right\} \right] = 0 \quad (3.28)$$

o que leva a $\log P_i = 1 + \lambda_0$ note que o resultado é que a probabilidade é independente do estado i , como esperado. Escolhendo o multiplicador de Lagrange, temos a resposta $P_1 = P_2 = 1/2$. (Você precisou fazer a conta?).

Se o vínculo $P_1 + P_2 = 1$ for explicitamente introduzido na expressão da entropia, teremos, fazendo $P_1 = p$ e $P_2 = 1 - p$

$$H(p) = -p \log p - (1 - p) \log(1 - p) \quad (3.29)$$

Desenhe a função e verifique que o máximo está em $p = 1/2$. Note que $\lim_{p \rightarrow 0^+} p \log p = 0$ e $\lim_{p \rightarrow 1^-} (1 - p) \log(1 - p) = 0$.

Não há muito o que fazer se for dada informação do tipo “o valor de $\langle x \rangle = m$ ”, pois não há mais que uma distribuição que satisfaz o vínculo.

3.3.3 3 estados

Suponha que seja dado “ X toma valores 0, 1 ou 2 e o valor médio de x , é conhecido: $\langle x \rangle = m$ ”. Determine a distribuição de máxima entropia compatível com o vínculo. Faça o problema. Resultado

$$P_i = \frac{e^{-\lambda_1 x_i}}{Z(\lambda_1)} \quad (3.30)$$

com λ_1 determinado pela eq. 3.27 e Z pela eq. 3.25 . Portanto, chamando $u = \exp(-\lambda_1)$,

$$Z = 1 + u + u^2 \quad (3.31)$$

e

$$P_i = \frac{u^i}{1 + u + u^2} \quad (3.32)$$

$$m = \frac{u + 2u^2}{1 + u + u^2}$$

$$u^2(m - 2) + u(m - 1) + m = 0 \quad (3.33)$$

$$u = \frac{1 - m \pm \sqrt{6m - 3m^2 + 1}}{2(m - 2)} \quad (3.34)$$

A equação 3.33 tem duas raízes. Discuta a escolha da solução correta. A figura 3.2 mostra o valor máximo da entropia H como função de m , e do lado esquerdo as probabilidades. Note que para o gráfico de $H[P]$ deveria ser desenhado num espaço tridimensional (bidimensional se o vínculo da normalização for incluído). Mas desenhamos o valor do máximo de $H[P]$ que é função de um único parâmetro.

3.3.4 Variáveis reais: a gaussiana

Queremos encontrar $P(x)$ que maximize $H[P]$ sujeito à informação I

$$\begin{aligned} \int P(x) dx &= 1 \\ \int xP(x) dx &= \mu \\ \int x^2 P(x) dx &= \sigma^2 + \mu^2 \end{aligned} \quad (3.35)$$

Isto é, qual é a densidade de probabilidade que devemos atribuir a uma variável quando sabemos que a sua média e variância tem valor finito dado? Novamente

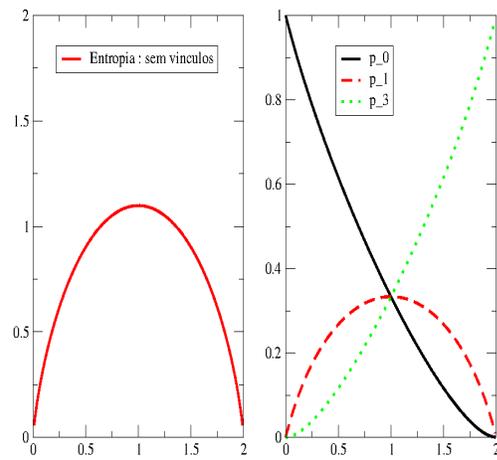


Figura 3.2: Esquerda: Máximo da entropia como função da média m . Direita: as probabilidades p_0 , p_1 e p_2 que maximizam a entropia de Shannon como função da informação $\langle x \rangle = m$.

introduzimos os vínculos através dos multiplicadores de Lagrange. Procuramos o máximo de “ $H[p]$ + vínculos”, através do seguinte problema variacional:

$$\delta \left[- \int dx P(x) \log \left(\frac{P(x)}{p_0(x)} \right) + \lambda_0 \left\{ \int P(x) dx - 1 \right\} + \right. \quad (3.36)$$

$$\left. \lambda_1 \left\{ \int x P(x) dx - \mu \right\} + \lambda_2 \left\{ \int x^2 P(x) dx - \sigma^2 - \mu^2 \right\} \right] = 0 \quad (3.37)$$

O diferença importante com relação ao caso de variáveis discretas é que aparece a distribuição p_0 . Supomos que neste caso o problema físico é tal que quando descrito usando a variável X , antes de levar em consideração a informação I ⁷ não temos nenhuma preferência, ou seja tomamos p_0 como constante, que pode ser esquecida. Novamente fazemos mudanças $P \rightarrow P + \delta P$, tratando as variações $\delta P(x)$ para valores de x diferentes, como independentes ,

$$\{1 + \log P(x) + \lambda_0 + \lambda_1 x + \lambda_2 x^2\} \delta P(x) = 0, \quad (3.38)$$

lembrando que as variações são independentes, o termo entre chaves deve ser igual a zero, o que leva novamente a uma forma exponencial

$$P(x) = e^{1 - \lambda_0 - \lambda_1 x - \lambda_2 x^2} \quad (3.39)$$

Agora escolhemos os multiplicadores de Lagrange para que os vínculos sejam satisfeitos. Use $\int_{-\infty}^{\infty} \exp(-y^2/2) dy / \sqrt{2\pi} = 1$, as propriedades das integrais ante mudanças de variáveis e mostre que

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3.40)$$

O resultado é distribuição gaussiana. Vemos um dos motivos da utilidade de distribuições gaussianas em tantos lugares em análise de dados: se só soubermos os valores finitos da média e da variância, então o uso de uma gaussiana é forçada por máxima entropia, qualquer outra distribuição significaria o uso implícito de informação não conhecida.

Exemplo: Bayes e Máxima Entropia

Foram feitas N medidas da variável Y para diferentes valores da variável de controle X , o que constitui o conjunto de dados $D = \{(x_i, y_i)\}_{i=1, \dots, N}$. Supomos conhecida a lei de Y em função de X , mas desconhecemos o parâmetro θ :

$$y = f_\theta(x)$$

e sabemos que há erros na medida dos valores de Y , que supomos independentes uns dos outros. Supomos ainda que as medidas de X não têm erros:

$$y_i = f_\theta(x_i) + \eta_i \quad (3.41)$$

⁷Ou seja voltamos à entropia de Shannon. O que significa uma distribuição uniforme no eixo real inteiro? Veja Jeffreys para priors não normalizáveis.

Sabemos, pelo estudo do aparelho de medida, que a variância de η é finita e que não há desvios sistemáticos: $\langle \eta \rangle = 0$ e $\langle \eta^2 \rangle = \sigma^2 < \infty$. Supomos que temos um conhecimento prévio de θ codificado na distribuição a priori: $p_0(\theta)$, logo, pelo teorema de Bayes determinamos

$$P(\theta|(x_i, y_i), I) = \frac{p_0(\theta)P(\{y_i\}|\theta, \{x_i\}, I)}{P(D)}, \quad (3.42)$$

Qual é a expressão para a verossimilhança $P(\{y_i\}|\theta, \{x_i\}, I)$?

Note por independência que deve ser o produto $\prod_i P(y_i|\theta, x_i, I)$, logo só devemos nós preocupar com um termo.

Se não houvesse ruído η :

$$P(y_i|\theta, x_i, I) = \delta(y_i - f_\theta(x_i))$$

Devido à equação 3.41 podemos ver que

$$P(y_i|\theta, x_i, I) = P(\eta)$$

Dada a informação que temos, qual é $P(\eta)$?

Este é um caso para o método de máxima entropia. A distribuição que surge é a gaussiana e se o tivéssemos dito desde o começo, poucos iriam reclamar.

Conclua o problema supondo que na região de relevância de θ a distribuição a prior $P(\theta)$ é uniforme. Obtenha então a posterior. Se olharmos para o valor θ_M que maximiza a posterior, neste caso a verossimilhança, teremos uma estimativa de *Máxima Verossimilhança*. Neste caso é o resultado de Mínimos Quadrados, que vimos no capítulo 1.

3.4 Entropia e convexidade

A divergência de Kullback-Leibler ou entropia cruzada entre duas distribuições de probabilidade que são não nulas no conjunto de n estados $\{i\}$ é definida por

$$K[p|q] = \sum_i p_i \log \frac{p_i}{q_i}. \quad (3.43)$$

Se considerarmos $q_i = 1/n$ a distribuição uniforme, termos

$$K[p|q] = \sum_i p_i \log(np_i), \quad (3.44)$$

$$K[p|q] = \log n - H[p_i], \quad (3.45)$$

Um número notável de resultados pode ser obtido a partir de considerações de convexidade do logaritmo, que leva por exemplo a

$$\log x \leq x - 1, \quad (3.46)$$

e que pode ser facilmente visto do gráfico das duas funções. Mas ver o gráfico não é suficiente, prove este resultado. A igualdade só ocorre se $x = 1$. A equação 3.46 usada na forma de $H[p]$ leva a

$$\begin{aligned} H[p] &= -\sum_i p_i \log(p_i), \\ &\geq -\sum_i p_i(p_i - 1) \\ &\geq 1 - \sum_i p_i^2 \geq 0, \end{aligned} \quad (3.47)$$

que a entropia é não negativa. Só será nula se $\sum_i p_i^2 = 1$, o que só pode ocorrer se $p_i = 0$ para todo i exceto para um estado, e.g. i' , para o qual $p_{i'} = 1$, ou seja temos informação completa que o valor de X é $x_{i'}$.

Se usarmos a mesma cota na expressão de $K[p|q]$ obteremos

$$\begin{aligned} K[p|q] &= -\sum_i p_i \log \frac{1}{np_i}, \\ &\geq -\sum_i p_i \left(\frac{1}{np_i} - 1 \right) \\ &\geq \sum_i \frac{1}{n} - \sum_i p_i = 0, \end{aligned} \quad (3.48)$$

com a igualdade ocorrendo somente se $p_i = q_i$. Juntando as duas cotas com a relação entre K e H , equação 3.45, temos que

$$0 \leq H[p] \leq \log n, \quad (3.49)$$

com a igualdade à esquerda ocorrendo quando temos informação completa sobre X e à direita quando a informação é simétrica com respeito aos n estados.

3.4.1 Independência

Anteriormente vimos que a regra para análise do produto lógico ou conjunção em função das probabilidades de asserções mais simples é

$$P(X, Y|I) = P(X|YI)P(Y|I). \quad (3.50)$$

Suponha o caso em que a informação sobre Y não diz nada sobre X , isto é, sob a informação I , saber algo sobre Y não muda nada sobre as probabilidades que atribuímos a X , isto é X é independente de Y , ou seja $P(X|YI) = P(X|I)$. É trivial mostrar que então $P(Y|XI) = P(Y|I)$ e que $P(X, Y|I) = P(X|I)P(Y|I)$ a probabilidade conjunta se fatoriza. É interessante considerar $K[P(X, Y|I)|P(X|I)P(Y|I)]$ que está relacionada à diferença de informação que temos quando temos a probabilidade conjunta e as marginais, $P(X|I) =$

$\sum_Y P(X, Y|I)$ e $P(Y|I) = \sum_X P(X, Y|I)$.

$$\begin{aligned} K[P(X, Y|I)|P(X|I)P(Y|I)] &= \sum_{X,Y} P(X, Y|I) \log \frac{P(X, Y|I)}{P(X|I)P(Y|I)} \\ &= -H[P(X, Y|I)] + \sum_{X,Y} P(X, Y|I) \log P(X|I) + \sum_{X,Y} P(X, Y|I) \log P(Y|I) \\ &= -H[P(X, Y|I)] + H[P(X|I)] + H[P(Y|I)] \end{aligned} \quad (3.51)$$

e como $K \geq 0$ obtemos que

$$H[P(X, Y|I)] \leq H[P(X|I)] + H[P(Y|I)]. \quad (3.52)$$

A equação 3.51 mostra que $K[P(X, Y|I)|P(X|I)P(Y|I)]$ mede a diferença entre a informação que falta se só olharmos para X e Y separadamente ou conjuntamente. A equação 3.52, que mostra a propriedade chamada de subaditividade, mostra que a informação que falta quando as variáveis são analisadas conjuntamente é menor ou igual que quando o são em separado. A não ser que as variáveis sejam independentes a análise da distribuição conjunta incorpora mais informação ou seja, falta menos informação.

3.4.2 Exercício sobre independência: Variáveis Gaussianas correlacionadas

Considere a densidade conjunta de probabilidades de X e Y que tomam valores no eixo real x e y :

$$P(x, y|\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)}. \quad (3.53)$$

O parâmetro ρ está entre 0 e 1 e no que resta do exercício todas as probabilidades serão condicionadas a conhecimento de ρ que não será escrito explicitamente.

Calcule as distribuições marginais $P(x)$ e $P(y)$, e.g. integrando $P(x) = \int dy P(x, y)$. Mostre que são gaussianas com média nula e variância 1.

Encontre a distribuição condicional $P(x|y)$ (use o teorema de Bayes).

Calcule a entropia das marginais (gaussiana univariadas (sic)) Calcule a entropia da distribuição conjunta (gaussiana multivariada)

Discuta o significado de ρ como (A) correlação entre X e Y , (comece calculando $\langle xy \rangle$) e (B) como medida da diferença de informação que temos nos seguintes dois casos: (a) dado $P(x, y)$ e (b) dados $P(x)$ e $P(y)$.

Use a sua linguagem preferida e trace as curvas de nível da distribuição conjunta para diferentes valores de ρ para mais uma forma de entender o papel de ρ .

Aqui acaba este exercício. O que segue é mais difícil, pode ser pulado, ou deixado até que seja discutido quando estudarmos o problema de aprendizagem de máquinas usando idéias de mecânica estatística. Suponha que é dado o valor do sinal de y , i.e $\sigma_y = \text{sin}(y)$. Calcule $P(\sigma_x|\sigma_y)$.

3.5. CASO GERAL PARA INFORMAÇÃO SOBRE VALORES ESPERADOS 67

Suponha que queremos usar σ_x como estimador de σ_y . Calcule como função de ρ a probabilidade e_σ de fazer um *erro*, i.e. de errar o sinal de Y quando o sinal de X é usado no seu lugar ou que os sinais de X e Y sejam diferentes. A resposta $e_\sigma = \frac{1}{\pi} \arccos \rho$. Fica fácil de entender se fizermos um desenho. Trace duas linhas ξ_x e ξ_y que fazem um ângulo θ . Trace dois vetores perpendiculares ‘as linhas \vec{J}_x e \vec{J}_y . Também fazem um ângulo θ . Escolha uma direção qualquer, desenhe um vetor \vec{S} do ponto de cruzamento nessa direção. A pergunta é qual é a probabilidade que a projeção desse vetor nas duas perpendiculares tenha o mesmo sinal. A resposta é $2\theta/2\pi$. Generalize o desenho par N dimensões. A resposta (deve ser provado) continua sendo θ/π . Note que θ está relacionado ao produto escalar entre os dois vetores perpendiculares aos (hiper-)planos $N - 1$ dimensionais. Portanto $\theta = \arccos \vec{J}_x \cdot \vec{J}_y$. É uma consequência do teorema do limite central que as projeções $x = \vec{J}_x \cdot \vec{S}$ e $y = \vec{J}_y \cdot \vec{S}$ são variáveis cuja distribuição conjunta é dada por 3.53.

O discutido neste exemplo é extremamente útil em problemas de aprendizagem de máquina, onde queremos aprender uma direção desconhecida \vec{J}_y só com informação sobre σ_y . O que se faz é construir uma aproximação \vec{J}_x a partir de σ_y para vários exemplos \vec{S} . Isto é, a máquina aprende e a medida da qualidade da aprendizagem, $P(\sigma_x|\sigma_y)$ é chamado de erro de generalização, que cai quanto maior for a correlação ρ . Sobre isto, muito mais será dito em capítulos posteriores ⁸.

3.5 Caso Geral para informação sobre valores esperados

Em geral podemos incluir vários vínculos

$$\langle g_k(x) \rangle = G_k, \quad (3.54)$$

com $k = 1 \dots K$ e chamando $1 = g_0$, o problema variacional será

$$\delta \left[- \sum P_i \log P_i - \sum_{k=0}^K \lambda_k \left\{ \sum_i g_k(x_i) P_i - \mu \right\} \right] = 0 \quad (3.55)$$

que leva a

$$- \sum_i \delta P_i \left\{ \log P_i + 1 + \sum_{k=0}^K \lambda_k g_k(x_i) \right\} = 0 \quad (3.56)$$

Como anteriormente, as variações são independentes e cada termo entre chaves é nulo, o que leva a

$$P_i = \frac{e^{-\sum_{k=0}^K \lambda_k g_k(x_i)}}{Z} \quad (3.57)$$

onde a função de partição $Z = \exp(-1 - \lambda_0)$ toma a forma

⁸Mas não neste curso introdutório.

$$Z = \sum_i e^{-\sum_{k=1}^K \lambda_k g_k(x_i)} \quad (3.58)$$

para satisfazer a normalização. Os multiplicadores de Lagrange remanescentes poderão ser calculados usando

$$\langle g_k \rangle = G_k = - \left(\frac{\partial \log Z}{\partial \lambda_k} \right)_{\lambda_i, i \neq k} \quad (3.59)$$

Exercício Mostre o resultado acima.

As equações acima formam um conjunto de equações não lineares que determinam os valores de $\{\lambda_k\}$ se os G_k forem conhecidos. No entanto nos dá ainda mais. Suponha que os valores de G_k não sejam conhecidos, mas que só tenhamos o conhecimento que se os soubessemos teríamos determinado uma distribuição adequada para algum fim. Isso significa que se, em vez de G_k tivéssemos acesso aos valores dos $\{\lambda_k\}$, então poderíamos calcular os G_k , e outros valores esperados de interesse teórico e experimental. Isto será possível quando tivermos uma interpretação quanto ao significado do λ s. Isto será possível quando for feita a conexão com a termodinâmica. O valor máximo da entropia de Shannon será identificada com a entropia termodinâmica de Clausius e o multiplicador de Lagrange associado à energia do sistema terá as propriedades do inverso da temperatura. O fato que não sabemos o valor esperado da energia em nada nos atrapalha, mas indica que a temperatura é uma quantidade útil e seu conhecimento, adquirido através de mais uma medida serve para determinar, junto com outros parâmetros (e.g. volume, número de partículas) o estado de equilíbrio que pode de forma robusta ser estudado em diferentes laboratórios. Para poder fazer esta ligação precisamos ver mais algumas propriedades da distribuição de máxima entropia.

3.6 Conexão com a Termodinâmica: estrutura matemática

3.6.1 Transformações de Legendre

Denote o máximo da entropia de Shannon por S . Este valor foi encontrado a partir da informação dada, portanto é uma função dela, ou seja do conjunto $\{G_k\}$. Substituindo o valor da probabilidade que maximiza H mais vínculos, $p_i = \frac{\exp(-\sum_{k=1}^K \lambda_k g_k)}{Z}$ na forma $H = -\sum p_i \log p_i$ obtemos

$$S(G_1, G_2, \dots, G_K) = \log Z(\lambda_1, \lambda_2, \dots, \lambda_K) + \sum_{k=1}^K \lambda_k G_k \quad (3.60)$$

Este resultado é importante porque mostra a relação análoga à relação termodinâmica $S = -\frac{F}{T} + \frac{E}{T}$, ou mais comumente escrita com $F = E - TS$. Na forma

3.6. CONEXÃO COM A TERMODINÂMICA: ESTRUTURA MATEMÁTICA 69

geral, mostra como podemos passar de uma descrição em termos do conjunto de variáveis $\{G_k\}$ para o conjunto $\{\lambda_k\}$.

Considerando uma pequena variação dos valores G_k , a partir do resultado acima podemos mostrar que:

$$\frac{\partial S(G_1, G_2, \dots, G_K)}{\partial G_i} = \sum_{j=1}^K \frac{\partial \log Z}{\partial \lambda_j} \frac{\partial \lambda_j}{\partial G_i} + \sum_{j=1}^K \frac{\partial \lambda_k}{\partial G_i} G_i + \lambda_i \quad (3.61)$$

Usando a equação 3.59, obtemos finalmente que

$$\lambda_i = \frac{\partial S(G_1, G_2, \dots, G_K)}{\partial G_i} \quad (3.62)$$

Um caso particular desta relação é $\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_{V,N}$ que veremos na seção 3.9.1.

Os resultados 3.60 e 3.62 mostram como podemos descrever o sistema em termos dos G_k s ou dos λ_k s, ou seja $S(\{G_k\})$ e $\log Z(\{\lambda_k\})$ são obtidos um do outro através de uma transformada de Legendre.

exercício Suponha $K = 1$ e $\langle g_1 \rangle = E$, a energia de um sistema com graus de liberdade X e estados $\{x_i\}$ e Hamiltoniano g_1 . (A) Mostre que $F = E - TS$, onde $F = -T \log Z$ e $T = 1/\lambda_1$. (B) Mostre que $\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_{V,N}$.

Este exercício mostra a utilidade de tudo isto para a termodinâmica. Mas ainda há muita física para explicar. Neste ponto deve ser entendido como um exercício formal.

3.6.2 Relações de Maxwell

Podemos continuar estudando as derivadas parciais superiores e obter a estrutura das relações de Maxwell. Olhemos para as segundas derivadas, a partir das derivadas da equação 3.62:

$$\frac{\partial \lambda_i}{\partial G_j} = \frac{\partial^2 S(G_1, G_2, \dots, G_K)}{\partial G_i \partial G_j} = \frac{\partial^2 S(G_1, G_2, \dots, G_K)}{\partial G_j \partial G_i} = \frac{\partial \lambda_j}{\partial G_i} \quad (3.63)$$

e da equação 3.59

$$\frac{\partial G_k}{\partial \lambda_i} = -\frac{\partial^2 \log Z}{\partial \lambda_k \partial \lambda_i} = \frac{\partial G_i}{\partial \lambda_k} \quad (3.64)$$

Aqui é interessante fazer uma pausa e lembrar a enorme variedade de resultados de interesse experimental que seguem das relações de Maxwell (e.g. Callen). Do ponto de vista matemático não passam da hipótese que as derivadas mistas são contínuas, essencialmente uma trivialidade. Do ponto de vista físico, justificam a teoria e a escolha dos graus de liberdade $\{x_i\}$ e a energia para descrever o equilíbrio termodinâmico do sistema.

Suponha que para um dado sistema as previsões falhem no campo experimental. O que fazer? Podemos jogar fora o método de máxima entropia ou podemos pensar que talvez nos tenhamos enganado ao escolher os graus de liberdade para descrever o problema, escolher de outra forma e começar a calcular

novamente para comparar com os resultados experimentais. Historicamente há precedentes importantes que nos levam a pensar na segunda possibilidade com grande atenção. Afinal foi a falta de concordância entre a termodinâmica obtida da Mecânica Estatística de Gibbs e Boltzmann, com seus métodos similares ao exposto acima, calculada a partir da Mecânica Clássica e o Eletromagnetismo Clássico⁹ e a experiência (e.g. corpo negro, calor específico de sólidos a baixas temperaturas), que levou às primeiras idéias da Mecânica Quântica. O que foi abandonado não foi a forma consistente de calcular, introduzida pela Mecânica Estatística, mas a forma errada de descrever os graus de liberdade da física clássica. A lição disso é que se uma aplicação dos métodos entrópicos levarem a problemas, não devemos jogar fora o método, mas desconfiar da maneira como foi aplicado.

Suponha que jogamos com um dado comum que Gabriel ganhou de brinde como lembrancinha de uma festa de aniversário. Suponha que eu acredite que o dado tem sete faces e que a informação que tenho sobre elas seja simétrica. Espero que a probabilidade de cada uma seja $1/7$. Você acredita que eu terei sucesso no jogo? Por exemplo, apostamos no seguinte jogo. Se sair, na próxima jogada do dado o número sete eu ganho, se sair o seis, perco. Se o resultado for outro número, o dado é jogado novamente. Não há pior erro que a enumeração errada dos estados acessíveis ao sistema. Dados de um jogo realizado mostram que eu perdi. Talvez seja hora de mudar minha maneira de descrever os estados possíveis do dado. O Eletromagnetismo Clássico fez uma aposta semelhante e perdeu o jogo ao descrever o espectro de cavidade do corpo negro. As mudanças necessárias, feitas inicialmente por Planck começaram a mostrar a maneira correta, que finalmente levou à Mecânica Quântica.

3.7 Conexão com a Termodinâmica: Micro e Macroestados

O uso de uma palavra num novo contexto requer que o proponente justifique seu novo uso. Aparentemente von Neumann teria dito a Shannon que poderia usar a palavra entropia para denotar a sua medida de informação por dois motivos. Primeiro porque a forma $-\sum p_i \log p_i$ aparece em trabalhos de Gibbs e também de Boltzmann e é efetivamente uma entropia. Segundo, porque já que ninguém sabe exatamente o que é entropia, não seria possível rebater o seu uso do nome.

O objetivo principal da Mecânica Estatística é obter as propriedades termodinâmicas de sistemas físicos. Aparentemente a primeira e mais importante informação que devemos ter sobre o sistema se refere às propriedades microscópicas e devemos então indagar sobre quais são os constituintes microscópicos e como interagem, isto é, caracterizar os graus de liberdade. Na linguagem da Mecânica Estatística, isto é definir quais são os *microestados* do sistema. Por exemplo, “este sistema esta feito de tal tipo de molécula”, “estas interagem entre si de tal forma”. Por que parar em moléculas? Elas são

⁹Na época não era necessário o adjetivo clássico.

3.7. CONEXÃO COM A TERMODINÂMICA: MICRO E MACROESTADOS 71

feitas de átomos. Devemos descrever as posições dos átomos. Deve ser necessário entender e caracterizar os estados eletrônicos. E os núcleos? Claro, são feitos de nêutrons, prótons. O estado do núcleo é certamente informação microscópica e deve ser incorporada. Os quarks que formam os núcleons, são eles microscópicos? Certamente e então devemos nos perguntar se a estrutura do quark também deverá ser levada em conta. Ai há dúvidas. Em alguns modelos são fundamentais, em outros são compostos. Como decidir? A coisa está ficando exagerada. Esta discussão levaria a pensar que iremos atribuir probabilidades a microestados especificados pelos valores de

$$\{qp\}_{mol}\{qp\}_{atom}\{qp\}_{elet}\{qp\}_{nucl}\{qp\}_{prot}\{qp\}_{neutr}\{qp\}_{quark}\{qp\}_{gluon}\{qp\}? \quad (3.65)$$

As escalas de energias em que os efeitos de cada tipo de estrutura se faz sentir são muito diferentes. Se houver necessidade de uma especificação tão detalhada dos graus de liberdade, não seria possível obter a termodinâmica. Do ponto de vista histórico vemos que não seria possível obter nenhum resultado no século 19, pois então o conhecimento da estrutura da matéria estava longe do que é hoje. Também não devemos esperar nenhum resultado hoje, pois certamente o conhecimento que temos é incompleto e deverá mudar. A especificação dos graus de liberdade relevantes é o principal problema em Mecânica Estatística, pois ela determinará as perguntas que queremos responder e “nada restringe tanto uma resposta como a própria pergunta”¹⁰.

Devemos parar e começar de novo. Qual é a pergunta que queremos responder? A resposta inclui questões como determinar quais são as propriedades termodinâmicas numa certa escala de temperatura. Imaginemos duas caixas de paredes rígidas, impermeáveis e isolantes que não permitem trocas de energia, também chamadas adiabáticas, com volumes, número de moles e energia iguais em dois laboratórios diferentes. Queremos caracterizar o estado dos sistemas de forma que *certas* perguntas tenham a mesma resposta. A velocidade da partícula 17 será a mesma nas duas caixas? Certamente essa não é uma das perguntas que queremos que tenham a mesma resposta. Queremos saber por exemplo, o calor específico ou quanta energia é necessária para elevar a temperatura de 14,5C até 15,5C de um grama de fluido. Não parece razoável que seja necessário, nesse regime de temperaturas, que os detalhes dos estados dos quarks tenham alguma influência. Devemos incluí-los, mesmo que não tenham importância? Se a resposta fosse sim como poderíamos justificar a inclusão da descrição dos graus de liberdade que ainda não sabemos como descrever, ou será que já sabemos tudo sobre a estrutura microscópica da matéria? ou como foi que no século 19, sem saber nada sobre a estrutura atômica foi possível a obtenção dos primeiros resultados?

Suponha que jogamos N moedas e queremos saber por exemplo o número de caras para cima. A resposta que você dará é baseada na distribuição binomial $P(m|N) = N!/(m!(N-m)!p^m q^{N-m})$. Suponha que agora é dada mais informação. Cada moeda tem um rosto, e para cada moeda é dado em que direção aponta o nariz. Temos dois graus de liberdade para cada moeda, um o

¹⁰A.C.

valor de $\sigma = \pm 1$, que indica se é cara ou coroa. Outro, a direção do nariz, está entre θ e $\theta + d\theta$.

A probabilidade da ocorrência de um dado resultado, para moedas independentes é

$$P(\sigma_1 = +1, \dots, \sigma_N = -1, \theta_1, \dots, \theta_N) \prod_i d\theta_i = p \dots q P(\theta_1) \dots P(\theta_N) \prod_i d\theta_i \quad (3.66)$$

Queremos a probabilidade que o número de caras seja m . Seja n o de coroas. Então $m + n = N$ e $m - n = \sum_i \sigma_i$ o que leva a $m = (N + \sum_i \sigma_i)/2$. De acordo com o capítulo 2 esperamos que

$$\begin{aligned} P(m|N\{p, q, P(\theta)\}) &= \sum_{\sigma_1, \dots, \sigma_N} \int \prod_i d\theta_i P(\sigma_1, \dots, \sigma_N, \theta_1, \dots, \theta_N) \delta_{m, (N + \sum_i \sigma_i)/2} \\ &= \frac{N!}{m!(N-m)!} p^m q^{N-m} \end{aligned} \quad (3.67)$$

As moedas são claramente distinguíveis entre si. Mas $P(m|N\{p, q, P(\theta)\})$ tem fatoriais para eliminar a contagem de combinações mais de uma vez. Isto não indica que as moedas são idênticas? A pergunta que queremos responder, sobre estados de $\sum_{i=1, N} \sigma_i$ é independente do valor de θ e a distinguibilidade (?) das moedas é irrelevante.¹¹ Qual é a entropia que devemos atribuir a esse sistema? A resposta, $-\sum p_i \log p_i$ depende explicitamente do que chamamos de estado i . Se as perguntas que quisermos responder não tratam de θ , então duas moedas com mesmo valor de σ são indistinguíveis.

A lição do parágrafo anterior está em que a resposta depende da pergunta. O tipo de pergunta feita determina quais são os graus de liberdade que devem ser usados para descrever o sistema. O leitor terá percebido que se perguntarmos sobre m , nem o valor de θ nem o estado eletrônico do material que compõe a moeda, que tínhamos esquecido de considerar, interessa. É comum neste ponto introduzir a noção de microestado, embora fique claro que sua definição é um tanto quanto vaga. Um microestado das moedas poderia ser descrito pelos valores de $\{(\theta_i, \sigma_i)\}_{i=1, N}$ se as perguntas que queremos responder tem algo a ver com a orientação dos narizes. Pode ser que haja um campo magnético e os narizes tendam a apontar em alguma direção privilegiada. Mas o microestado que nós interessa quando queremos responder sobre o número total de caras é determinado simplesmente por $\{\sigma_i\}_{i=1, N}$. A probabilidade do microestado de interesse $P(\{\sigma_i\}_{i=1, N})$ é obtida por marginalização dos θ_i

$$P(\{\sigma_i\}_{i=1, N}|I) = \int \prod_i d\theta_i P(\{(\theta_i, \sigma_i)\}_{i=1, N}|I) \quad (3.68)$$

e também por marginalização de outras variáveis irrelevantes para a pergunta presente. Este um tipo de agrupamento: todos os valores de θ são agrupados e

¹¹As moedas são distintas mas são tratadas de forma que mostra que não faz diferença se a moeda que tem $\sigma = 1$ tem um nariz que aponta numa ou outra direção Voltaremos a falar destas moedas ao discutir gases quânticos

a probabilidade total do estado é obtido pela soma dos elementos do conjunto agrupado.

Toda vez que falamos num microestado tomamos como óbvio que as variáveis em escalas menores foram agrupadas. É difícil exagerar a importância desta idéia, se não fosse assim não haveria como falar de nenhum sistema, pois não seríamos capazes de acabar com a sequência de escalas de forma satisfatória.

O estado descrito por $P(m|N)$ chamamos de macroestado, mas este também depende da pergunta que fazemos. Pode ser que hajam agrupamentos intermediários que podem ser chamados de mesoestados.

O que vai definir o que chamamos de macroestado é a pergunta que queremos responder. Perguntas em física são experiências. O macroestado é escolhido pelo arranjo experimental que inclui fixar alguns parâmetros e a medida de outros. No mesmo sistema podemos mudar o que é fixo e o que se mede. A informação que usaremos para determinar as probabilidades é o próprio valor das quantidades que são fixas pelo experimental durante a experiência.

3.8 Conexão com a Termodinâmica: Física

3.8.1 O Ensemble Microcanônico

A importância do conceito de energia é devida à sua conservação. O célebre teorema de Noether mostra a relação entre simetrias e leis de conservação. Se sob certas condições o sistema tem uma simetria contínua, concluímos que numa experiência realizada nessas condições haverá uma quantidade conservada. Suponha um sistema isolado de tal maneira que as propriedades são invariantes antes translações temporais, a energia será conservada. A preparação do arranjo experimental é feita com o objetivo de responder certas perguntas. Isso define o macroestado. Pode ser que não saibamos como fixá-lo, mas essencialmente esse é o objetivo de nosso estudo, encontrar quais as condições experimentais, ou quais os vínculos impostos ao sistema, para que o macroestado seja um macroestado específico. O problema consiste agora em descrever de forma correta os microestados. Mas o que é um microestado? Suponhamos que o experimento será feito numa certa região de energias e que nessa faixa os graus de liberdade relevantes sejam descritos por $\{q_i, p_i\}$ e que a dinâmica, seja em termos dessas variáveis relevantes, descrita por um Hamiltoniano $\mathcal{H}(q_i, p_i)$. Suponhamos então para ser específicos que colocamos o sistema em condições que sua energia é E fixa. Fixar E num valor determinado é muito difícil experimentalmente. Talvez seja melhor dizer que sua energia é fixa entre E e $E + \delta E$. Para alguns sistemas, fluídos simples, é também necessário fixar o número de partículas N e o volume do sistema V . Devemos caracterizar os microestados, por exemplo determinar se número. Mas o que é um microestado? Suponhamos que o experimento será feito numa certa região de energias e que nessa faixa os graus de liberdade relevantes sejam descritos por $\{q_i, p_i\}$, o espaço de fases, e que a dinâmica, seja em termos dessas variáveis relevantes, descritos por um Hamiltoniano $\mathcal{H}(q_i, p_i)$. Podemos associar à esses microestados densidades de probabilidades $P(q_i, p_i)$

usando a informação disponível e o método de máxima entropia

$$H[P] = -k \int_{E \leq \mathcal{H}(q_i, p_i) \leq E + \delta E, V} \prod_i dq_i dp_i P(q_i, p_i) \log P(q_i, p_i) \quad (3.69)$$

sujeito unicamente ao vínculo de normalização. Tomaremos $k = 1$. O resultado de maximizar a entropia de Shannon quando a informação sobre os microestados é indiferente é simplesmente que a distribuição é uniforme sobre os microestados permitidos:

$$P(q_i, p_i) = c \quad (3.70)$$

como esperado pelo princípio da razão insuficiente.

O fluido simples, introduzido por Gibbs, é um sistema onde o estado termodinâmico é determinado por algumas poucas variáveis. Suponhamos um gás isolado numa caixa de paredes adiabáticas. O experimental pode controlar separadamente a energia E , o volume V e o número de constituintes (e.g moléculas) N . A constante c na equação 3.70 é determinada pela imposição do vínculo de normalização,

$$P(q_i, p_i) = \frac{1}{\Omega(E, N, V)} \quad (3.71)$$

onde $\Omega(E, N, V)$ é o volume no espaço de fases e para variáveis contínuas é

$$\Omega(E, N, V) = \int_{E \leq \mathcal{H}(q_i, p_i) \leq E + \delta E} \prod_i dq_i dp_i. \quad (3.72)$$

Uma vez determinada a probabilidade de cada microestado substituímos de volta na expressão da entropia de Shannon para obter o valor máximo da entropia:

$$H_{extr}[P] = \int_{E \leq \mathcal{H}(q_i, p_i) \leq E + \delta E} \prod_i dq_i dp_i \frac{1}{\Omega(E, N, V)} \log \Omega(E, N, V) \quad (3.73)$$

e o extremo da entropia é

$$S(E, N, V) = \log \Omega(E, N, V). \quad (3.74)$$

Esta expressão foi escrita primeiramente por Planck, que a chamou de entropia de Boltzmann, por estar implícita nos seus escritos. Expressão equivalente está gravada no seu túmulo ($S = k \log W$).

Não é incomum que os livros de Mecânica Estatística comecem por *postular* que todos os microestados com energia E são igualmente prováveis. A conexão entre a Mecânica e a Termodinâmica é também *postulada* através da equação 3.74.

Ainda não está claro porque S , o máximo da entropia de Shannon, está relacionada (é) a entropia de Clausius, feito isso iremos descobrir as relações entre quantidades mecânicas e termodinâmicas. E então teremos tipicamente, para avançar em Física Estatística, que olhar o problema da determinação de

$\Omega(E, N, V)$ ou a quantidade equivalente em cada sistema e situação experimental. Também veremos que certas situações experimentais são melhor descritas por vínculos que vem na forma de estipulação de valores esperados de certas quantidades. A percepção que há interesse teórico e experimental em descrever vínculos experimentais diferente de simplesmente impor que a energia ou o volume ou o número de partículas é fixo, é devida a Gibbs e à sua teoria dos *ensembles*. Diferentes tipos de vínculos experimentais levam a resultados diferentes, cada situação, chamada de *ensemble* recebe um nome específico. O que acabamos de ver, onde a energia do sistema sistema é fixa, porque o sistema está isolado, é chamado de *ensemble* microcanônico. Outros *ensembles* são obtidos por extensões da análise deste caso.

3.9 Conexão com a Termodinâmica: T, P, μ

3.9.1 Temperatura

Um sistema isolado, numa caixa de volume V é dividido em dois subsistemas, o sistema 1 de graus de liberdade $\{q_i, p_i\}$, N_1 partículas e volume V_1 e o sistema 2 de graus de liberdade $\{Q_i, P_i\}$, N_2 partículas e volume V_2 . Isto pode ser feito pela inclusão de uma parede ideal que não permite a passagem de átomos de um lado para outro (impermeável) e rígida, não há variações de volume. A separação é feita mantendo

$$E_1 + E_2 = E \quad (3.75)$$

$$V_1 + V_2 = V, \quad (3.76)$$

$$N_1 + N_2 = N, \quad (3.77)$$

tal que todos os termos das equações 3.76 e 3.77 conhecidos enquanto só o lado direito da equação 3.75 é conhecida. Como é feita a distribuição de energia entre os dois subsistemas? Suponhamos que o Hamiltoniano é bem descrito por

$$\mathcal{H}(\{q_i, p_i, Q_i, P_i\}) = \mathcal{H}_1(\{q_i, p_i\}) + \mathcal{H}_2(\{Q_i, P_i\}) \quad (3.78)$$

onde desprezamos um termo de interação que é muito pequeno e talvez so dependa da superfície da região de contato entre os dois sistemas. Supomos as interações de curto alcance ou se de longo alcance blindadas por cargas opostas. A energia do sistema se dividirá de forma ainda desconhecida entre os dois subsistemas, mas uma vez dividida entre eles, cada subsistema terá uma distribuição uniforme, da mesma forma que no caso da seção anterior. Uma vez estabelecido o equilíbrio cada sistema terá uma energia, um volume e um número de partículas fixos, análogo ao caso já visto. Desprezadas as interações inter sistema, esperamos que as variáveis $\{q_i, p_i\}$ e $\{Q_i, P_i\}$ sejam independentes e a distribuição de probabilidades seja um produto: $P(\{q_i, p_i, Q_i, P_i\}) = P(\{q_i, p_i\})P(\{Q_i, P_i\})$,

assim

$$\begin{aligned}
H[P(\{q_i, p_i, Q_i, P_i\})] &= - \int_{E_1 \leq \mathcal{H}_1(q_i, p_i) \leq E_1 + \delta E_1, E_2 \leq \mathcal{H}_2(Q_i, P_i) \leq E_2 + \delta E_2} \\
&\times P(\{q_i, p_i, Q_i, P_i\}) \log P(\{q_i, p_i, Q_i, P_i\}) \prod_i dq_i dp_i dQ_i dP_i \\
&= - \int_{E_1 \leq \mathcal{H}_1(q_i, p_i) \leq E_1 + \delta E_1} \prod_i dq_i dp_i P(\{q_i, p_i\}) \log P(\{q_i, p_i\}) \\
&\quad - \int_{E_2 \leq \mathcal{H}_2(Q_i, P_i) \leq E_2 + \delta E_2} \prod_i dQ_i dP_i P(\{Q_i, P_i\}) \log P(\{Q_i, P_i\}) \\
H[P(\{q_i, p_i, Q_i, P_i\})] &= H[P(\{q_i, p_i\})] + H[P(\{Q_i, P_i\})] \tag{3.79}
\end{aligned}$$

Note que na mesma hipótese sobre o hamiltoniano do sistema e de divisão de energia, o volume no espaço de fase se fatoriza

$$\begin{aligned}
\Omega(E, N, V) &= \int_{E_1 \leq \mathcal{H}_1(\{q_i, p_i\}) \leq E_1 + \delta E_1, E_2 \leq \mathcal{H}_2(\{Q_i, P_i\}) \leq E_2 + \delta E_2} \prod_i dq_i dp_i dQ_i dP_i \\
&= \int_{E_1 \leq \mathcal{H}_1(\{q_i, p_i\}) \leq E_1 + \delta E_1} \prod_i dq_i dp_i \int_{E_2 \leq \mathcal{H}_2(\{Q_i, P_i\}) \leq E_2 + \delta E_2} \prod_i dQ_i dP_i \\
&= \Omega(E_1, N_1, V_1) \Omega(E_2, N_2, V_2), \tag{3.80}
\end{aligned}$$

e os máximos de entropia de cada um dos subsistemas, para uma dada divisão de energia $E = E_1 + E_2$ serão dados pela análise anterior

$$S(E_1, E_2, N_1, N_2, V_1, V_2) = S(E_1, N_1, V_1) + S(E_2, N_2, V_2) \tag{3.81}$$

onde $S(E_1, N_1, V_1,) = \log \Omega(E_1, N_1, V_1)$ e forma análoga para o sistema 2. Este resultado parece muito bom até o ponto que percebemos que não sabemos a divisão de energias, só o vínculo que $E = E_1 + E_2$.

O que podemos afirmar sobre a partição de energias? A única forma que temos de escolher as distribuições de probabilidade é através da determinação do máximo da entropia e impomos que a entropia do sistema conjunto seja máximo. Basta olhar para o valor da entropia conjunta para uma dada divisão e dentre essas escolher o valor de E_1 que leva ao máximo valor:

$$S(E_1, E_2, N_1, N_2, V_1, V_2) = \log \Omega(E_1, N_1, V_1) \Omega(E_2, N_2, V_2), \tag{3.82}$$

$$S(E_1, E_2, N_1, N_2, V_1, V_2) = S(E_1, N_1, V_1,) + S(E_2, N_2, V_2) \tag{3.83}$$

portanto a maximização da energia sujeita ao vínculo $E = E_1 + E_2$ dará

$$0 = \frac{\partial S(E_1, E_2, N_1, N_2, V_1, V_2)}{\partial E_1} = \frac{\partial S(E_1, N_1, V_1,)}{\partial E_1} + \frac{\partial S(E_2, N_2, V_2)}{\partial E_1} \tag{3.84}$$

$$= \frac{\partial S(E_1, N_1, V_1,)}{\partial E_1} - \frac{\partial S(E_2, N_2, V_2)}{\partial E_2} \tag{3.85}$$

$$\left(\frac{\partial S(E_1, N_1, V_1)}{\partial E_1} \right)_{V_1, N_1} = \left(\frac{\partial S(E_2, N_2, V_2)}{\partial E_2} \right)_{V_2, N_2}. \quad (3.86)$$

Na última equação escrevemos de forma explícita o que é mantido constante ao tomar as derivadas parciais. Reconhecemos a relação que define a temperatura em termodinâmica,

$$\left(\frac{\partial S(E, N, V)}{\partial E} \right)_{V, N} = \frac{1}{T} \quad (3.87)$$

e reescrevemos a equação 3.86 como $1/T_1 = 1/T_2$, os sistemas 1 e 2 estarão em equilíbrio quando suas temperaturas forem iguais. Isso também nos permitirá interpretar (ver seções 3.6.1 e 3.11) do ponto de vista experimental o parâmetro de Lagrange associado à um vínculo, no caso a temperatura com um vínculo sobre o energia. Os parâmetros intensivos da termodinâmica serão associados a multiplicadores de Lagrange que impoem vínculos sobre valores de grandezas extensivas.

3.9.2 Pressão

Agora e na próxima subseção olharemos para situações experimentais onde um vínculo é relaxado. O sistema conjunto será descrito por uma nova distribuição de probabilidades obtida por maximização da entropia. O vínculo que liberamos agora diz respeito à rigidez da parede que separa o volume V em V_1 e V_2 . Ainda temos

$$E_1 + E_2 = E \quad (3.88)$$

$$V_1 + V_2 = V, \quad (3.89)$$

$$N_1 + N_2 = N, \quad (3.90)$$

mas do lado esquerdo, só os termos da equação 3.90 são conhecidos. Novamente procuramos o máximo da equação

$$S(E_1, E_2, N_1, N_2, V_1, V_2) = S(E_1, N_1, V_1,) + S(E_2, N_2, V_2) \quad (3.91)$$

mas agora devemos satisfazer duas condições:

$$\left(\frac{\partial S(E_1, E_2, N_1, N_2, V_1, V_2)}{\partial E_1} \right)_{V_1, N_1, N_2} = 0; \quad \left(\frac{\partial S(E_1, E_2, N_1, N_2, V_1, V_2)}{\partial V_1} \right)_{E_1, N_1, N_2} = 0 \quad (3.92)$$

que levam a

$$\left(\frac{\partial S(E_1, N_1, V_1)}{\partial E_1} \right)_{V_1, N_1} = \left(\frac{\partial S(E_2, N_2, V_2)}{\partial E_2} \right)_{V_2, N_2}. \quad (3.93)$$

e

$$\left(\frac{\partial S(E_1, N_1, V_1)}{\partial V_1} \right)_{V_1, N_1} = \left(\frac{\partial S(E_2, N_2, V_2)}{\partial V_2} \right)_{V_2, N_2}. \quad (3.94)$$

que significam respectivamente que

$$\frac{1}{T_1} = \frac{1}{T_2} \text{ e } \frac{P_1}{T_1} = \frac{P_2}{T_2} \quad (3.95)$$

no equilíbrio, não só as temperaturas serão iguais, mas também as pressões, a parede mudará de posição até que as pressões de ambos os lados se igualem.

3.9.3 Potencial químico

Consideremos a situação em que a parede que separa os dois subsistemas é rígida mas não impermeável. Só os termos da equação 3.89 são determinados pelo arranjo experimental. Podemos supor que o sistema é composto por vários tipos de moléculas

$$N_1^a + N_2^a = N^a \quad (3.96)$$

$$N_1^b + N_2^b = N^b \quad (3.97)$$

$$N_1^c + N_2^c = N^c \quad (3.98)$$

mas a parede só permite a passagem de moléculas do tipo a , sendo os termos da equações 3.97 e 3.98 conhecidos e fixos. Obteremos, por maximização da entropia do sistema conjunto que

$$\left(\frac{\partial S(E_1, N_1^a, N_1^b, N_1^c, V_1)}{\partial N_1^a} \right)_{E_1, N_1^b, N_1^c, V_1} = \left(\frac{\partial S(E_2, N_2^a, N_2^b, N_2^c, V_2)}{\partial N_2^a} \right)_{E_2, N_2^b, N_2^c, V_2} \quad (3.99)$$

e lembrando da definição termodinâmica do potencial químico $\mu(E, V, N)$:

$$-\frac{\mu}{T} = \left(\frac{\partial S(E, N, V)}{\partial N} \right)_{E, V} \quad (3.100)$$

$$\frac{1}{T_1} = \frac{1}{T_2} \text{ e } -\frac{\mu_1^a}{T_1} = -\frac{\mu_2^a}{T_2} \quad (3.101)$$

e o equilíbrio só é obtido quando a temperatura e o potencial químico da espécie a são iguais dos dois lados da parede.

Exercício

Discuta o problema que resulta ao considerar subsistemas separados por uma parede que permita mudança de volume V mas não de E : a parede adiabática móvel. Pense se é possível mudar volumes, mas não haver mudança de energia...

3.10 Conexão com a Termodinâmica

As condições de equilíbrio que vimos nas seções anteriores: igualdade de temperatura, pressão ou potencial químico, permitem fazer o pulo e dizer que o

extremo da entropia de Shannon, sob os vínculos adequados, isto é, a entropia de Boltzmann-Gibbs é a entropia termodinâmica de Clausius. Não é por acaso que a forma funcional da entropia como função de (E, V, N) - os parâmetros chamados *extensivos* - recebe o nome de relação fundamental em termodinâmica. Se a tivermos, qualquer quantidade de interesse termodinâmico pode ser calculado. Temos uma relação fundamental $S(E, V, N) = \log \Omega$, onde o número Ω de microestados relevantes para o conjunto de experiências em questão pode ser calculado a partir do conhecimento do hamiltoniano do sistema. Temos, então a conexão da Termodinâmica com a Mecânica, seja ela Clássica ou Quântica. O trabalho que segue, no que diz respeito às aplicações consiste em calcular a entropia. É comum que este cálculo não seja trivial. Muitas das técnicas de aproximação que serão desenvolvidas requerem uma compreensão profunda do problema. Os capítulos que seguem darão exemplos de aplicações.

O formalismo apresentado traz um paralelo total com a termodinâmica. A determinação do estado de equilíbrio em termodinâmica é feito a partir de um princípio de extremo. Ao liberar o sistema de um vínculo, ou ao impor um novo vínculo, um novo estado de equilíbrio resulta. A escolha dentro de todos os possíveis estados compatíveis com os vínculos é feito pela maximização da entropia, expressa pela relação fundamental. Note que os outros estados compatíveis com os vínculos impostos, não ocorrem. Mas cada tal estado poderá ocorrer se for feita a imposição de algum novo vínculo específico adicional. A escolha do estado termodinâmico é feita dentre aqueles que satisfazem os vínculos que sabemos existirem e não mais que isso.

Qual é a justificativa por trás desse princípio de extremo? Uma possibilidade é que o estudante de Física ao chegar a esse ponto esteja acostumado a princípios de extremo, que terá visto em Mecânica Clássica ou Quântica, Eletromagnetismo e suas aplicações, e.g. Óptica. Devido a esse costume talvez não ousará discutir a possibilidade de mais um. A conexão com a teoria de informação nos traz a justificativa do princípio de extremo. O estado escolhido, não pelo sistema, mas por nós, é aquele que faz menos hipóteses não justificadas. Pela primeira vez um princípio de extremo está sendo apresentado sem dizer que essa é a forma como funciona a natureza. Neste caso esta é a forma como funciona a maneira de fazer previsões o mais honestas possíveis.

Neste ponto o estudante de Física pode voltar a argumentar: “Porque o que eu sei (vínculos impostos) tem alguma influência sobre o que o experimental mede no laboratório?”. A resposta é simples. Não tem nenhuma influência! Mas aquilo que *sabemos* tem uma influência direta sobre as *previsões* que fazemos. Se a informação que temos não for boa, suficiente, relevante...as previsões teóricas sobre as experiências, serão igualmente ruins. Voltamos ao debate entre subjetivo e objetivo? Só mais uma vez. O método pode parecer subjetivo ao ser olhado do ponto de vista que as previsões que eu faço dependem do que sei. Como poderia ser diferente? Deveriam depender do que não sei? A Mecânica Estatística fornece um conjunto de regras, se duas pessoas tiverem a mesma informação sobre o sistema e aplicarem as regras de forma adequada, farão previsões iguais. Nada pode ser mais objetivo que isso. Se eles tiverem informações diferentes, muito possivelmente farão previsões diferentes e o resul-

tado experimental servirá de juiz. Se a contagem de estados for feita usando a informação dada pela Física Clássica poderá haver erros ao comparar com a experiência. Repetimos, os primeiros resultados da Mecânica Quântica foram obtidos de forma a que as previsões da Mecânica Estatística refletissem os resultados experimentais no problema do Corpo Negro. Não foi mudada a ME mas sim a informação usada.

Os parâmetros extensivos escalam com o tamanho do sistema. Isto está claro para E , V e N . Devemos provar a extensividade da entropia. Isto é postulado na termodinâmica. Neste trabalho também o é, isso ocorreu no momento que colocamos aditividade ante agrupamentos. Há possibilidade de mudar este “postulado” e obter outra termodinâmica? Este é um tópico de grande discussão no momento mas que não nos interessa aqui ¹².

A equação 3.78 mostra sob que condições ocorre a extensividade. Consideremos um sistema num estado (E, V, N) com interações de curto alcance. Imaginemos, sem fazer, uma separação em dois sistemas 1 e 2 como nas seções anteriores. Façamos a separação em K partes

$$S(E, V, N) = \sum_{i=1, K} S(E_i, V_i, N_i) \quad (3.102)$$

com $E = \sum_{i=1, K} E_i$, $V = \sum_{i=1, K} V_i$ e $N = \sum_{i=1, K} N_i$. Se os E_i , V_i e N_i das partes forem iguais, teremos

$$S(E, V, N) = KS(E/K, V/K, N/K) \quad (3.103)$$

Podemos estender isto para números reais em geral e não só para inteiros e concluímos que se E , V e N forem proporcionalmente mudados a entropia também o será e chamamos essa propriedade de extensividade. Além dos parâmetros extensivos temos os intensivos T , P e μ identificados pelo papel que desempenham na determinação do estado de equilíbrio de dois sistemas em contato entre si, isolados do resto do mundo. Os parâmetros intensivos, que são derivadas parciais de quantidades extensivas com respeito a outras extensivas, não mudam com a escala. Seja T a temperatura de um sistema. Consideremos que é composto de duas partes, S , E , V e N de cada parte muda mas a temperatura é a mesma.

Para pequenas mudanças nos parâmetros extensivos, a entropia ($S = \log \Omega$) muda de acordo com

$$dS(E, V, N) = \left(\frac{\partial S}{\partial E} \right)_{VN} dE + \left(\frac{\partial S}{\partial V} \right)_{EN} dV + \left(\frac{\partial S}{\partial N} \right)_{EV} d\mu \quad (3.104)$$

¹²Faremos apenas um breve comentário. A entropia cruzada pode ser usada para definir uma distância, ao menos para pequenas diferenças. Há muitas formas de introduzir distância. Há algumas preferíveis a outras. A forma de Kullback-Leibler é a única invariante ante Markov embeddings (Ver Cenkov) Se há um princípio geral de inferência, então entropias não extensivas podem levar a resultados no mínimo estranhos ao lidar com sistemas independentes. Previsões de medidas termodinâmicas feitas sobre um sistema na terra sob a hipótese que existe a estrela A dariam resultados diferentes sob a hipótese que não existe a estrela. Fazendo a medida poderíamos decidir se existe ou não a estrela. Aparentemente este tipo de argumento não convence todos os pesquisadores e ainda há cálidas discussões. (inserir referências)

e usando as definições dos parâmetros intensivos

$$dS(E, V, N) = \frac{1}{T}dE + \frac{P}{T}dV - \frac{\mu}{T}d\mu \quad (3.105)$$

3.11 Ensemble Canônico

Há vezes em que uma experiência termodinâmica é feita numa situação diferente daquela em que o sistema está dentro de paredes isolantes, rígidas e impermeáveis. Por exemplo podemos olhar uma transformação a temperatura constante, a pressão constante ou a potencial químico constante. Queremos investigar o que significa isto em termos de vínculos de informação. Devemos olhar de novo a seção 3.5. Obtivemos resultados gerais para casos onde o valor esperado de certas quantidades é conhecido. O ensemble canônico de Gibbs é obtido quando o sistema não está isolado, mas a temperatura é fixa. Como implementar o vínculo informacional que a temperatura é fixa? Usamos a equação 3.62,

$$\lambda_i = \frac{\partial S(G_1, G_2, \dots, G_K)}{\partial G_i}$$

que nos dá indícios de como prosseguir.

Suponhamos que não fixamos a temperatura mas sim o valor esperado da energia. Como isso é feito numa experiência? Ficará claro a seguir. É dado que o valor esperado da energia

$$\langle \mathcal{H} \rangle = \int \prod_i dq_i dp_i P(q_i, p_i) \mathcal{H}(q_i, p_i) \quad (3.106)$$

é fixo, podemos supor que não o conhecemos, mas tem um valor definido. Se soubessemos esse valor poderíamos aplicar os resultados da seção 3.5 e obter a distribuição - de Boltzmann, de Boltzmann-Gibbs, canônica - em analogia à equação 3.57, que agora toma a forma

$$P(\{q_i, p_i\}) = \frac{e^{-\beta \mathcal{H}}}{Z} \quad (3.107)$$

onde o multiplicador de Lagrange que chamamos β em lugar de λ , pode ser obtido da equação 3.62

$$\beta = \frac{\partial S(E, V, N)}{\partial E} \quad (3.108)$$

e usamos a notação $E = \langle \mathcal{H} \rangle$. Mas a derivada da entropia com respeito à energia é o inverso da temperatura. Portanto identificamos β com o inverso da temperatura:

$$\beta = \frac{1}{T} \quad (3.109)$$

O curioso deste resultado é que ao fazer previsões para uma experiência onde conhecemos o multiplicador de Lagrange, mas não o valor esperado, agimos da

mesma forma que se soubessemos o valor esperado. Saber o valor do multiplicador de Lagrange, do ponto de vista da informação, é equivalente a saber o valor esperado da energia.

A função de partição Z adquire uma interpretação importante. Note que a equação 3.25 toma a forma

$$Z = \int \prod_i dq_i dp_i e^{-\beta \mathcal{H}} \quad (3.110)$$

e a equação 3.59 fica assim

$$E = \langle \mathcal{H} \rangle = - \frac{\partial \log Z}{\partial \beta} \quad (3.111)$$

A expressão para a entropia (o extremo), que obtemos substituindo a distribuição canônica no funcional de entropia de Shannon nos dá

$$S(E, V, N) = - \int \prod_i dq_i dp_i P(q_i, p_i) \log P(q_i, p_i) \quad (3.112)$$

$$= - \int \prod_i dq_i dp_i \frac{e^{-\beta \mathcal{H}}}{Z} \log \left(\frac{e^{-\beta \mathcal{H}}}{Z} \right) \quad (3.113)$$

$$= - \int \prod_i dq_i dp_i \frac{e^{-\beta \mathcal{H}}}{Z} (-\beta \mathcal{H} - \log Z) \quad (3.114)$$

$$= \beta \langle \mathcal{H} \rangle - \beta \log Z \quad (3.115)$$

$$= E - \beta \log Z. \quad (3.116)$$

Rearranjando obtemos $-\log Z/\beta = E - TS$, ou seja podemos identificar o logaritmo da função de partição e a energia livre de Gibbs

$$F(T, V, N) = - \frac{1}{\beta} \log Z(\beta, V, N) \quad (3.117)$$

$$F = E - TS \quad (3.118)$$

3.11.1 Comparação da Entropia nos diferentes Ensembles

Encontramos duas formulações, o ensemble microcanônico e o canônico, e ainda encontraremos outras, que descrevem situações experimentais diferentes. O que tem em comum e o que tem de diferente? Esperamos que as entropias encontradas sejam as mesmas? A resposta é sim e não. Não de forma trivial, a informação é diferente, a situação experimental sob análise é diferente e portanto o valor numérico também poderá sê-lo. No entanto ao descrever sistemas físicos que incluem um número muito grande de graus de liberdade os valores sob condições “iguais” e as derivadas ou seja, as equações de estado, serão iguais.

Exercício: Ensemble Grande Canônico

Considere a situação em os vínculos de informação sobre a experiência são o valor esperado da energia e do número de partículas. A notação que usamos é que $P(q_i, p_i, n) = P(n)P(\{q_i, p_i\}|n)$ denota a probabilidade atribuímos a que o sistema tenha n partículas e elas estejam no estado $\{q_i, p_i\}$. Obviamente dado um valor de n o número de variáveis depende de n , e.g. poderia ser $6n$ se o sistema estiver em três dimensões. O equivalente aos vínculos $\langle g_k \rangle$ da equação 3.54 tomam a forma

- $\sum_{n=1}^{\infty} \int P(q_i, p_i, n) \prod dq_i dp_i = 1$
- $\sum_{n=1}^{\infty} \int \mathcal{H}(\{q_i, p_i\}, n) P(q_i, p_i, n) \prod dq_i dp_i = \langle \mathcal{H} \rangle = E$
- $\sum_{n=1}^{\infty} \int n P(q_i, p_i, n) \prod dq_i dp_i = \langle n \rangle = N$

Encontre a distribuição de probabilidades chamada grande canônica ou grã canônica que maximiza o funcional de entropia sob os vínculos acima. Mostre que o multiplicador de Lagrange para o vínculo sobre o número de partículas está relacionado ao potencial químico $\lambda = \mu/T$, veja a equação 3.100. Voltaremos ao ensemble grande canônico em problemas importantes como o gás ideal clássico e quântico que veremos nos próximos capítulos.

3.12 Apêndice: Multiplicadores de Lagrange

O problema de encontrar pontos extremos ou só estacionários de funções sujeitos a vínculos é muito vasto. Damos algumas idéias básicas sem preocupação com rigor, para lembrar o estudante de técnicas que deveriam ser vistas em Cálculo 2 ou curso equivalente.

Seja o problema

- P_{livre} : Queremos encontrar um ponto (x^*, y^*) dentro de uma certa região C no plano real onde uma função $f(x, y)$ tem localmente um valor estacionário.

Fácil, tome as derivadas parciais e resolva o sistema $\partial_x f = 0$, $\partial_y f = 0$.

Queremos, a seguir resolver um caso mais difícil.

- P_{vinc} : Suponha que não procuramos o resultado em qualquer lugar de C , mas especificamente queremos um ponto estacionário entre aqueles pontos que satisfazem $\phi(x, y) = c$, que supomos descreva uma curva no plano que está parcialmente contida em C e chamaremos γ .

A solução do parágrafo anterior dificilmente nos dá a resposta pois seria uma coincidência se (x^*, y^*) caísse encima dessa curva.

A solução a esta classe de problema foi proposta por Lagrange. Considere a classe de funções $F_\lambda(x, y)$ que dependem do chamado multiplicador de Lagrange λ :

$$F_\lambda(x, y) = f(x, y) + \lambda(\phi(x, y) - c) \quad (3.119)$$

Note que se o ponto de coordenadas x e y estiver na curva γ então F_λ e f tem o mesmo valor. Repetindo: F_λ e f tem o mesmo valor se o vínculo que “ x e y estão sobre γ ” for respeitado.

Consideremos o P_{ivre} mas para a função $F_\lambda(x, y)$. O problema é novamente simples ¹³. Resolvemos o sistema

$$\frac{\partial F_\lambda}{\partial x} = 0 \quad (3.120)$$

$$\frac{\partial F_\lambda}{\partial y} = 0, \quad (3.121)$$

onde λ ainda não foi especificado. A resposta depende de valor escolhido para λ , isto é define uma curva ρ , parametrizada por $\lambda : (x^*(\lambda), y^*(\lambda))$ onde F_λ é extremo. Agora voltamos ao problema P_{vinc} . Da resposta à dupla pergunta “onde f é máximo?” e “onde o vínculo é satisfeito?”, quando as duas são respondidas simultaneamente, decorre a solução. Substituímos a primeira por “onde F é máximo?” (resposta: em ρ) junto com a afirmação “ $f = F_\lambda$ sob a condição de estar em γ ”. Segue que queremos encontrar o cruzamento de γ e ρ . Basta escolhermos $\lambda = \lambda_*$ tal que $\phi(x^*(\lambda_*), y^*(\lambda_*)) = c$, o resultado é um extremo para f e satisfaz o vínculo.

Agora procure um livro de cálculo e preencha os detalhes. Discuta também como lidar com casos em que o extremo está na borda de C . Há vínculos que são representados por desigualdades, os nomes de Kuhn e Tucker estão associados a esta extensão. Em muitos casos isto pode útil mas não no curso introdutório.

¹³a não ser que não seja....

Capítulo 4

Aplicações Simples

Devemos aplicar os resultados do último capítulo a sistemas físicos. Começamos pelos mais simples e gradualmente olharemos para sistemas mais ricos.

4.1 Sistemas Paramagnéticos

Os modelos de spin clássicos tem um papel importante na Física Estatística. Suponha um cristal, com impurezas diluídas, que tem um momento magnético (spin) localizado. Os graus de liberdade são discretos, por exemplo estudaremos o caso em que a variável σ_i toma um de dois valores possíveis, que representa o momento alinhado na direção de um campo magnético externo, ou na direção oposta. Note que a origem deste tipo de variável deverá ser procurado dentro da Mecânica Quântica, mas o interessante é podemos tratar isto simplesmente como uma variável que toma dois estados sem mais necessidade de Mecânica Quântica. Este modelo é muito mais difícil de justificar do que seu tratamento matemático a seguir. A palavra diluída foi usada para justificar que os estados individuais dos spins são independentes, sem interação entre eles. Este sistema é chamado Paramagnético, em oposição a por exemplo um sistema Ferromagnético, onde as interações entre spins podem levar a sistemas com fases termodinâmicas com propriedades coletivas diferentes, separadas por transições de fase. Mais sobre este tipo de modelo no capítulo ??.

Para descrever um sistema paramagnético com N spins clássicos não interagentes entre si, na presença de um campo magnético h , o Hamiltoniano relevante é uma soma de termos que representam a energia de um único spin:

$$\mathcal{H} = -\mu_0 h \sum_{i=1}^N \sigma_i \quad (4.1)$$

onde $\sigma = 1$ ou -1 para spin meio e μ_0 é uma constante que caracteriza o momento magnético da impureza.

4.1.1 Ensemble Microcanônico: Energia fixa

O número de microestados compatíveis com a informação $\mathcal{H} = E$ é calculado assim. Sejam N_+ o número de spins no estado 1 e N_- o número no estado -1 . Temos que

$$N = N_+ + N_- \quad (4.2)$$

$$E = -\mu_0 h(N_- + N_+) \quad (4.3)$$

de onde segue que

$$N_+ = \frac{1}{2}(N - \mathcal{E}) = \frac{N}{2}(1 - \epsilon) \quad (4.4)$$

$$N_- = \frac{1}{2}(N + \mathcal{E}) = \frac{N}{2}(1 + \epsilon) \quad (4.5)$$

onde $\epsilon = \mathcal{E}/N = E/(N\mu_0 h)$. O número de estados será dado por

$$\Omega(E, N) = \frac{N!}{N_+!N_-!} = \frac{N!}{\left(\frac{1}{2}(N - \mathcal{E})\right)!\left(\frac{1}{2}(N + \mathcal{E})\right)!}. \quad (4.6)$$

Usando a fórmula de Stirling para a expansão do fatorial obtemos

$$\begin{aligned} S(E, N) &= N \log N - N_+ \log N_+ - N_- \log N_- \\ &= -N \left(\frac{1 - \epsilon}{2} \log(1 - \epsilon) + \frac{1 + \epsilon}{2} \log(1 + \epsilon) \right) \end{aligned} \quad (4.7)$$

Note que a entropia é proporcional a N e portanto extensiva. A temperatura é

$$\begin{aligned} \frac{1}{T} &= \frac{\partial S(E, N)}{\partial E} = \frac{\partial \epsilon}{\partial E} \frac{\partial S}{\partial \epsilon} \\ &= \frac{1}{2\mu_0 h} \log \left(\frac{1 - \epsilon}{1 + \epsilon} \right) \end{aligned} \quad (4.8)$$

que pode ser invertida, dando

$$E = \mu_0 h N \epsilon = -\mu_0 h N \tanh \mu_0 h T \quad (4.9)$$

e para a magnetização por grau de liberdade

$$m := \left\langle \frac{\sum_{i=1}^N \sigma_i}{N} \right\rangle = \mu_0 \tanh \frac{\mu_0 h}{T} \quad (4.10)$$

O interesse do experimental é o de determinar como o sistema responde a mudanças dos parâmetros de controle. Por exemplo a susceptibilidade magnética descreve como muda a magnetização quando o campo externo muda mantida a temperatura constante

$$\chi_T = \left(\frac{\partial m}{\partial h} \right)_T = \frac{\mu_0^2}{T} \frac{1}{\cosh^2 \frac{\mu_0 h}{T}} \quad (4.11)$$

mostrada na figura 4.3. Note que para campo h constante χ tem um pico que se desloca cada vez mais para valores de $T \rightarrow 0$ quando h diminui. Para $h = 0$ o comportamento de $\chi \propto T^{-1}$ é conhecido como lei de Curie.

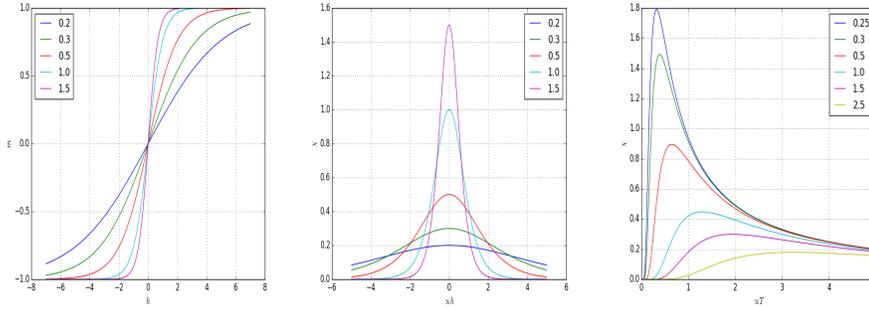


Figura 4.1: Esquerda: A magnetização por grau de liberdade m como função do campo h externo para diferentes valores da temperatura inversa β ($\mu_0 = 1$). Centro: Suscetibilidade magnética (equação 4.11) como função do campo h , para diferentes valores de β . Direita: $\chi(T)$ para diferentes valores de h .

4.1.2 Ensemble Canônico: valor esperado da Energia fixo

Estudamos novamente o sistema paramagnético descrito na secção anterior nas condições experimentais em que a temperatura é mantida fixa no valor $T = \beta^{-1}$. Novamente, maximizando a entropia sujeita a que o valor da energia tem um valor fixo, obtemos a distribuição canônica, dada por

$$P(\{\sigma_i\}) = \frac{e^{\beta\mu_0 h \sum_{i=1}^N \sigma_i}}{Z} \quad (4.12)$$

com a função de partição

$$Z(\beta, N, h) = \sum_{\{\sigma_i\}} e^{\beta\mu_0 h \sum_{i=1}^N \sigma_i} \quad (4.13)$$

esta somatória é sobre os 2^N microestados possíveis de N spins. Note a diferença com o microcanônico, onde o número de microestados considerados eram só aqueles com um dada energia.

A soma pode ser feita porque na exponencial as variáveis σ entram numa soma de termos que não incluem mais que uma variável. Veremos no próximo capítulo que quando isso não acontece, a situação é bem mais complicada. Agora temos

$$\begin{aligned} Z(\beta, N, h) &= \prod_{i=1}^N \sum_{\sigma_i = \pm 1} e^{\beta\mu_0 h \sigma_i} \\ &= (2 \cosh \beta\mu_0 h)^N \end{aligned} \quad (4.14)$$

Note que da equação 4.13 obtemos

$$\frac{\partial \log Z}{\partial h} = \beta\mu_0 \left\langle \sum_{i=1}^N \sigma_i \right\rangle \quad (4.15)$$

e da equação 4.14

$$\frac{\partial \log Z}{\partial h} = \beta \mu_0 N \tanh(\beta \mu_0 h) \quad (4.16)$$

e portanto chegamos a uma equação igual a equação 4.10.

Exercício Resfriamento magnético: Na figura 4.3 vemos curvas de magnetização para diferentes valores da temperatura. Suponha que o sistema paramagnético esteja em contato térmico com um sistema maior, chamado reservatório, a temperatura T_{alta} e campo inicialmente nulo. O campo é elevado até h_{alto} . A sua temperatura será T_{alta} e a magnetização pode ser obtida da figura 4.3. A seguir o sistema é isolado termicamente e o campo é lentamente reduzido até um valor pequeno h_{peq} . É possível supor que a magnetização se manteve constante? Como será a temperatura do sistema paramagnético? É possível usar esse método no laboratório? Procure referências sobre *magnetic cooling*.

4.2 Gás Ideal: Microcanônico

Vimos no último capítulo que para começar a descrever um sistema físico, isto é, fazer previsões a respeito de experiências, precisamos saber que tipo de experiências queremos abordar. Começaremos por olhar os sistemas físicos mais simples, gases a baixa pressão em equilíbrio. O que significa baixa pressão? Baixo em relação ao que? Se a pressão for baixa a distância entre as moléculas será grande e as interações, que decaem com a distância poderão ser desprezadas. O gás ideal é o nome que se dá a um modelo de gás de moléculas ou átomos que não interagem entre si. Sem interações, a energia é puramente cinética. O problema de determinar a relação fundamental dentro do formalismo do ensemble microcanônico para o gás ideal consiste simplesmente em calcular o volume do espaço de fase sob as seguintes condições:

- Supomos que os únicos graus de liberdade são as variáveis que descrevem as coordenadas e momentos do centro de massa de cada molécula.
- O gás está dentro de uma caixa de paredes isolantes, rígidas e impermeáveis de volume V .
- O número de moléculas é N .
- O hamiltoniano é $\mathcal{H}(\{q_i, p_i\}) = \sum_{i=1}^N \frac{p_i^2}{2m}$ e tem valor próximo a E :

$$E \leq \mathcal{H}(\{q_i, p_i\}) \leq E + \delta E, \quad (4.17)$$

uma relação que denotaremos simplesmente por $\mathcal{H} = E$

Precisamos descrever com mais cuidado o que significa o estado que queremos estudar. Se as posições e momentos das moléculas com os rótulos k e k' fossem trocadas entre si, isto daria lugar a um novo estado? Não poderíamos perceber que houve uma troca, portanto não deveria ser considerado diferente. Se a pergunta experimental pudesse distinguir entre microestados onde moléculas foram

trocadas, deveríamos levar em consideração que as moléculas são distinguíveis. Se em lugar de moléculas estivessemos falando de moedas e de novo perguntássemos sobre a probabilidade de m caras para cima quando N foram jogadas, a informação sobre a direção aonde apontam os narizes seria irrelevante. As moedas são distinguíveis entre si e no entanto houve uma divisão por fatoriais que eliminam a recontagem de microestados que diferem por troca das direções dos narizes, como se fossem indistinguíveis. Suponha que as moléculas, ou átomos, tem narizes pintados, faria diferença? Ainda não saberíamos se houve a troca. Quer que consideremos os átomos indistinguíveis ou distinguíveis num nível de detalhes mais profundo, não interessa neste ponto: as perguntas que queremos responder não levam em conta este ponto sobre a possível distinguibilidade dos átomos. No contexto das perguntas que queremos responder eles são indistinguíveis. Voltando à idéia de agrupamentos que Shannon considerou ao ser levado à forma da entropia, definimos um estado intermediário, um *mesoestado*, que agrupa as $N!$ configurações onde uma partícula tem posição e momento q_1, p_1 , outra q_2, p_2 , etc. sem que importe qual das partículas é a que tem q_1, p_1 , etc. Isto significa que o volume do espaço de fases, como medida do número de *mesoestados* diferentes compatíveis com os vínculos, não é

$$\Omega_{dist}(E, V, N) = \int_{\mathcal{H}=E, V} \prod dq_i dp_i \quad (4.18)$$

mas

$$\Omega(E, V, N) = \int_{\mathcal{H}=E, V} \frac{\prod dq_i dp_i}{N! h^{3N}} \quad (4.19)$$

onde h é por agora uma constante com dimensões $[h] = [qp]$, introduzida para que Ω seja adimensional. Por agora a deixamos livre de interpretações, mas voltaremos a ela ao estudar gases quânticos. Isto está de acordo com a idéia que não atribuímos uma entropia ao sistema. Há várias formas de descrever um sistema, cada uma identificará os estados e lhes atribuirá probabilidades, e finalmente uma entropia. Qual dessas entropias terá relevância experimental? Depende da experiência sendo feita.

A parte das coordenadas da integral na expressão 4.19 é muito simples. Para cada partícula temos uma integral sobre os valores possíveis das coordenadas, portanto cada partícula contribui com um fator V .

$$\Omega(E, V, N) = \frac{V^N}{N!} \int_{\mathcal{H}=E} \prod dp_i, \quad (4.20)$$

a integral que resta é sobre um casca da hipersfera de $3N$ dimensões, raio \sqrt{E} e espessura $\delta R = \delta E / (2\sqrt{E})$. Transforme para coordenadas esféricas, a parte radial é fácil porque o raio é fixo. Para a parte angular, considere uma integral auxiliar em n dimensões que sabemos calcular, e.g um produto de n integrais unidimensionais da gaussiana: $1 = \prod_{i=1, n} \int \exp(-x_i^2/2) dx_i / \sqrt{2\pi}$. Mude para coordenadas esféricas e separe a parte angular da radial, obtendo 1/radial=angular. Da integral para a parte radial aparece a função $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$, vista na próxima equação.

Assim

$$\Omega(E, V, N) = \frac{V^N}{N!} \frac{2\pi^{\frac{3N}{2}}}{\Gamma(3N/2)} E^{\frac{3N-1}{2}} \delta E \quad (4.21)$$

Novamente voltamos a insistir que a discussão acima não tem **nada** a ver com um tratamento quântico das partículas em oposição a um tratamento clássico. Trabalhando assim não chegaremos a nenhum dos resultados característicos de sistemas de partículas quânticas (férmions ou bósons), no entanto há vários livros que não colocam o fator $1/N!$ durante um tratamento clássico e depois, como ficará claro daqui a pouco, encontram entropias não extensivas e chamam o resultado de *paradoxo*¹. Como foi Gibbs o primeiro a chamar atenção para este fato, chamam-no de paradoxo de Gibbs e resolvem o problema introduzindo a Mecânica Quântica que faz a contagem de estados de forma diferente, lida diretamente com os mesoestados. O fator $1/N!$ aparece e resolve o problema, levando o estudante a acreditar que o tratamento clássico do gás ideal estava errado porque não era quântico. Claro que o tratamento clássico estará errado a baixas temperaturas, mas a altas temperaturas o tratamento clássico também está certo. Um argumento de apóio histórico pode ser dado dizendo que Gibbs resolveu este problema sem o uso (antes) da Mecânica Quântica.

Assim temos pela equação 3.74 que a entropia é

$$\begin{aligned} S(E, V, N) &= \log \Omega(E, V, N) \\ &= \log \left(\frac{V^N}{N!} \frac{2\pi^{\frac{3N}{2}}}{\Gamma(3N/2)} E^{\frac{3N-1}{2}} \delta E \right), \end{aligned}$$

usando a expansão de Stirling para o fatorial: $\log n! = n \log n - n$ e $\Gamma(n) = (n-1)!$ obtemos

$$S(E, V, N) = N \left(\log \frac{V}{N} + \frac{3}{2} \log \frac{E}{N} - \frac{1}{2} + \frac{3}{2} \log \frac{3\pi}{2} \right) + \dots \quad (4.22)$$

onde os termos desprezados incluem termos de ordem N^0 e N^{-1} . Como estamos interessados em N grande podemos desprezá-los. Usando o fato que a forma dada pela expressão é homogênea, podemos introduzir as densidades $u = E/N$ de energia, $s = S/N$ de entropia e o volume por partícula $v = V/N$.

$$s(u, v) = \frac{S}{N} = \log v + \frac{3}{2} \log u - \frac{1}{2} + \frac{3}{2} \log \frac{3\pi}{2}. \quad (4.23)$$

¹Uma definição pragmática de paradoxo: algo que está errado mas é fruto de um raciocínio a primeira vista correto.

A seguir investigamos as equações de estado obtidas ao tomar as derivadas da entropia com respeito à energia, volume e número de partículas

$$\left(\frac{\partial S}{\partial E}\right)_{V,N} = \frac{1}{T} = \frac{3N}{2E} \quad (4.24)$$

$$\left(\frac{\partial S}{\partial V}\right)_{E,N} = \frac{P}{T} = \frac{N}{V} \quad (4.25)$$

$$\left(\frac{\partial S}{\partial N}\right)_{E,V} = \frac{-\mu}{T} = \log \frac{V}{N} + \frac{3}{2} \log \frac{E}{N} - 3 + \frac{3}{2} \log \frac{3\pi}{2} \quad (4.26)$$

(VERIFICAR CONTAS) onde reconhecemos as célebres equações de estado:

$$PV = NT \quad (4.27)$$

$$E = \frac{3}{2}NT \quad (4.28)$$

4.3 Gás ideal: Canônico

Analisamos a situação experimental onde a temperatura é mantida fixa no valor T , assim como o volume e o número de partículas.

Sabemos como encontrar distribuições que satisfazem certos vínculos sobre valores esperados e sabemos que se o valor esperado for o da energia, a distribuição depende do multiplicador de Lagrange β que é o inverso da temperatura: $\beta = 1/T$. Agimos então como se soubessemos que o valor esperado do hamiltoniano $\mathcal{H}(\{q_i, p_i\}) = \sum_{i=1}^N \frac{p_i^2}{2m}$ fosse

$$E = \langle \mathcal{H}(\{q_i, p_i\}) \rangle. \quad (4.29)$$

A densidade de probabilidades é

$$P(\{q_i, p_i\}) = \frac{\exp(-\beta \mathcal{H}(\{q_i, p_i\}))}{Z} \quad (4.30)$$

onde

$$Z = \int \frac{\prod dq_i dp_i}{N!} \exp(-\beta \mathcal{H}), \quad (4.31)$$

que impõe a normalização

$$1 = \int \frac{\prod dq_i dp_i}{N!} P(\{q_i, p_i\}). \quad (4.32)$$

A equação relevante é

$$E = -\frac{\partial \log Z}{\partial \beta} \quad (4.33)$$

que mostra que a devemos calcular a função de partição $Z(T, V, N)$, que é possível neste caso pois as integrais envolvidas são fáceis. Devido a que o hamiltoniano só tem a parte cinética e não depende das coordenadas dos átomos,

a integração é simples. Introduzimos, para referência posterior a função de partição ζ_i da partícula i :

$$\zeta_i = \int dq_i dp_i \exp(-\beta \mathcal{H}_i), \quad (4.34)$$

onde \mathcal{H}_i é a contribuição de uma única partícula ao Hamiltoniano e se todas as partículas forem idênticas ζ_i não depende do índice i . Segue que

$$Z = \frac{1}{N!} \zeta^N. \quad (4.35)$$

A integração sobre as coordenadas é análoga à secção anterior, dando V^N .

Para integrar os momentos usamos:

$$\int_{-\infty}^{\infty} e^{-\frac{\beta}{2m} p^2} dp = \sqrt{2\pi \frac{m}{\beta}} \quad (4.36)$$

dN vezes, onde d é geralmente 3, a dimensão do espaço, mas poderia ter outro valor se os átomos do gás ideal se restringissem a uma superfície.

Assim

$$Z = \frac{V^N}{N!} \left(2\pi \frac{m}{\beta}\right)^{\frac{3N}{2}} \quad (4.37)$$

$$\zeta = V \left(2\pi \frac{m}{\beta}\right)^{\frac{3}{2}} \quad (4.38)$$

e a equação 4.33 dá:

$$E = \frac{3N}{2\beta} = 3N \frac{T}{2}. \quad (4.39)$$

Este é um exemplo do chamado *Teorema*² da Equipartição: Cada grau de liberdade que contribui de forma quadrática ao hamiltoniano leva a uma contribuição à energia média $T/2$. Se for usada a escala de temperatura Kelvin, $\beta^{-1} = k_B T$, e o número de átomos do gás for escrito $N = nN_0$ em termos do número de Avogadro N_0 e o número de moles n , a energia média será $E = \frac{3N}{2} N_0 k_B T = \frac{3n}{2} RT$, onde a *constante universal dos gases* $R = N_0 k_B$. Este é um caso particular do princípio de equipartição da energia: cada grau de liberdade quadrático contribui $k_B T/2$ para a energia e $k_B/2$ para o calor específico a volume constante C_V .

A função de partição está relacionada à energia livre (equação 3.116 através de $-\beta F = \log Z$, portanto $S = \log Z + \beta E$. A pressão no ensemble canônico, obtida através a equação 4.25 $\left(\frac{\partial S}{\partial V}\right)_{V,N} = \frac{P}{T}$, nos leva novamente à equação de estado $PV = NT$ ou nas unidades usuais $PV = Nk_B T_k = nRT_k$.

Usando a expansão de Stirling para $\log N!$

$$-\beta F = N \log V + \frac{3N}{2} \log \frac{2\pi m}{\beta} - N \log N + N \quad (4.40)$$

²Não devemos usar a palavra teorema em Física, teoremas se aplicam a estruturas matemáticas e não a questões empíricas. A falha do *teorema* em prever corretamente resultados empíricos levou à conclusão, não que a matemática estava errada, mas que essa particular estrutura matemática não era útil para acomodar os fatos empíricos.

e definindo as densidades $f = F/N$, $v = V/N$ e $\epsilon = E/N$

$$-\beta f = \log v + \frac{3}{2} \log \frac{2\pi m}{\beta} + 1 \quad (4.41)$$

de onde fica claro que se o fatorial não estivesse presente as densidades não seriam intensivas.

4.4 Mistura de gases ideais clássicos

Num volume V temos as espécies químicas C_1, C_2, \dots, C_m e suas populações, em número de moléculas são N_1, N_2, \dots, N_m . Usaremos os índices $l = 1, \dots, m$ para denotar espécies e $i = 1, \dots, \sum_i^m N_l$, para indexar as moléculas. O hamiltoniano pode ser aproximado por

$$\mathcal{H} = \sum_i \mathcal{H}_l(i), \quad (4.42)$$

e podemos supor que

$$\mathcal{H}_l(i) = \frac{\mathbf{p}_i^2}{2m_l} + \varepsilon_l(s_i) \quad (4.43)$$

onde s_i representa os graus de liberdade internos da molécula i que é do tipo l e $\varepsilon_l(s_i)$ a energia associada. A função de partição associada a essa molécula do tipo l é

$$\zeta_l = \int d\mathbf{q}d\mathbf{p} \sum_s e^{-\beta \mathcal{H}_l}, \quad (4.44)$$

e a função de partição do sistema

$$Z(\beta, V, N_1, \dots, N_m) = \prod_{l=1, \dots, m} \frac{\zeta_l^{N_l}}{N_l!}, \quad (4.45)$$

dado que duas moléculas do mesmo tipo não podem ser distinguidas experimentalmente, mas poderão se forem de tipo diferentes.

A energia livre é dada por

$$-\beta F = \log Z = \sum_l (N_l \log \zeta_l - \log N_l + 1). \quad (4.46)$$

A pressão pode ser facilmente calculada pois cada fator ζ_l contribui com um fator V , portanto

$$-\beta F = \sum_l (N_l \log V) + \dots, \quad (4.47)$$

onde não escrevemos termos independentes do volume, segue que

$$\beta p = \left(\frac{\partial \log Z}{\partial V} \right)_{T, N_1, \dots, N_l, \dots, N_m} \quad (4.48)$$

$$\beta p = \sum N_l \frac{1}{V} = \beta \sum_{l=1}^m p_l, \quad (4.49)$$

ou seja a pressão total é a soma das pressões parciais que cada espécie teria, caso fosse a única espécie no volume V à temperatura β^{-1} .

4.4.1 Reações e potenciais químicos

Suponha que as moléculas possam reagir de tal forma que numa dada reação os parâmetros c_l são números inteiros que denotam o número de moléculas que são consumidas ($c_l < 0$) ou produzidas ($c_l > 0$) cada vez que uma dada reação ocorre. É claro que se ocorrem muitas reações, as variações dN_l devem satisfazer

$$\frac{dN_1}{c_1} = \frac{dN_2}{c_2} = \dots = \frac{dN_l}{c_l} = \dots = \frac{dN_m}{c_m} \quad (4.50)$$

A energia livre $F(T, V, \{N_l\}_{l=1\dots m})$ é mínima no equilíbrio, portando flutuações dN_l devem satisfazer

$$0 = \Delta F = \sum_l \left(\frac{\partial F}{\partial N_l} \right)_{T, V, N_1, \dots, N_{l'} \neq l, \dots, N_m} dN_l. \quad (4.51)$$

Lembrando que estas derivadas de F são os potenciais químicos

$$\mu_l = \left(\frac{\partial F}{\partial N_l} \right)_{T, V, N_{l'} \neq l} \quad (4.52)$$

podemos ver que

$$\sum_l c_l \mu_l = 0. \quad (4.53)$$

E dada a relação entre a energia livre e a função de partição ($-\beta F = \log Z$), obtemos

$$\mu_l = -T \log \frac{\zeta_l}{N_l} = f_l - T \log N_l \quad (4.54)$$

onde introduzimos $f_l = -\beta^{-1} \log \zeta_l$, a soma $F_0 = \sum_l f_l$ e temos

$$0 = T \sum_l c_l \log \frac{\zeta_l}{N_l} = - \sum_l f_l - T \sum_l \log N_l^{c_l} \quad (4.55)$$

$$\prod_l N_l^{c_l} = e^{-\beta F_0} = K(T, V, \{C\}), \quad (4.56)$$

onde a função $K = e^{-\beta F_0}$ é comumente chamada de constante de equilíbrio. Esta expressão permite calcular as populações da mistura de espécies. Devemos esperar que funcione melhor tanto em solução diluída, ou na forma de gases a

baixa pressão, para que se possam desprezar as interações entre as moléculas. Como exemplo suponha a reação



portanto $c_A = 2$, $c_B = 1$, $c_C = -2$ e

$$\frac{N_A^2 N_B}{N_C^2} = K(T, V, A, B, C) \quad (4.58)$$

4.5 Sólido Clássico

Uma pequena extensão nos permite obter resultados para um sólido. Consideramos um sistema formado por átomos localizados numa rede cristalina periódica. À diferença com o caso do gás, temos pouca incerteza sobre a posição de cada átomo: deve estar perto de um sítio da rede. Supomos que o átomo está sujeito a forças harmônicas que se originam da interação com os outros átomos. Isso é um salto muito grande pois estamos dizendo que o hamiltoniano

$$\mathcal{H}(\{q_i, p_i\}) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} + V(\{q_i\}) \quad (4.59)$$

pode ser aproximado por

$$\mathcal{H}(\{q_i, p_i\}) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} + \sum_i V(q_i) \quad (4.60)$$

onde o potencial $V(q_i)$ só depende das coordenadas do átomo i e ainda mais, que é harmônico

$$\mathcal{H}(\{q_i, p_i\}) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} + \sum_i \frac{m\omega^2}{2} \mathbf{q}_i^2 \quad (4.61)$$

A função de partição é novamente gaussiana e além da equação 4.36 usamos

$$\int_{-\infty}^{\infty} e^{-\frac{\beta m \omega^2}{2} q^2} dq = \sqrt{2\pi \frac{1}{\beta m \omega^2}}. \quad (4.62)$$

A função de partição

$$Z = \left(\sqrt{2\pi \frac{m}{\beta}} \sqrt{2\pi \frac{1}{\beta m \omega^2}} \right)^{3N} \quad (4.63)$$

A energia, obtida derivando o logaritmo de Z dá:

$$E = \frac{6TN}{2} = 3NT (= 3nRT_K), \quad (4.64)$$

e $C_V = 3R$ por mol. Temos de novo a equipartição, cada grau de liberdade quadrático no Hamiltoniano contribui com $k_B T_K/2$ para a energia. Se em lugar de seis graus de liberdade, como no sólido, ou três, como no gás, o sistema for descrito por mais graus de liberdade quadráticos poderemos ter outros “ $k_B T/2$ ”. Considere por exemplo moléculas diatómicas. Podemos ter rotação em torno de eixos perpendiculares à linha que une os núcleos dos átomos com contribuições cinéticas do tipo $I\omega^2/2$, onde I é um momento de inércia e ω é a frequência angular. E para rotações em torno dessa linha? Dependendo da temperatura em que a experiência é feita esses termos contribuirão ou não. Classicamente poderíamos dizer que o momento de inércia para rotações em torno do eixo longitudinal da molécula é pequeno e pode ser desprezado. Mas isso não deveria convencer ninguém pois a dedução acima não pergunta sobre o tamanho dos coeficientes dos termos quadráticos. Para entender porque podem ser desprezados precisaremos um pouco de mecânica quântica.

Ainda experimentalmente, sob certas condições (pressão baixa, temperatura alta) o calor específico de um gás é aproximadamente constante. Mas decresce com a temperatura e não há como explicar isso classicamente. A primeira explicação, devida a Einstein, usando uma quantização da energia pré-mecânica quântica, nos dá uma boa idéia porque o calor específico diminui com a temperatura.

4.6 Sólido de Einstein

Einstein considera os $3N$ osciladores descritos acima quantizados e independentes, o que significa que o microestado do sistema é descrito por $\{n_1, n_2, \dots, n_N\}$ e a energia

$$\mathcal{H}(\{n_1, n_2, \dots, n_{3N}\}) = \hbar\omega \sum_{i=1}^{3N} \left(\frac{1}{2} + n_i \right) \quad (4.65)$$

4.6.1 Microcanônico

Temos a situação experimental em que o sistema está isolado e a energia é fixa, assim como o número de osciladores e o volume. As perguntas experimentais colocadas ao sistema não distinguem entre situações em que os osciladores i e j estão no estado (n_i, n_j) ou no estado (n'_i, n'_j) , desde que $n_i + n_j = n'_i + n'_j$. Assim devemos olhar para $K = \sum_{i=1}^{3N} n_i$ e N como as quantidades que efetivamente determinam o estado do sistema. Em termos delas a energia é dada por

$$\mathcal{H}(\{n_1, n_2, \dots, n_{3N}\}) = E = \frac{3N}{2}\hbar\omega + K\hbar\omega \quad (4.66)$$

O número de microestados $\Omega(E, N)$ é dado pelas combinações indistinguíveis do ponto de vista experimental de $3N + K$ objetos, formados por duas classes de

³N osciladores tridimensionais



Figura 4.2: Exemplo: uma das 1287 configurações distintas com $K = 8$ quanta e $N = 5$ osciladores

objetos diferentes mas indistinguíveis entre si dentro da classe: $3N$ osciladores e K quanta de energia $\hbar\omega$:⁴

$$\Omega(E, N) = \frac{(3N + K)!}{(3N)!K!} \quad (4.67)$$

Definimos $\varepsilon = E/(N\hbar\omega)$, a energia por partícula medida em unidades de $\hbar\omega$. Para a entropia temos

$$S(E, N) = (3N + K) \log(3N + K) - 3N \log 3N - K \log K \quad (4.68)$$

$$= N \left(\left(\varepsilon + \frac{3}{2} \right) \log \left(\varepsilon + \frac{3}{2} \right) - \left(\varepsilon - \frac{3}{2} \right) \log \left(\varepsilon - \frac{3}{2} \right) \right) \quad (4.69)$$

onde usamos $\varepsilon = 3/2 + K/N$. Assim podemos calcular a temperatura

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E} \right)_N = \frac{1}{\hbar\omega} \log \frac{\varepsilon + \frac{3}{2}}{\varepsilon - \frac{3}{2}} \quad (4.70)$$

e invertendo temos

$$\varepsilon = \frac{3}{2} + \frac{3}{e^{\frac{\hbar\omega}{T}} - 1} \quad (4.71)$$

$$E = \frac{3N\hbar\omega}{2} + \frac{3N\hbar\omega}{e^{\frac{\hbar\omega}{T}} - 1} \quad (4.72)$$

onde o primeiro termo é a energia de ponto zero, que Einstein não tinha como descobrir, e a segundo pode ser interpretado definindo $\langle n \rangle$ o número médio de quanta em cada oscilador

$$E = \frac{3N\hbar\omega}{2} + 3N\hbar\omega \langle n \rangle \quad (4.73)$$

e $\langle n \rangle = \frac{1}{e^{\frac{\hbar\omega}{T}} - 1}$, que é chamada distribuição de Planck.

Podemos agora calcular o calor específico a volume constante por partícula

$$c_v = \frac{1}{N} \frac{\partial E}{\partial T} = 3 \frac{(\hbar\omega)^2 e^{\frac{\hbar\omega}{T}}}{T^2 (e^{\frac{\hbar\omega}{T}} - 1)^2}, \quad (4.74)$$

e tomando os limite de pequenas temperaturas vemos que o calor específico não é constante mas cai com a temperatura

$$c_v \approx 3 \frac{(\hbar\omega)^2 e^{-\frac{\hbar\omega}{T}}}{T^2}, \quad (4.75)$$

⁴Na realidade devemos considerar $3N - 1$ osciladores mas não faz diferença

enquanto que para valores altos da temperatura

$$c_v \approx 3 \quad (4.76)$$

devemos colocar as unidades costumeiras para obter $c_v \approx 3R$ por mol.

4.6.2 Canônico

Consideremos o mesmo modelo acima, mas nas condições experimentais onde a temperatura é fixa. Sabemos que do ponto de vista de informação devemos agir como se soubessemos que o valor esperado da energia tem um valor fixo, a densidade de probabilidade atribuída aos microestados será obtida maximizando a entropia de Shannon sujeita aos vínculos de normalização e de $\langle \mathcal{H} \rangle = E$:

Exercício

Mostre que

$$P(\{n_i\}) = \frac{e^{-\beta \mathcal{H}(\{n_i\})}}{Z} \quad (4.77)$$

onde

$$Z(\beta, N) = \sum_{n_1 \dots n_{3N}} e^{-\beta \mathcal{H}} \quad (4.78)$$

impõe o vínculo de normalização. Mostre ainda que a temperatura é o inverso do multiplicador de Lagrange β : $T^{-1} = \beta$

O cálculo da função de partição é simples porque o hamiltoniano é uma soma de termos e cada um deles depende somente de uma variável de ocupação, ou seja os osciladores são independentes entre si. Assim

$$Z(\beta, N) = \left[\sum_{n=0}^{\infty} e^{-\beta(n+\frac{1}{2})\hbar\omega} \right]^{3N}. \quad (4.79)$$

$$Z(\beta, N) = e^{-\frac{3N\beta}{2}\hbar\omega} [1 + a + a^2 + a^3 + \dots]^{3N}, \quad (4.80)$$

onde $a = e^{-\beta\hbar\omega}$ e somando a série geométrica

$$Z(\beta, N) = e^{-\frac{3N\beta}{2}\hbar\omega} \left[\frac{1}{1 - e^{-\beta\hbar\omega}} \right]^{3N}. \quad (4.81)$$

Usando $E = -\partial \log Z / \partial \beta$, obtemos

$$E = \frac{3N\hbar\omega}{2} + \frac{3N\hbar\omega}{e^{\beta\hbar\omega} - 1} \quad (4.82)$$

que é exatamente o resultado obtido no ensemble microcanônico

4.6.3 Gases ideais clássicos com estrutura interna quântica

a ser escrito ...

4.7 Ensemble Grande Canônico

Revisitamos o ensemble grande canônico que será útil no estudo de gases quânticos. Numa situação experimental em que o sistema não está isolado mas pode trocar energia com um reservatório e o número de partículas pode mudar ao longo do tempo, os vínculos informacionais podem ser escritos

$$E = \langle E_\alpha \rangle, \quad N = \langle N_\alpha \rangle \quad V = \text{fixo}. \quad (4.83)$$

Vai ficar claro que os possíveis valores de E_α serão influenciados pelo valor de V . Denotamos por α os estados do sistema de número indefinido de partículas, portanto α não só diz respeito à energia do sistema, mas ao número de partículas. Mais tarde faremos uma especificação cuidadosa do significado de cada estado α , por agora a integração sobre todos os possíveis valores de α será denotada por uma soma:

$$E = \sum_{\alpha} P_{\alpha} E_{\alpha} \quad N = \sum_{\alpha} P_{\alpha} N_{\alpha}, \quad 1 = \sum_{\alpha} P_{\alpha}, \quad (4.84)$$

que inclui a normalização.

Maximizamos a entropia de Shannon sujeita a estes vínculos

$$H[P_{\alpha}] = - \sum_{\alpha} P_{\alpha} \log P_{\alpha} + \lambda_0 (1 - \sum_{\alpha} P_{\alpha}) + \beta (E - \sum_{\alpha} P_{\alpha} E_{\alpha}) + \lambda_1 (N - \sum_{\alpha} P_{\alpha} N_{\alpha}), \quad (4.85)$$

e obtemos

$$P_{\alpha} = e^{1-\lambda_0} e^{-\beta E_{\alpha} - \lambda_1 N_{\alpha}}. \quad (4.86)$$

Resta impor os vínculos para determinar o valor dos multiplicadores de Lagrange. Como já deve parecer usual ao leitor, a introdução da função de partição permite eliminar λ_0

$$P_{\alpha} = \frac{1}{\Xi(\beta, \lambda_1, V)} e^{-\beta E_{\alpha} - \lambda_1 N_{\alpha}} \quad (4.87)$$

onde

$$\Xi(\beta, \lambda_1, V) = \sum_{\alpha} e^{-\beta E_{\alpha} - \lambda_1 N_{\alpha}}. \quad (4.88)$$

Usamos, como é costume, a letra ξ maiúscula Ξ para enfatizar que isto levará a um potencial termodinâmico diferente da energia livre de Helmholtz F , que aparece no ensemble canônico. A entropia termodinâmica será identificada com o máximo da entropia de Shannon, ou seja substituímos a distribuição de probabilidades, equação 4.87, na entropia de Shannon 4.85

$$S(E, N, V) = - \sum_{\alpha} P_{\alpha} \log P_{\alpha}, \quad (4.89)$$

para obter

$$S(E, N, V) = \log \Xi + \beta E + \lambda_1 N. \quad (4.90)$$

O leitor deveria ver isto como semelhante à equação 3.116. Não é a mesma equação porque estamos descrevendo um sistema em situação experimental diferente, onde o número de partículas está fixo somente através do seu valor médio. Para interpretar o significado de λ_1 , devemos fazer o análogo à equação 3.60. Primeiro notemos que

$$\frac{\partial \log \Xi}{\partial \beta} \Big|_{\lambda_1, V} = -E \quad (4.91)$$

$$\frac{\partial \log \Xi}{\partial \lambda_1} \Big|_{\beta, V} = -N \quad (4.92)$$

$$\frac{\partial \log \Xi}{\partial V} \Big|_{\beta, \lambda_1} = -\frac{P}{T} \quad (4.93)$$

Agora, usando a regra da cadeia

$$\frac{\partial S}{\partial E} \Big|_N = \frac{\partial \log \Xi}{\partial \beta} \frac{\partial \beta}{\partial E} + \frac{\partial \log \Xi}{\partial \lambda_1} \frac{\partial \lambda_1}{\partial E} + \frac{\partial \beta}{\partial E} E + \frac{\partial \lambda_1}{\partial E} N + \beta \quad (4.94)$$

$$\frac{\partial S}{\partial E} \Big|_N = \beta, \quad (4.95)$$

obtemos novamente a interpretação que β é o inverso da temperatura. Derivando com respeito a N :

$$\frac{\partial S}{\partial N} \Big|_E = \frac{\partial \log \Xi}{\partial \beta} \frac{\partial \beta}{\partial N} + \frac{\partial \log \Xi}{\partial \lambda_1} \frac{\partial \lambda_1}{\partial N} + \frac{\partial \beta}{\partial N} E + \frac{\partial \lambda_1}{\partial N} N + \lambda_1 \quad (4.96)$$

$$\frac{\partial S}{\partial N} \Big|_E = \lambda_1, \quad (4.97)$$

de onde obtemos, usando a definição do potencial químico (ver equação 3.100)

$$\lambda_1 = -\frac{\mu}{T} = -\beta\mu. \quad (4.98)$$

Segue que

$$\Xi(\beta, \mu) = \sum_{\alpha} e^{-\beta(E_{\alpha} - \mu N_{\alpha})}. \quad (4.99)$$

Para referência futura note que

$$E = \langle E_{\alpha} \rangle = -\frac{\partial \log \Xi}{\partial \beta} \Big|_{\mu, V} \quad (4.100)$$

$$N = \langle N_{\alpha} \rangle = \beta^{-1} \frac{\partial \log \Xi}{\partial \mu} \Big|_{\beta, V} \quad (4.101)$$

$$P = -\beta^{-1} \frac{\partial \log \Xi}{\partial V} \Big|_{\beta, \mu} \quad (4.102)$$

4.8 Gases Quânticos

Consideremos um gás ideal de partículas (átomos, moléculas, elétrons) numa caixa. A parte mais difícil no que segue é a descrição dos microestados relevantes para nossa descrição de um gás quântico. Precisamos descrever estados de N_α partículas mas só sabemos resolver (nesta altura do curso) problemas de uma partícula única num potencial. O potencial é simples, infinito fora de uma caixa cúbica de tamanho L^3 , zero dentro. Resolvemos o problema de uma partícula única, usando a equação de Schrödinger independente do tempo:

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r}) + V(\mathbf{r})\psi(\mathbf{r}) = \epsilon\psi(\mathbf{r}) \quad (4.103)$$

obtendo os autoestados de energia $\epsilon_{\mathbf{q}}$, associados ao momento $\mathbf{p} = \hbar\mathbf{q}$

$$\epsilon_{\mathbf{q}} = \frac{\hbar^2\mathbf{q}^2}{2m}. \quad (4.104)$$

As soluções são construídas como superposições de ondas planas que satisfaçam as condições de contorno. O mais simples é usar condições periódicas de contorno (PBC) tal que se $\mathbf{r} \rightarrow \mathbf{r} + \mathbf{u}_L$ onde as componentes \mathbf{u}_L são números inteiros vezes L , a função de onda não deve se alterar: $\psi(\mathbf{r}) = \psi(\mathbf{r} + \mathbf{u}_L)$. A primeira vista as PBC não são intuitivas. Em duas dimensões o espaço é a superfície de um toroide. Imaginamos que uma partícula ao bater numa parede sai da caixa e entra pela parede oposta com o mesmo momento e na mesma posição. Mas a vida tende a ser mais fácil ao usar PBC e vale a pena o esforço de se acostumar com elas. Poderíamos impor que a função de onda vai a zero nas paredes. O resultado será o mesmo no limite de volumes grandes, chamado de limite termodinâmico. O resultado é que os vetores de onda permitidos são

$$\mathbf{q} = (q_x, q_y, q_z), \text{ com } q_x = \frac{2\pi}{L}l_x, q_y = \frac{2\pi}{L}l_y, q_z = \frac{2\pi}{L}l_z \quad (4.105)$$

onde l_x, l_y e l_z são inteiros que tomam valores $0, \pm 1, \pm 2, \dots$

Estes estados de uma partícula são os tijolos de construção dos estados de N_α partículas. Para N_α partículas não interagentes a função de onda seria a solução da equação de Schrödinger independente do tempo:

$$\left(-\frac{\hbar^2}{2m}\sum_i^{N_\alpha}\nabla_i^2 + V(\mathbf{r})\right)\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{N_\alpha}) = E_\alpha\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{N_\alpha}), \quad (4.106)$$

onde o Laplaciano com índice i atua sobre as coordenadas da partícula i . Separação de variáveis resolve quase todo o problema:

Exercício: Mostre que o produto $\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{N_\alpha}) = \prod_i \psi(\mathbf{r}_i)$ de soluções da equação 4.103 é solução da equação 4.106.

Mas isto não resolve o problema. Suponha duas partículas e duas soluções de 4.103, $\psi_a(\mathbf{r}_1)$ e $\psi_b(\mathbf{r}_2)$. A função $\Psi_{a1,b2} = \psi_a(\mathbf{r}_1)\psi_b(\mathbf{r}_2)$ é solução de 4.106

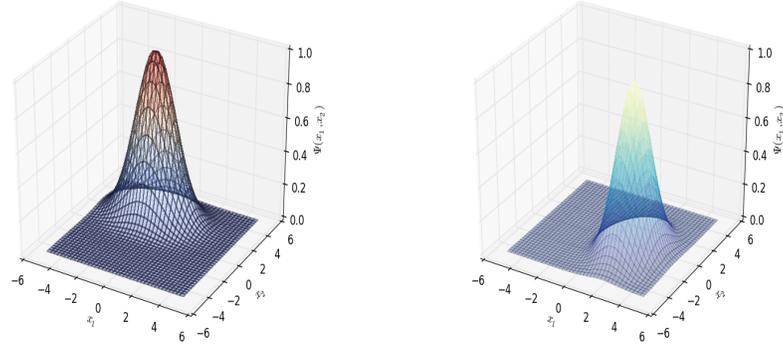


Figura 4.3: (linhas grossas) $\Psi_{a1,b2} = \psi_a(x_1)\psi_b(x_2) \neq \psi_a(x_2)\psi_b(x_1) = \Psi_{a2,b1}$ (linhas finas)

mas não leva em conta a indistinguibilidade das partículas. Note que colocar a partícula 1 no estado a e a 2 no b é diferente de colocar a 1 em b e a 2 em a , i.e.

$$\Psi_{a1,b2} \neq \Psi_{a2,b1} \quad (4.107)$$

Embora satisfaçam as equações de Schrödinger para uma e duas partículas estas funções não são aceitáveis como soluções do problema de duas partículas. No entanto a combinação

$$\Psi_{\pm}(\mathbf{r}_1, \mathbf{r}_2) \propto \psi_a(\mathbf{r}_1)\psi_b(\mathbf{r}_2) \pm \psi_a(\mathbf{r}_2)\psi_b(\mathbf{r}_1) \quad (4.108)$$

satisfaz $\Psi_{\pm}(\mathbf{r}_1, \mathbf{r}_2) = \pm \Psi_{\pm}(\mathbf{r}_2, \mathbf{r}_1)$, logo as densidades de probabilidade permanecem iguais quando as duas partículas são intercambiadas.

$$|\Psi_{\pm}(\mathbf{r}_1, \mathbf{r}_2)|^2 = |\Psi_{\pm}(\mathbf{r}_2, \mathbf{r}_1)|^2 \quad (4.109)$$

Isto sugere, mas não prova, que para N_{α} partículas os estados serão descritos por combinações de produtos de N_{α} funções do tipo $\psi(\mathbf{r}_i)$ que descrevem estados de uma partícula. As combinações devem ser tais que a generalização da equação 4.109 deva ser satisfeita. Isto em princípio poderia levar ao aparecimento de uma fase

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_k, \mathbf{r}_l, \dots) = e^{i\phi} \Psi(\mathbf{r}_1, \dots, \mathbf{r}_l, \mathbf{r}_k, \dots) \quad (4.110)$$

para uma fase qualquer ϕ . Não discutiremos o motivo, mas o fato é que $\phi = 0$ ou $\phi = \pi$. Isto ocorre para partículas com spin inteiro, chamados de bósons ($\phi = 0$), ou spin seminteiro, férmions $\phi = \pi$, respectivamente. Acho que este é o primeiro lugar nestas notas em que se pede do leitor a crença numa asserção para a qual não são apresentadas as evidências. Estes nomes honram a memória de Satyendra Bose e Enrico Fermi.

Historicamente Bose está localizado antes da Mecânica Quântica e portanto não tinha à sua disposição os argumentos acima. A discussão dele é simplificada pela forma como escolheu enumerar os estados de N_α partículas. Na função de onda de $N - \alpha$ as combinações de produtos de funções de uma partícula tem a seguinte similaridade: O número de vezes que uma função de onda de um determinado estado de uma partícula aparece num termo é o mesmo para todos os termos. Este número é chamado de número de ocupação do estado de uma partícula, pois indica não quais partículas estão nesse estado mas quantas. Portanto o estado $-N_\alpha$ será caracterizado por o conjunto de números de ocupação. Como cada estado -1 é caracterizado por um conjunto de números quânticos \mathbf{q} , então o estado $-N_\alpha$ é representado por $:(n_{\mathbf{q}_1}, n_{\mathbf{q}_2}, \dots, n_{\mathbf{q}_l}, \dots)$, com

$$N_\alpha = \sum_{\mathbf{q}} n_{\mathbf{q}}. \quad (4.111)$$

No caso do gás ideal o \mathbf{q} são simplesmente os momentos permitidos na caixa pelas condições de contorno e portanto

$$E_\alpha = \sum_{\mathbf{q}} n_{\mathbf{q}} \epsilon_{\mathbf{q}}, \quad (4.112)$$

pois a energia total do estado $-N_\alpha$ é a soma das energias dos estados -1 ocupados e simplesmente juntamos os termos iguais.

Exercício Mostre a equação 4.112 a partir da equação 4.106.

O que significa que uma função de onda seja simétrica (+) ou antisimétrica (-) ao intercambiar o índice de duas partículas? Consideremos

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_k, \mathbf{r}_l, \dots) = \pm \Psi(\mathbf{r}_1, \dots, \mathbf{r}_l, \mathbf{r}_k, \dots) \quad (4.113)$$

para o caso de duas partículas. Segue que

$$\Psi_{BE}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}} (\psi_a(\mathbf{r}_1)\psi_b(\mathbf{r}_2) + \psi_a(\mathbf{r}_2)\psi_b(\mathbf{r}_1)) \quad (4.114)$$

$$\Psi_{FD}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}} (\psi_a(\mathbf{r}_1)\psi_b(\mathbf{r}_2) - \psi_a(\mathbf{r}_2)\psi_b(\mathbf{r}_1)) \quad (4.115)$$

Para o sinal +, Ψ_{BE} não apresenta nenhum comportamento estranho se $a = b$, isto é se as duas partículas ocupam o mesmo estado-1. Já para o sinal -, Ψ_{FD} é zero se $a = b$, portanto os estados aceitáveis serão aqueles em não há dois férmions no mesmo estado. Temos o resultado que os estados $-N_\alpha$, que são descritos pelo conjunto de números de ocupação devem ser

$$\text{Bósons } \alpha \leftrightarrow (n_1, n_2, \dots) \text{ sem nenhuma restrição: } n_{\mathbf{q}} = 0 \dots \infty, \quad (4.116)$$

$$\text{Férmions } \alpha \leftrightarrow (n_1, n_2, \dots) \text{ restrição: } n_{\mathbf{q}} = 0, 1. \quad (4.117)$$

A generalização para mais de dois férmions foi feita por Slater que percebeu que a expressão 4.115 era um determinante. Se os estados de uma partícula

tiverem índice i e os vetores posição tiverem índice j , a função antissimétrica do estado de N_α partículas, será dada pelo determinante da matriz de Slater cujos elementos são $S_{ij} = \psi_i(\mathbf{r}_j)$.

Isto tem uma importância enorme sobre os resultados que obteremos a seguir. Podemos agora voltar a $\Xi(\beta, \mu, V) = \sum_\alpha e^{-\beta(E_\alpha - \mu N_\alpha)}$,

$$\begin{aligned}\Phi(\beta, \mu, V) &= -T \log \Xi \\ &= -T \log \sum_{n_{q_i}} e^{-\beta \sum_{q_i} n_{q_i} (\epsilon_{q_i} - \mu)} \\ &= -T \log \prod_{q_i} \sum_{n_{q_i}} e^{-\beta n_{q_i} (\epsilon_{q_i} - \mu)} \\ &= -T \sum_{q_i} \log \sum_{n_{q_i}} e^{-\beta n_{q_i} (\epsilon_{q_i} - \mu)}\end{aligned}\quad (4.118)$$

$$= \sum_{q_i} \Phi_{q_i}.\quad (4.119)$$

Ainda não sabemos como fazer a soma sobre os q e trataremos disso mais tarde. Por agora nos concentraremos nas somas sobre n_{q_i} , levando em conta 4.116 e 4.117. Para o caso de Bósons, a estatística Bose-Einstein será governada pelo potencial

$$\begin{aligned}\Phi_{q_i}^{BE} &= -T \log \sum_{n_{q_i}=0}^{\infty} e^{-\beta n_{q_i} (\epsilon_{q_i} - \mu)} \\ &= -T \log(1 + x + x^2 + x^3 + \dots) \\ &= -T \log\left(\frac{1}{1-x}\right) \\ &= T \log(1 - e^{-\beta(\epsilon_{q_i} - \mu)})\end{aligned}\quad (4.120)$$

A convergência da série geométrica só ocorrerá caso $x = e^{-\beta(\epsilon_{q_i} - \mu)}$ seja menor que 1. Como o menor valor de ϵ_{q_i} , no limite de volume infinito, é zero, devemos ter um potencial químico negativo

$$x < 1 \rightarrow \mu_{BE} < 0\quad (4.121)$$

As consequências físicas desta trivialidade matemática são impressionantes.

Para Férmions teremos a chamada estatística de Fermi-Dirac, a soma é ainda mais simples, e o potencial termodinâmico será:

$$\begin{aligned}\Phi_{q_i}^{FD} &= -T \log \sum_{n_{q_i}=0}^1 e^{-\beta n_{q_i} (\epsilon_{q_i} - \mu)} \\ &= -T \log(1 + x) \\ &= -T \log(1 + e^{-\beta(\epsilon_{q_i} - \mu)}),\end{aligned}\quad (4.122)$$

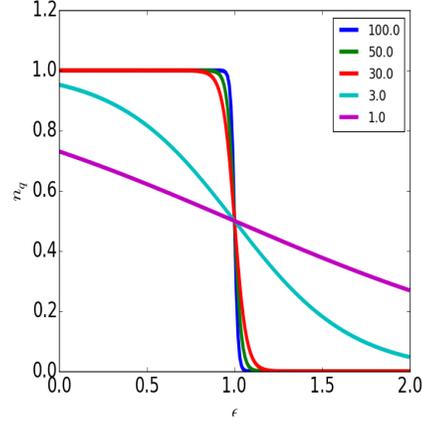


Figura 4.4: A função de Fermi: $n = (\exp(\beta(\epsilon - 1) + 1))^{-1}$, para diferentes valores de β .

Juntando os resultados

$$\Phi_{\mathbf{q}_i} = \mp T \log(1 \pm e^{-\beta(\epsilon_{\mathbf{q}_i} - \mu)}) \quad (4.123)$$

com $-+$ para férmions e $+-$ para bósons. Se a exponencial na expressão acima for pequena $e^{-\beta(\epsilon_{\mathbf{q}_i} - \mu)} \ll 1$ a expansão do logaritmo ($\log(1 + x) \approx x$) leva ao mesmo resultado tanto para férmions quanto bósons:

$$\Phi_{\mathbf{q}_i}^C = -T e^{-\beta(\epsilon_{\mathbf{q}_i} - \mu)} \quad (4.124)$$

que veremos, recupera os resultados para o limite clássico.

Uma quantidade central na discussão que segue é o *número médio de ocupação* $\bar{n}_{\mathbf{q}}$ que pode ser calculado a partir de (ver equação 4.101)

$$\bar{n}_{\mathbf{q}} = \langle n_{\mathbf{q}} \rangle = \frac{\partial \Phi_{\mathbf{q}}}{\partial \mu} \Big|_{\beta, V} . \quad (4.125)$$

Obtemos para bósons

$$\bar{n}_{\mathbf{q}}^{BE} = \frac{1}{e^{\beta(\epsilon_{\mathbf{q}_i} - \mu)} - 1} \quad (4.126)$$

e para férmions

$$\bar{n}_{\mathbf{q}}^{FD} = \frac{1}{e^{\beta(\epsilon_{\mathbf{q}_i} - \mu)} + 1} \quad (4.127)$$

Estas expressões são aparentemente parecidas, mas como veremos são muito diferentes para baixas temperaturas. A diferença ocorre devido a que o potencial químico se comporta de maneira muito diferente nos dois casos, para ambos os

casos é claro que pode ser determinado de forma implícita pelo vínculo que o número médio de partículas é dado

$$N_{BE} = \sum_{\mathbf{q}} \bar{n}_{\mathbf{q}} = \sum_{\mathbf{q}} \frac{1}{e^{\beta(\epsilon_{\mathbf{q}} - \mu)} - 1} \quad (4.128)$$

$$N_{FD} = \sum_{\mathbf{q}} \bar{n}_{\mathbf{q}} = \sum_{\mathbf{q}} \frac{1}{e^{\beta(\epsilon_{\mathbf{q}} - \mu)} + 1} \quad (4.129)$$

Na figura 4.4 mostramos a função de Fermi. Não é o número de ocupação porque ainda não sabemos como depende o potencial químico. O número de ocupação para BE é mais difícil de desenhar, e será deixado para após a análise do potencial químico. Para esta discussão precisamos entender a soma sobre os estados de uma partícula e para tanto é necessário introduzir a idéia de densidade de estados.

4.9 Densidade de Estados

Para um sistema unidimensional numa caixa de tamanho L , o número de estados entre os valores de momento q_x e $q_x + \Delta q_x$ é

$$\mathcal{N}_1 = \Delta l_x = \frac{L}{2\pi} \Delta q_x = \frac{L}{2\pi\hbar} \Delta p_x \quad (4.130)$$

Ao fazer a conta integrando sobre momentos num intervalo pequeno Δp_x

$$\int_{L, \Delta p_x} \frac{dx dp_x}{h} = \frac{L \Delta p_x}{h}. \quad (4.131)$$

Na equação 4.19 h era uma constante adimensional sem interpretação, agora vemos que pode ser interpretada como a constante de Planck $h = 2\pi\hbar$, o que não poderíamos ter feito classicamente.

Ao somar sobre estados quânticos substituiremos uma soma sobre autovalores por uma integral. Algumas vezes isso será permitido, mas outras não, portanto avançar cegamente substituindo somas por integrais pode ser perigoso.

Para três dimensões teremos simplesmente o produto das contribuições de cada uma

$$\mathcal{N}_3 = \frac{L^3 \Delta p_x \Delta p_y \Delta p_z}{(2\pi\hbar)^3}, \quad (4.132)$$

e em coordenadas esféricas, tomando os Δp como diferenciais, teremos a medida de integração

$$\frac{4\pi V p^2 dp}{(2\pi\hbar)^3}, \quad (4.133)$$

onde $V = L^3$ é o volume onde se encontra o gás. As somas da secção anterior são feitas sobre níveis de energia, e precisamos transformar esta medida de

integração sobre os momentos. No caso de interesse atual, com partículas livres dentro de uma caixa, a equação 4.104

$$\epsilon_{\mathbf{q}} = \frac{\hbar^2 \mathbf{q}^2}{2m}, \quad (4.134)$$

leva, usando $d\epsilon = \frac{d\epsilon}{dp} dp = \frac{p dp}{m}$, a

$$\mathcal{N}_3 = \frac{4\pi V}{(2\pi\hbar)^3} \sqrt{2m}^{\frac{3}{2}} \epsilon^{\frac{1}{2}} d\epsilon, \quad (4.135)$$

e introduzindo $\rho(\epsilon)$, a densidade de estados de energia de uma partícula não relativística em 3 dimensões

$$\rho(\epsilon)d\epsilon = \frac{4\pi V}{(2\pi\hbar)^3} \sqrt{2m}^{\frac{3}{2}} \epsilon^{\frac{1}{2}} d\epsilon = CV \epsilon^{\frac{1}{2}} d\epsilon, \quad (4.136)$$

onde

$$C = \frac{4\pi}{(2\pi\hbar)^3} \sqrt{2m}^{\frac{3}{2}} \quad (4.137)$$

Exploraremos a seguir a aproximação de somas por integrais:

$$\sum_{\mathbf{q}} (\cdot) \rightarrow \int (\cdot) \rho(\epsilon) d\epsilon. \quad (4.138)$$

Exercício Encontre a densidade de estados para partículas confinadas a uma região bidimensional no caso não relativístico ($\epsilon = p^2/2m$).

Exercício Para partículas (ultra)relativísticas, ou excitações sem massa, $\epsilon = cp$. Obtenha $\rho(\epsilon)$ em duas e três dimensões.

Há duas condições que devem ser satisfeitas para que o cálculo da soma pela integral seja possível. Uma, o espaçamento entre os níveis discretos de energia deve ser muito menor que as energias de uma partícula, que devem ser da ordem de $k_B T$. A segunda é que o número médio de ocupação de cada estado que contribui para a integral seja pequeno. A primeira é trivialmente satisfeita. A segunda é satisfeita no caso de férmions, mas como veremos mais adiante, não no caso de bósons, resultando em um fenômeno físico muito interessante.

O comprimento de onda térmico é calculado para uma partícula com energia cinética da ordem de $k_B T$, lembrando que $p = \hbar q = 2\pi\hbar/\lambda$

$$\frac{p^2}{2m} = k_B T \rightarrow \epsilon = \frac{1}{2m} \left(\frac{2\pi\hbar}{\lambda_T} \right)^2 \quad (4.139)$$

$$\lambda_T = \frac{2\pi\hbar}{\sqrt{2mk_B T}}. \quad (4.140)$$

O intervalo típico entre níveis de energia é

$$\delta\epsilon = \frac{1}{2m} \left(\frac{2\pi\hbar}{L} \right)^2 \quad (4.141)$$

e em geral $L \gg \lambda_T$ portanto desde este ponto de vista a integral é uma boa aproximação da soma. Para Férmions o maior número de ocupação possível é um, pois $N \gg 1$, então esta parte também é satisfeita. Veremos que para bósons isto nem sempre vale.

4.10 Gás de Fermi-Dirac

Tipicamente estaremos interessados em sistemas de elétrons. Primeiro um gás de elétrons que não interagem parece não muito natural, mas um gás de elétrons em um cristal com núcleos positivos pode se comportar como não interagente, pelo menos para alguns propósitos. O ponto de partida são as expressões 4.119 e 4.122. O limite de temperatura $T \rightarrow 0$ é o ponto de partida, pois permite passar ao estudo de pequenas temperaturas. Isso parece óbvio, mas o que não é óbvio é qual o regime de valores para considerar temperaturas altas ou baixas. Um resultado surpreendente é que em muitos problemas a escala da temperatura que determina pequeno ou grande é da ordem de alguns milhares de graus Kelvin. Ou seja, quase todo o regime de interesse é em muitos casos o de baixa temperatura.

O número de partículas médio do sistema é dado por uma integral sobre os estados de energia com densidade $\rho(\epsilon)$

$$\begin{aligned} N &= \sum_{\mathbf{q}} \bar{n}_{\mathbf{q}} = \sum_{\mathbf{q}} \frac{1}{e^{\beta(\epsilon_{\mathbf{q}} - \mu)} + 1} \\ &\rightarrow \int_0^{\infty} \rho(\epsilon) \frac{1}{e^{\beta(\epsilon - \mu)} + 1} d\epsilon, \end{aligned} \quad (4.142)$$

onde lembramos que junto com vínculo

$$E = \int_0^{\infty} \rho(\epsilon) \frac{\epsilon}{e^{\beta(\epsilon - \mu)} + 1} d\epsilon, \quad (4.143)$$

determinam o estado macroscópico do sistema. O potencial químico μ é uma função de T e N . No limite $T \rightarrow 0$, $\mu \rightarrow \mu_0$ que pode ser obtido de

$$N = \lim_{T \rightarrow 0} \int_0^{\infty} \rho(\epsilon) \frac{1}{e^{\beta(\epsilon - \mu)} + 1} d\epsilon, \quad (4.144)$$

usando que o número de ocupação se torna um função degrau, 1 para energias abaixo de μ_0 e zero acima

$$N = \int_0^{\mu_0} \rho(\epsilon) d\epsilon, \quad (4.145)$$

Isto define a maior energia que uma partícula poderá ter, chamada de energia de Fermi e uma temperatura (de Fermi) associada : $\mu_0 = \epsilon_F = k_B T_F$. Esta temperatura definirá a escala do que é grande ou pequeno. Para partículas não relativísticas em três dimensões é fácil calcular usando a expressão 4.136. Podemos definir o módulo do momento da partícula com maior energia, o momento de ... Fermi = p_F , definido por $\mu_0 = p_F^2/2m$. Mas falta um detalhe. O elétron tem spin 1/2, isso faz com que haja dois estados diferentes com a mesma energia. Em geral se o spin for s o número será $g = 2s + 1$. portanto a densidade deve ser multiplicada por g :

$$\begin{aligned}
N &= g \int_0^{\mu_0} \rho(\epsilon) d\epsilon, \\
&= gV \int_0^{\mu_0} C\epsilon^{\frac{1}{2}} d\epsilon = gVC \frac{2}{3} \epsilon_F^{\frac{3}{2}}, \quad (4.146)
\end{aligned}$$

$$\begin{aligned}
&= g \frac{4\pi V}{(2\pi\hbar)^3} \sqrt{2m} \frac{2}{3} \epsilon_F^{\frac{3}{2}} \\
&= \frac{gV(4\pi p_F^3)}{3(2\pi\hbar)^3}. \quad (4.147)
\end{aligned}$$

Segue que

$$p_F = 2\pi\hbar \left(\frac{3}{4\pi g}\right)^{\frac{1}{3}} \left(\frac{N}{V}\right)^{\frac{1}{3}} \propto \left(\frac{N}{V}\right)^{\frac{1}{3}} \quad (4.148)$$

e

$$\epsilon_F = \left(\frac{3}{2gC}\right)^{\frac{2}{3}} \left(\frac{N}{V}\right)^{\frac{2}{3}} \quad (4.149)$$

$$\mu_0 = \epsilon_F = \frac{(2\pi\hbar)^2}{2m} \left(\frac{3}{4\pi g}\right)^{\frac{2}{3}} \left(\frac{N}{V}\right)^{\frac{2}{3}} \propto \left(\frac{N}{V}\right)^{\frac{2}{3}} \quad (4.150)$$

A energia média por partícula

$$\bar{\epsilon}(T, N) = \frac{g}{N} \int_0^{\infty} \rho(\epsilon) \frac{\epsilon}{e^{\beta(\epsilon-\mu)} + 1} d\epsilon \rightarrow \bar{\epsilon} = \frac{\int_0^{\epsilon_F} \epsilon \rho(\epsilon) d\epsilon}{\int_0^{\epsilon_F} \rho(\epsilon) d\epsilon}, \quad (4.151)$$

$$\bar{\epsilon} = \frac{\int_0^{\epsilon_F} \epsilon^{\frac{3}{2}} d\epsilon}{\int_0^{\epsilon_F} \epsilon^{\frac{1}{2}} d\epsilon} = \frac{3}{5} \epsilon_F. \quad (4.152)$$

A energia total

$$E = \frac{3}{5} N \epsilon_F = 3 \frac{(2\pi\hbar)^2}{10m} \left(\frac{3}{4\pi g}\right)^{\frac{2}{3}} \left(\frac{N}{V}\right)^{\frac{2}{3}} N \propto (V)^{-\frac{2}{3}} N^{\frac{5}{3}} \quad (4.153)$$

O cálculo da entropia é educativo. Quanto esperamos que seja? Há quantas formas de arranjar os elétrons no estado fundamental? Eles são indistinguíveis. Dois no primeiro estado, dois no segundo estadoaté a energia de Fermi. Só uma maneira, logo $S(T=0) = 0$. A pressão é mais interessante ainda. Quanto esperamos que seja a pressão num gás ideal? Se o gás for clássico de $PV = NT$ temos $P(T=0) = 0$. Mas no caso de férmions o resultado é diferente:

$$P = - \left. \frac{\partial E}{\partial V} \right|_N = \frac{2E}{3V}. \quad (4.154)$$

No caso clássico tínhamos $E = 3NT/2 = \frac{3}{2}PV$ zero energia a temperatura zero, agora também temos

$$E = \frac{3}{2}PV, \quad (4.155)$$

mas neste caso não é zero.

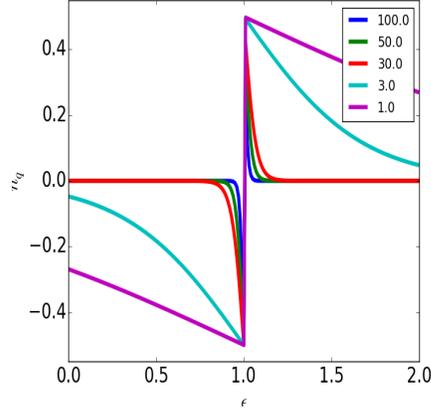


Figura 4.5: A função de Fermi a temperatura finita menos a função de Fermi a $T = 0$ (degrau): $f(\epsilon, \beta) = (\exp(\beta(\epsilon - 1)) + 1)^{-1} - \Theta(1 - \epsilon)$, para diferentes valores de β . A energia é medida em unidades do potencial químico.

Exercício(verificar numeros) No caso de metais $m = 9 \cdot 10^{-28}$ g, $g = 2$, 1 Mol de cobre = 63 g, densidade = 9 g/cm³, 1 elétron de condução por átomo. Isso leva a $N = 8.5 \cdot 10^{22}$ por cm³. Mostre que a temperatura de Fermi da ordem de 10^5 graus Kelvin.

4.10.1 Gás ideal de Férmions a temperatura finita

Pelo exercício acima vemos que temperaturas da ordem de 300K ainda são muito baixas. Olhe para a figura 4.4 para ver a forma da dependência do número de ocupação com a energia. Para os valores de $\beta = 100$, 30 é quase um degrau. A diferença $f(\epsilon, T) = n(\epsilon, \beta = T^{-1}) - n(\epsilon, \beta = \infty)$ aparece na figura 4.5. O significado prático desta figura é que se a temperatura for muito menor que a temperatura de Fermi T_F podemos calcular propriedades físicas usando expansões em série de potências de um parâmetro pequeno T/T_F . Do ponto de vista físico podemos esperar que nem todos os elétrons estarão envolvidos em excitações térmicas e a contribuição para, e.g. o calor específico, será devida a elétrons em níveis de energia onde a curva da figura 4.5 é diferente de zero.

As equações relevantes são 4.142 e 4.143 que dão respectivamente N e E como função de β e μ , ou seja são os vínculos informacionais em função dos multiplicadores de Lagrange. Vamos considerar a situação experimental em que a temperatura T e o número de partículas são dados. Em geral estaremos interessados em calcular integrais para uma função $a(\epsilon)$ bem comportada, do tipo

$$A(T, \mu) = \int_0^{\infty} \frac{a(\epsilon)}{e^{\beta(\epsilon - \mu)} + 1} d\epsilon, \quad (4.156)$$

onde $\mu = \mu(T, N)$. Lembrando que o parâmetro relevante $x = T/T_F$ é pequeno é razoável olhar a expansão em potências de x . Seguindo Sommerfeld, a expansão fica mais fácil olhando para

$$A(T, \mu) = A(0, \mu) + (A(T, \mu) - A(0, \mu)) \quad (4.157)$$

onde, usando a função f (ver figura 4.5)

$$A(0, \mu) = \int_0^\mu a(\epsilon) d\epsilon \quad (4.158)$$

$$\begin{aligned} \delta A = (A(T, \mu) - A(0, \mu)) &= \int_0^\infty a(\epsilon) \left(\frac{1}{e^{\beta(\epsilon-\mu)} + 1} - \Theta(\mu - \epsilon) \right) d\epsilon \\ &= \int_{-\infty}^\infty a(\epsilon) f(\epsilon, T) d\epsilon. \end{aligned} \quad (4.159)$$

onde usamos que f cai rapidamente para valores que se afastam de μ para estender o extremo de integração inferior até $-\infty$ e podemos expandir em série

$$\int_{-\infty}^\infty a(\epsilon) f(\epsilon, T) d\epsilon = \sum_k \frac{T^{k+1}}{k!} \left. \frac{d^k a}{d\epsilon^k} \right|_\mu \int_{-\infty}^\infty \left(\frac{\epsilon - \mu}{T} \right)^k f(\epsilon, T) \frac{d\epsilon}{T} \quad (4.160)$$

onde colocamos alguns fatores de T para poder introduzir a variável $u = (\epsilon - \mu)/T$. Devido a que f é uma função ímpar para reflexões em torno de $u = 0$ teremos que as ordens pares se anulam

$$\int_{-\infty}^\infty u^{2k} f(u, T) du = 0 \quad (4.161)$$

se anulam e o primeiro termo na expansão que contribui é:

$$\begin{aligned} \int_{-\infty}^\infty u f(u, T) du &= \int_{-\infty}^\infty u \left(\frac{1}{e^u + 1} - \Theta(-u) \right) du \\ &= \int_{-\infty}^0 u \left(\frac{1}{e^u + 1} - 1 \right) du + \int_0^\infty \frac{u}{e^u + 1} du \\ &= - \int_{-\infty}^0 u \left(\frac{1}{1 + e^{-u}} \right) du + \int_0^\infty \frac{u}{e^u + 1} du \\ &= 2 \int_0^\infty \frac{u}{e^u + 1} du = \frac{\pi^2}{6}. \end{aligned} \quad (4.162)$$

O próximo termo na expansão será o cúbico

$$2 \int_0^\infty \frac{u^3}{e^u + 1} du = \frac{7\pi^4}{60}. \quad (4.163)$$

Note que o expoente de T em 4.160 é $k + 1$, portanto a série em potências de T para $A(\beta, \mu)$ na equação 4.156 tem expoentes pares:

$$A(T, \mu) = \int_0^\mu a(\epsilon) d\epsilon + \left. \frac{da}{d\epsilon} \right|_\mu \frac{\pi^2}{6} T^2 + \left. \frac{d^3 a}{d\epsilon^3} \right|_\mu \frac{7\pi^4}{360} T^4 + \dots \quad (4.164)$$

Agora vamos a brincar com a expressão acima:

O potencial termodinâmico Φ e a energia E

Lembre que

$$\Phi(\beta, \mu, V) = -T \int_0^\infty \rho(\epsilon) \log(1 + e^{-\beta(\epsilon-\mu)}) d\epsilon.$$

Suponha que $\rho = CV\epsilon^b$, como é nos casos de interesse. Integramos por partes:

$$\Phi = -\frac{CV}{b+1} \epsilon^{b+1} \frac{1}{e^{\beta(\epsilon-\mu)} + 1} \Big|_0^\infty - \int_0^\infty \frac{CV}{b+1} \epsilon^{b+1} \frac{1}{e^{\beta(\epsilon-\mu)} + 1} d\epsilon$$

o termo de superfície é nulo, e o segundo termo (que muda duas vezes de sinal) é proporcional à energia

$$\begin{aligned} \Phi &= -\int_0^\infty \frac{\epsilon}{b+1} \rho(\epsilon) \frac{1}{e^{\beta(\epsilon-\mu)} + 1} d\epsilon \\ &= -\frac{1}{b+1} \int_0^\infty \epsilon \rho(\epsilon) \bar{n}(\epsilon) d\epsilon \\ &= -\frac{E}{b+1} \end{aligned} \tag{4.165}$$

$$= -\frac{2}{3} E \quad (b = 1/2) \tag{4.166}$$

onde a última linha vale para partículas com massa, em três dimensões, não relativísticas. Tanto para Φ quanto E teremos $a(\epsilon) = CV\epsilon^{b+1}$, segue que, para b geral e $b = 1/2$, respectivamente

$$\begin{aligned} \int_0^\mu CV\epsilon^{b+1} d\epsilon &= \frac{CV}{b+2} \mu^{b+2} = \frac{2CV}{5} \mu^{\frac{5}{2}} \\ \frac{da}{d\epsilon} \Big|_\mu &= (b+1)CV\mu^b = \frac{5}{2} CV\mu^{\frac{1}{2}} \\ \frac{d^3a}{d\epsilon^3} \Big|_\mu &= (b+1)b(b-1)CV\mu^{b-2} = -\frac{5}{8} CV\mu^{-\frac{3}{2}} \end{aligned}$$

e a expansão 4.164 até a primeira ordem não trivial (primeira derivada de a) fica, para $b = 1/2$:

$$\begin{aligned} \Phi(\beta, \mu, V) &= -\frac{2CV}{3} \left(\frac{2}{5} \mu^{\frac{5}{2}} + \frac{\pi^2}{4} \mu^{\frac{1}{2}} T^2 \right) \\ &= -\frac{4CV}{15} \mu^{\frac{5}{2}} \left(1 + \frac{5\pi^2}{8} \left(\frac{T}{\mu} \right)^2 \right) \end{aligned} \tag{4.167}$$

o termo fora do parêntese é a expressão para Φ em $T = 0$ com μ_0 em lugar de μ . A correção é de ordem $(\frac{T}{\mu})^2$, que efetivamente é de ordem $(T/T_F)^2$ pois $\mu \approx \mu_0$. Finalmente colocando a expressão de C dada pela equação 4.136

$$\Phi(\beta, \mu, V) = -\frac{4}{15} \frac{4\pi V g}{(2\pi\hbar)^3} \sqrt{2m^{\frac{3}{2}} \mu^{\frac{5}{2}}} \left(1 + \frac{5\pi^2}{8} \left(\frac{T}{\mu} \right)^2 \right) \tag{4.168}$$

O número de partículas N e o potencial químico μ

O número de partículas N para um dado valor de μ pode ser calculado com a função $a(\epsilon)$ apropriada ou usando a relação termodinâmica

$$\begin{aligned} N = -\frac{\partial\Phi}{\partial\mu}|_T &= \frac{2CV}{3} \left(\mu^{\frac{3}{2}} + \frac{\pi^2}{8} \mu^{-\frac{1}{2}} T^2 \right) \\ &= \frac{2CV}{3} \mu^{\frac{3}{2}} \left(1 + \frac{\pi^2}{8} \left(\frac{T}{\mu} \right)^2 \right). \end{aligned} \quad (4.169)$$

Se mantivermos μ fixo, o número N varia com T . Se considerarmos N fixo, μ dependerá de T . Em $T = 0$, para um dado μ teremos

$$N(T = 0) = N_0 = \frac{2CV}{3} \mu^{\frac{3}{2}}, \quad (4.170)$$

ou para um dado N o potencial químico será μ_0 , obtido invertendo 4.170

$$\mu_0 = \left(\frac{3N}{2CV} \right)^{\frac{2}{3}}. \quad (4.171)$$

Considerando N fixo temos duas expressões (4.169 e 4.171) que permitem, ao eliminar N obter uma relação entre os potenciais químicos a temperatura zero e finita:

$$N = \frac{2CV}{3} \mu_0^{\frac{3}{2}} = \frac{2CV}{3} \mu^{\frac{3}{2}} \left(1 + \frac{\pi^2}{8} \left(\frac{T}{\mu} \right)^2 \right). \quad (4.172)$$

Mantendo a expansão até segunda ordem em T , esta expressão pode ser invertida para obter

$$\mu = \mu_0 \left(1 + \frac{\pi^2}{8} \left(\frac{T}{\mu} \right)^2 \right)^{-\frac{2}{3}} \quad (4.173)$$

$$= \mu_0 \left(1 - \frac{\pi^2}{12} \left(\frac{T}{\mu} \right)^2 \right) \quad (4.174)$$

$$= \mu_0 \left(1 - \frac{\pi^2}{12} \left(\frac{T}{T_F} \right)^2 \right) \quad (4.175)$$

onde na segunda linha expandimos novamente $(1+x)^a \approx 1+ax$, e na terceira linha substituímos $\mu \approx \mu_0 = T_F$, pois o erro cometido ao fazer isso é de terceira ordem. O sinal negativo da correção mostra que o potencial químico diminui quando aumenta a temperatura, o que é esperado, pois a altas temperaturas no limite clássico, o resultado deve ser igual ao de bósons, que deve ser sempre negativo. Lembramos que $\mu = \partial E / \partial N$ a entropia constante, isto é a energia necessária para adicionar uma partícula a entropia constante. A altas temperaturas, se apenas se coloca uma nova partícula, o número de estados possíveis do sistema aumenta. Logo calor deve ser retirado para que a entropia desça até seu valor inicial. A baixas temperaturas a nova partícula só poderá entrar acima da energia de Fermi, sem que isso aumente o número de estados distinguíveis do sistema. Portanto $\mu > 0$.

A entropia S e o calor específico C_V

Usamos

$$S = -\left.\frac{\partial\Phi}{\partial T}\right|_{\mu,V}$$

e a expressão 4.167

$$S = \frac{CV\pi^2}{3}\mu^{\frac{1}{2}}T. \quad (4.176)$$

Mas por 4.170, $\mu^{\frac{1}{2}} = (3/2C)^{1/3}(N/V)^{1/3}$ e $\epsilon_F = (\frac{3}{2C})^{\frac{2}{3}}(\frac{N}{V})^{\frac{2}{3}}$

$$S = \frac{\pi^2 N}{2} \frac{2C}{3} \frac{V}{N} (3/2C)^{1/3} (N/V)^{1/3} T = \frac{\pi^2 N}{2} \frac{T}{T_F} \quad (4.177)$$

e o calor específico

$$C_V = T \left(\frac{\partial S}{\partial T} \right)_V = \frac{\pi^2 N}{2} \frac{T}{T_F}. \quad (4.178)$$

Podemos escrevê-lo de uma forma que lembra a equipartição

$$C_V = \frac{3}{2} N_{ef} \quad (4.179)$$

onde $N_{ef} = \frac{\pi^2}{3} \frac{T}{T_F}$ é o número efetivo de elétrons que contribui para o calor específico. São aqueles que podem ser excitados a estados desocupados por energias da ordem de $k_B T$ ⁵, os outros estão muito afundados nos estados de menor energia para poder sair da superfície de Fermi.

Exercício Calcule para cobre o valor de previsto de C_V/Nk_bT e compare com um valor experimental 0.8×10^{-4} . Suponha que a devido a interações com outros elétrons e com a rede cristalina os estados de uma partícula sejam melhor descritos por partículas livres com uma massa efetiva M_{ef} diferente da do elétron. Calcule M_{ef}/m . Esta prescrição da massa efetiva pode ser confrontada independentemente através de outro experimento. Suponha que ao colocar o cobre em um campo magnético H os elétrons de condução tenham estados com órbitas circulares. Estes estados tem uma frequência de ciclotron $\omega = eH/M_{ef}c$, que pode ser medida pela absorção de radio frequência, de onde podemos calcular M_{ef} . Há uma razoável acordo entre este resultado e a estimativa a partir do calor específico [?].

4.11 Bósons

4.11.1 Radiação de corpo negro

Historicamente foi este problema que levou Planck à ideia de quantização da energia. Uma cavidade com radiação eletromagnética pode ser tratada com os

⁵Essencialmente nestas notas, como medimos temperatura em unidades de energia, a constante de Boltzmann $k_B = 1$.

métodos acima. A equação de onda, com condições de contorno periódicas, leva exatamente à mesma contagem de modos normais em termos dos momentos. A relação entre energia e momento é porém diferente do caso não relativístico com massa. Um pequeno orifício permitira o escape de energia. Caso a temperatura não seja muito alta a radiação não terá frequência abaixo do visível e o orifício parecerá negro, justificando que seja chamado de corpo negro. Começamos com o número de estados de uma onda numa cavidade 4.132:

$$\rho(\epsilon)d\epsilon = \mathcal{N}_3 = 2 \frac{L^3 \Delta p_x \Delta p_y \Delta p_z}{(2\pi\hbar)^3}, \quad (4.180)$$

só que agora a relação de dispersão entre energia e momento é diferente e o prefator 2 é devido aos dois estados de polarização da luz ⁶.

$$\epsilon = \hbar\omega, \quad \mathbf{p} = \hbar\mathbf{q}, \quad |\mathbf{p}| = \frac{\hbar\omega}{c} \quad (4.181)$$

$$\rho(\epsilon)d\epsilon = 2 \frac{V 4\pi p^2}{(2\pi\hbar)^3} dp, \quad (4.182)$$

$$\rho(\epsilon)d\epsilon = \frac{8\pi V \omega^2}{(2\pi c)^3} d\omega, \quad (4.183)$$

O número de ocupação médio de um modo com frequência ω é obtido da expressão 4.126, só que com o detalhe que o potencial químico é nulo, pois como o fóton não tem massa, sempre podemos colocar um fóton a mais no extremo infravermelho, que para $V \rightarrow \infty$, tem frequência zero. Segue, o número de ocupação no problema de Planck:

$$\bar{n}_\omega^P = \frac{1}{e^{\beta\hbar\omega} - 1}. \quad (4.184)$$

A energia total será a integral sobre todas as frequências, da energia associada ao estado de frequência $\hbar\omega$, vezes o número de ocupação \bar{n}_ω^P e densidade de estados de frequência ω :

$$\begin{aligned} E &= \int_0^\infty \frac{8\pi V \omega^2}{(2\pi c)^3} \frac{\hbar\omega}{e^{\beta\hbar\omega} - 1} d\omega \\ &= \frac{8\pi V \hbar}{(2\pi c)^3} \int_0^\infty \frac{\omega^3}{e^{\beta\hbar\omega} - 1} d\omega \end{aligned} \quad (4.185)$$

Podemos escrever $E = \int u(\omega, T) d\omega$, definindo a densidade de energia espectral:

$$u(\omega, T) = \frac{8\pi V \hbar}{(2\pi c)^3} \frac{\omega^3}{e^{\beta\hbar\omega} - 1} \quad (4.186)$$

⁶O campo do fóton é um vetor, tem spin 1, mas não são 3 estados porque a massa é nula

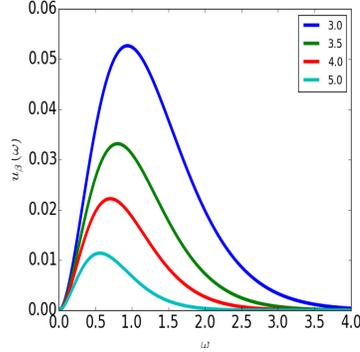


Figura 4.6: Resultado de Planck para a densidade espectral de radiação de corpo negro para diferentes valores de β . Unidades de densidade e β arbitrárias.

que é o famoso resultado de Planck. Mudando variáveis $x = \beta\hbar\omega$ temos

$$\begin{aligned} E &= T^4 \frac{V}{\pi^2(\hbar c)^3} \int_0^\infty \frac{x^3}{e^x - 1} dx \\ \frac{E}{V} &= \sigma_{SB} T^4 \end{aligned} \quad (4.187)$$

onde a última linha expressa o que é chamado de lei de Stephan-Boltzmann e a constante σ_{SB} pode ser calculada usando

$$\int_0^\infty \frac{x^3}{e^x - 1} dx = \frac{\pi^4}{15} \quad (4.188)$$

o que leva, após recolocar a constante de Boltzmann, a

$$\sigma_{SB} = \frac{\pi^2 k_B^4}{15(\hbar c)^3} \quad (4.189)$$

Exercício Mostre que a pressão do gás de fótons é $P = E/(3V)$.

Exercício Mostre que a frequência ω^* em que a função $u(\omega, T)$ atinge um máximo depende de β e que $\omega^*\beta = \text{constante}$, resultado obtido por Wein, experimentalmente e por Planck teoricamente.

4.11.2 Condensação de Bose-Einstein

Voltamos ao caso em que a densidade de estados é dada por 4.136. Estamos interessados em calcular

$$N = \sum_{\mathbf{q}} \bar{n}_{\mathbf{q}} \quad (4.190)$$

e precisamos substituir a soma por uma integral. Isso é válido? Como dissemos antes, nem sempre. Chamemos de N^* o número de partículas que pode ser calculado pela integral sobre os estados de energia com densidade $\rho(\epsilon)$

$$\begin{aligned} \sum_{\mathbf{q}} \bar{n}_{\mathbf{q}} &= \sum_{\mathbf{q}} \frac{1}{e^{\beta(\epsilon_{\mathbf{q}} - \mu)} - 1} \\ \rightarrow N^*(T, \mu) &= \int_0^{\infty} \rho(\epsilon) \frac{1}{e^{\beta(\epsilon - \mu)} - 1} d\epsilon. \end{aligned} \quad (4.191)$$

Podemos considerar duas situações experimentais diferentes ambas com volume V constante. Na primeira, mantemos T fixo e na segunda mantemos N fixo. Para T fixo, quando N^* cresce, a expressão 4.191 mostra que μ também deve crescer. Lembramos que o potencial químico não pode ser positivo (convergência da série geométrica), portanto o valor de μ não pode crescer indefinidamente e eventualmente chega ao seu limite $\mu = 0$. Temos

$$\begin{aligned} N^*(T, 0) &= \int_0^{\infty} \rho(\epsilon) \frac{1}{e^{\beta\epsilon} - 1} d\epsilon. \\ &= CV \int_0^{\infty} \epsilon^{\frac{1}{2}} \frac{1}{e^{\beta\epsilon} - 1} d\epsilon. \\ &= CVT^{\frac{3}{2}} \int_0^{\infty} x^{\frac{1}{2}} \frac{1}{e^x - 1} dx. \end{aligned} \quad (4.192)$$

O que impede de aumentar o número de partículas dentro do volume V , mantendo T fixo além do limite $N(T, 0)$? Nada. Mas devemos perceber que enquanto μ era negativo $N^*(T, \mu)$ calculado pela expressão 4.191 efetivamente calculava o número de partículas total, mas ao atingir a densidade $d_0(T)$

$$d_0(T) = \frac{N^*(T, 0)}{V} = C'T^{\frac{3}{2}} \quad (4.193)$$

com

$$C' = C \int_0^{\infty} x^{\frac{1}{2}} \frac{1}{e^x - 1} dx = C\Gamma\left(\frac{3}{2}\right)\zeta\left(\frac{3}{2}\right) \quad (4.194)$$

a integral não estima mais o número total, mesmo que este continue aumentando. O que acontece é que como o potencial químico é nulo, ao aumentar o número de partículas, estas entrarão no estado fundamental e portanto seu número de ocupação começará a crescer. A segunda condição para substituir a soma pela integral não é satisfeita. O número de ocupação para o estado de $\epsilon = 0$ é

$$N_0 \approx n_0 = \frac{1}{e^{-\beta\mu} - 1} \approx \frac{1}{1 - \beta\mu - 1} = -\frac{T}{\mu}$$

e a divergência quando $\mu \rightarrow 0$ sinaliza problemas.

Ainda usaremos a integral, mas só para os estados excitados. Assim escrevemos

$$N = N_0 + N^* \quad (4.195)$$

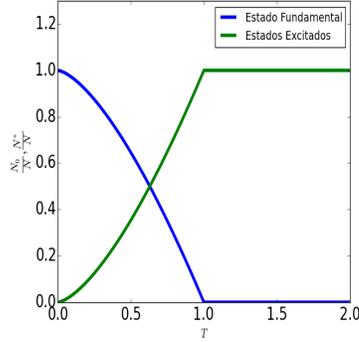


Figura 4.7: Frações dos números de partículas no estado fundamental e nos estados excitados com função de $\frac{T}{T_0}$, para N e V fixos.

onde o número total é a soma do número de partículas no estado fundamental N_0 mais o número nos estados excitados N^* estimado pela integral 4.192. Ao aumentar o número de partículas a temperatura fixa, N^* ficará constante enquanto N_0 aumentará.

Podemos analisar o segundo caso experimental, onde agora a densidade $d = N/V$ é fixa e a temperatura pode mudar. Ao abaixar a temperatura, o potencial químico irá aumentando, como pode ser visto da equação 4.191. A temperatura em que $\mu = 0$ pode é dada por uma relação sutilmente diferente de 4.193

$$d = \frac{N}{V} = C' T_0^{\frac{3}{2}} \quad (4.196)$$

A diferença entre as equações 4.193 e 4.196 é que a primeira descreve quando, para uma temperatura fixa, aumenta-se a densidade até bater em $\mu = 0$, enquanto que a segunda descreve o processo para um sistema de N e V fixos, a temperatura diminui até o potencial químico se tornar zero. Agora a pergunta é o que acontece quando T abaixa ainda além de T_0 ? Para $T < T_0$

$$\begin{aligned} N^* &= C' V T^{\frac{3}{2}} \\ &= C' V T_0^{\frac{3}{2}} \left(\frac{T}{T_0}\right)^{\frac{3}{2}} \\ &= N \left(\frac{T}{T_0}\right)^{\frac{3}{2}} \end{aligned} \quad (4.197)$$

o número de partículas nos estados excitado diminui com a temperatura (elevada a 3/2). Portanto

$$N_0 = N - N^* = N \left(1 - \left(\frac{T}{T_0}\right)^{\frac{3}{2}}\right) \quad (4.198)$$

Exercício Mostre que

$$\int_0^{\infty} x^s \frac{1}{e^x - 1} dx = \Gamma(s)\zeta(s) \quad (4.199)$$

onde $\Gamma(s) = \int_0^{\infty} e^{-t} t^{s-1} dt$ é a função Gama, e $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$ é a função Zeta de Riemann.

Exercício Mostre que $\zeta(3/2) =$.

4.11.3 Calor Específico

O cálculo da energia pode ser feito usando a aproximação da integral mesmo na região da condensação. A temperatura fixa, o aumento de partículas não mudará a energia total, pois entrarão no estado fundamental. Claro que a energia por partícula diminuirá. A densidade fixa, a diminuição da temperatura, influenciará a energia pois N^* mudará. Assim

$$\begin{aligned} E(T, \mu) &= \int_0^{\infty} \rho(\epsilon) \frac{\epsilon}{e^{\beta(\epsilon-\mu)} - 1} d\epsilon, \\ &= CVT^{\frac{5}{2}} \int_0^{\infty} x^{\frac{3}{2}} \frac{1}{e^x - 1} dx. \end{aligned} \quad (4.200)$$

$$\begin{aligned} E(T, \mu) &= CVT^{\frac{5}{2}} \Gamma\left(\frac{3}{2}\right) \zeta\left(\frac{3}{2}\right) \\ &= 0.770TN^* \end{aligned} \quad (4.201)$$

$$C_V = \left. \frac{\partial E}{\partial T} \right|_{V, N} \quad (4.202)$$

$$= 1.9N \left(\frac{T}{T_0}\right)^{\frac{3}{2}} \quad (4.203)$$

$$P = \frac{2}{3} \frac{E}{V} \propto N^* \frac{T}{V} \quad (4.204)$$

Capítulo 5

Aplicações: Modelo de Ising e Campo Médio

Continuando com as aplicações estudaremos variações sobre o tema de variáveis de dois estados, conhecidas como variáveis de Ising. Vários temas importantes serão introduzidos: transições de fase e quebra de simetria, campo médio, matriz de transferência e grupo de renormalização. Não há como tratar isto a não ser de forma introdutória e serve só para provocar os alunos a pensar, mas não devem achar que estes desenvolvimentos são mais do que simplesmente a ponta do iceberg. As técnicas de campo médio encontram aplicação em grande número de problemas. As vezes são as únicas formas analíticas, outras são suficientes para os propósitos desejados e outras se mostram insuficientes. De qualquer forma costuma ser a primeira forma de pensar sobre um problema em Mecânica Estatística.

Neste capítulo faremos uma pequena introdução ao modelo de Ising. Inicialmente proposto por Lenz para tópicos de tese do seu aluno Ising, como modelo para um sistema ferromagnético. O número de extensões possíveis desse modelo é sem dúvida finito, embora do ponto de vista prático, para o estudante que começa a estudá-las possa parecer que não. Faremos uma excursão por métodos de campo médio (CM), primeiro veremos um caso extremamente conhecido da aproximação de CM mas de um ponto de vista geométrico. Depois olharemos para o modelo de alcance infinito e introduziremos o método de integração de ponto de sela. Depois mostraremos que essa aproximação é exata no limite termodinâmico: os termos desprezados vão para zero no limite termodinâmico. Usaremos nas próximas seções, várias vezes a integração por ponto de sela para estudar vidros de spin e redes neurais. O leitor deverá levar em conta que este teorema não garantirá o rigor nas manipulações: o desafio máximo à intuição tanto física quanto matemática será trazido pelo truque de réplicas. O objetivo deste é tratar sistemas onde há pelo menos dois grupos de graus de liberdade e uma grande separação das escalas de tempo em que essas duas classes de variáveis evoluem. Uma delas evolui rapidamente e a outra classe evolui, numa

escala lenta, que por comparação ao tempo de observação experimental pode ser simplesmente levada a infinito. Esta classe permanece, para todos os efeitos, numa configuração desordenada. O objetivo inicial, nesta área, era o estudo de sistemas onde impurezas diluídas, na forma de átomos magnéticos, efetivamente fixos no espaço, num meio não magnético, ocupam posições aleatórias e interagem entre si com interações que dependem da distância e portanto são aleatórias, tanto em intensidade como sinal. A mecânica estatística muda, métodos e intuições devem ser de novo desenvolvidos.

mais blablabla - mudar para um capítulo inicial. daqui::: Os desafios à intuição são os melhores. Aqueles resultados e métodos que não fazem esse desafio, dificilmente merecem uma atenção detalhada. Estes métodos de estudos de sistemas desordenados permitirão e incentivarão à generalização dos modelos estudados, permitindo uma extensão dos problemas da mecânica estatística para áreas fora da física convencional. Em primeiro lugar permitiu o estudo de memorização e generalização em redes neurais artificiais. A desordem tem uma origem totalmente diferente, devida aos padrões ou mensagens que devem ser aprendidos, levando ao estudo de desordem correlacionada. Rapidamente se percebeu que em vários casos o problema estudado era efetivamente o de otimização combinatória o que levou ao estudo de problemas clássicos em ciência da computação. No capítulo ?? estudaremos especificamente o problema de K-sat usando os métodos desenvolvidos neste capítulo.

mais blablabla Finalmente mencionamos, embora no começo e que de forma bastante hesitante, problemas de possível interesse econômico, em ciência cognitiva e ciências sociais tem sido modelados de forma a poder ser estudados com estes métodos. Este estágio incipiente não deve afastar os estudantes. Pelo contrário parecerá atrativo, em especial àqueles que por natureza gostam de trabalhar em áreas não estabelecidas, onde os problemas não se chegaram ainda a questionar se as aproximações concordam com as medidas em menos que frações de 1%. Talvez seja muito ambicioso, mas certamente parece uma conjectura razoável, afirmar que é a partir de este tipo de modelagem, previsões poderão ser feitas e comparadas com medidas em condições, se talvez no começo de natureza muito restritiva, quem sabe, mais naturais no futuro. Este tipo de verificação das previsões contra medidas parece ser o único critério para denominar uma área de ciência. A biologia certamente já está nessa fase, mas a era das previsões quantitativas em, por exemplo, estudo de neurofisiologia de sistemas sensoriais, aparece mais como uma promessa do que uma realidade concreta. até agora...

Essa intuição pode ser levada a quase certeza em vários casos onde simulações confirmam os resultados. O que quer dizer confirmam? Quão satisfatória é essa confirmação, depende do gosto do usuário e de suas convicções epistemológicas. Para o matemático e o físico matemático não são suficientes, enquanto que para o físico teórico podem parecer argumentos contundentes, que permitirão ir frente para atacar outros problemas. Podem ainda não parecer suficientes para o físico experimental, interessado no comportamento de sistemas desordenados magnéticos encontrados no laboratório, com uma geometria bem diferente dos modelos que podem ser atacados com os métodos analíticos aqui estudados. Es-

tes resultados podem parecer curiosos para o cientista de computação que, ao estudar estes problemas do ponto de vista de *pior caso* não está acostumado ao estudo de comportamento *típico*, possibilidade que a mecânica estatística traz. Para o neurocientista os resultados obtidos aplicando as técnicas aprendidas no estudo de vidros de spin não serão interessantes, dada a enorme distância entre os objetos, métodos e linguagem que são do seu interesse. Mas a experiência adquirida pelo físico neste laboratório permitirá ir além e estudar o comportamento de sistemas dinâmicos muito mais próximos de sistemas de interesse biológico. Isto sim permitirá um diálogo maior. O interessante também que a perspectiva do estudo de sistemas desordenados como sistemas de processamento de informação permitirá estudar a caracterização de sistemas biológicos e apontar limites de desempenho, sugerindo o estudo do ponto de vista evolutivo e aprendendo com as falhas desses sistemas biológicos de atingirem os desempenhos ótimos.

5.1 O modelo de Ising

Considere N variáveis s_i que tomam valores 1 ou -1 . A probabilidade a priori é a medida de Bernoulli $p_0(s) = \frac{1}{2}(\delta(1-s) + \delta(1+s))$. As variáveis podem estar localizadas numa rede de Bravais ou em algum outro grafo. A cada configuração é associada uma energia através do Hamiltoniano:

$$\mathcal{H} = - \sum_{\langle ij \rangle} J_{ij} s_i s_j - \sum_i h_i s_i. \quad (5.1)$$

A primeira soma é sobre os pares (i, j) de spins que interagem entre si e a segunda a interação com um campo magnético externo, que pode depender do índice i . Já vimos no capítulo anterior o caso de $J_{ij} = 0$, que é o modelo paramagnético. O modelo é essencialmente definido pelos J_{ij} e os campos h_i . Casos que serão interessantes:

- Modelo ferromagnético de primeiros vizinhos

onde (i, j) são primeiros vizinhos numa rede hipercúbica de D dimensões e $J_{ij} = J > 0$. Para $D = 1$ é trivial, $D = 2$ são conhecidos vários resultados exatamente. Foi o primeiro modelo de rede onde se encontrou rigorosamente uma transição de fase (Peierls). A solução exata, para a energia livre foi encontrada inicialmente por Onsager. Para $D = 3$ se acredita atualmente que não seja exatamente solúvel, mas se conhecem rigorosamente vários resultados (rigorosos mas não exatos). Para $J_{ij} = J < 0$ o modelo é dito antiferromagnético. Estes casos representam possivelmente os modelos mais populares e básicos em Física Estatística. Estamos interessados neles e olharemos também outra classe de modelos relacionados.

- Modelo de alcance infinito

onde J_{ij} é diferente de zero para qualquer par de spins (i, j) . Se $J_{ij} = J/N > 0$ o modelo é exatamente solúvel. Estamos também interessados no caso em que J_{ij} é uma variável aleatória com valores tanto positivos como negativos:

- Vidro de spin

Ha vários modelos de Ising para vidros de spin. Em particular olharemos detalhadamente o model de Sherrington-Kirpatrick. Neste caso, o modelo de alcance infinito tem acoplamentos J_{ij} com distribuição gaussiana

$$P(J_{ij}) = \sqrt{\frac{N}{2\pi J^2}} e^{-\frac{N(J_{ij}-J_0/N)^2}{2J^2}}, \quad (5.2)$$

portanto

$$\begin{aligned} \langle\langle J_{ij} \rangle\rangle_J &= \frac{J_0}{N} \\ \sigma_J^2 &= \langle\langle J_{ij}^2 \rangle\rangle_J - \langle\langle J_{ij} \rangle\rangle_J^2 = \frac{J^2}{N} \end{aligned} \quad (5.3)$$

Note que a dependência com N é diferente do modelo onde todos os acoplamentos são iguais. O problema ainda não está bem posto. Precisamos definir o limite termodinâmico e olhar para os possíveis parâmetros de ordem. Antes disso, para aquecer, olharemos o problema mais simples, de análise de campo médio vista de um ponto de vista geométrico e para um modelo ferromagnético numa rede e depois para o modelo de alcance infinito com $J_{ij} = J/N$ constante.

5.2 Modelo de Ising em uma dimensão

Este é o problema original estudado por Ising e Lenz. Os sítios são indexados por $i = 1, \dots, N$. O modelo é de interação de primeiros vizinhos e invariante por translações. Isso é devido à uniformidade das interações $J_{ij} = J > 0$ se $j = i + 1$ e senão $J_{ij} = 0$, e também $h_i = h$ para todo i .

5.2.1 Matriz de Transferência

Vamos supor um modelo de um ferromagneto nas seguintes condições experimentais: Uma amostra é colocada em um campo magnético externo h .

$$\mathcal{H} = -J \sum_{\langle ij \rangle} s_i s_{i+1} - h \sum_i s_i. \quad (5.4)$$

Após isso podemos escolher a temperatura $T = \beta^{-1}$ e colocamos o sistema em contato com um reservatório a essa temperatura. O problema de atribuir uma probabilidade a cada configuração $\{s_1, s_2, \dots, s_N\}$ é feito, de acordo com o que vimos nos capítulos anteriores maximizando a entropia sujeito a um vínculo de energia total fixa e impondo normalização. O parâmetro de Lagrange que impõe esse vínculo é β . Estamos interessados em calcular a magnetização, primeiro em tamanho finito

$$m(\beta, h, N) = \left\langle \frac{1}{N} \sum_i s_i \right\rangle = \sum_{\{s_1, s_2, \dots, s_N\}} \left(\frac{1}{N} \sum_i s_i \right) P(s_1, s_2, \dots, s_N) \quad (5.5)$$

A soma externa é sobre todas as 2^N configurações possíveis. A soma interna sobre todos os sítios da rede. No limite termodinâmico $N \rightarrow \infty$,

$$m(\beta, h) = \lim_{N \rightarrow \infty} m(\beta, h, N),$$

e finalmente a magnetização espontânea é definida no limite de campo nulo:

$$m(\beta) = \lim_{h \rightarrow 0^+} m(\beta, h). \quad (5.6)$$

O limite $h \rightarrow 0^+$ permite quebrar a simetria da escolha do estado de menor energia livre, caso haja mais de um estado. Veremos o resultado que desapontou Ising, que em 1 dimensão só há um estado. Mas veremos que há dois em duas dimensões ou mais.

Se estivermos interessados em sistemas com N muito grande não deve fazer muita diferença adicionarmos um termo de interação entre o primeiro e o último spin da cadeia, $s_1 = s_{N+1}$, de forma que a podemos olhar para

$$\begin{aligned} \mathcal{H} &= -J \sum_{i=1}^N s_i s_{i+1} - \frac{h}{2} \sum_{i=1}^N (s_i + s_{i+1}) \\ &= - \sum_{i=1}^N \left(J s_i s_{i+1} - \frac{h}{2} (s_i + s_{i+1}) \right) = \sum_{i=1}^N \mathcal{H}_{s_i, s_{i+1}} \end{aligned} \quad (5.7)$$

Sabendo que o valor esperado de \mathcal{H} seria informação de grande valor, se a tivéssemos, procedemos por maximização da entropia, como se soubéssemos o vínculo

$$\langle H \rangle = E$$

e atribuímos a probabilidade

$$P(\{s_i\}) = \frac{e^{-\beta H}}{Z}. \quad (5.8)$$

que é a distribuição de Boltzmann. Como sempre $Z(\beta, h, N) = \sum_{\{s_i\}} \exp(-\beta H)$ é a função de partição. A energia livre por spin

$$f(\beta, h) = - \lim_{N \rightarrow \infty} \frac{1}{N\beta} \log Z(\beta, h, N). \quad (5.9)$$

e a magnetização

$$m(\beta, h) = - \frac{\partial f(\beta, h)}{\partial h} = \lim_{N \rightarrow \infty} \frac{1}{N\beta} \frac{\partial \log Z(\beta, h, N)}{\partial h} \quad (5.10)$$

Com as condições de contorno periódicas

$$Z(\beta, h) = \sum_{\{s_1, s_2, \dots, s_N\}} \prod_i e^{-\beta \mathcal{H}_{s_i, s_{i+1}}} \quad (5.11)$$

A matriz de transferência \mathbf{T} é introduzida definindo seus elementos através de

$$(\mathbf{T})_{s_i, s_{i+1}} = \exp \beta \mathcal{H}_{s_i, s_{i+1}} \quad (5.12)$$

$$\mathbf{T} = \begin{pmatrix} T_{++} & T_{+-} \\ T_{-+} & T_{--} \end{pmatrix}$$

$$\mathbf{T} = \begin{pmatrix} e^{\beta(J+h)} & e^{-\beta J} \\ e^{-\beta J} & e^{\beta(J-h)} \end{pmatrix}$$

Se $\mathbf{B} = \mathbf{A} \cdot \mathbf{A}$ os elementos de \mathbf{B} serão

$$B_{ij} = \sum_k A_{ik} A_{kj}$$

e o traço de \mathbf{B}

$$\text{Tr} \mathbf{B} = \sum_i B_{ii} = \sum_i \sum_k A_{ik} A_{ki}$$

A utilidade disto fica clara ao perceber que a função de partição pode ser escrita como o traço da matriz de transferência elevada à N -ésima potência.

$$\begin{aligned} Z(\beta, h) &= \sum_{\{s_1, s_2, \dots, s_N\}} T_{s_1, s_2} T_{s_2, s_3} \dots T_{s_i, s_{i+1}} \dots T_{s_N, s_1} \\ &= \text{tr}(\mathbf{T}^N) \end{aligned} \quad (5.13)$$

Usamos dois resultados simples de álgebra linear:

(i) o traço de produto de matrizes é invariante ante permutações cíclicas: $\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{A})$. Prove este resultado.

(ii) A matriz de transferência pode ser diagonalizada por uma transformação unitária \mathbf{U} :

$$\mathbf{T}_D = \mathbf{U}^{-1} \mathbf{T} \mathbf{U} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

Prove isto.

Segue que, inserindo $\mathbf{I} = \mathbf{U}\mathbf{U}^{-1}$ entre os fatores \mathbf{T}

$$\begin{aligned} Z(\beta, h, N) &= \text{tr}(\mathbf{T}^N) \\ &= \text{tr}(\mathbf{U}\mathbf{U}^{-1}\mathbf{T}^N) = \text{tr}(\mathbf{U}^{-1}\mathbf{T}^N\mathbf{U}) \\ &= \text{tr}(\mathbf{U}^{-1}\mathbf{T}\mathbf{U}\mathbf{U}^{-1} \dots \mathbf{T}\mathbf{U}\mathbf{U}^{-1}\mathbf{T}\mathbf{U}) \\ &= \text{tr}(\mathbf{T}_D^N) \\ &= \lambda_1^N + \lambda_2^N \end{aligned} \quad (5.14)$$

onde $\lambda_{1,2}$ são os autovalores de \mathbf{T} obtidos das raízes de

$$\det(\mathbf{T} - \lambda \mathbf{I}) = 0,$$

então

$$\lambda^2 - 2\lambda e^{\beta J} \cosh(\beta h) + 2 \sinh(2\beta J) = 0$$

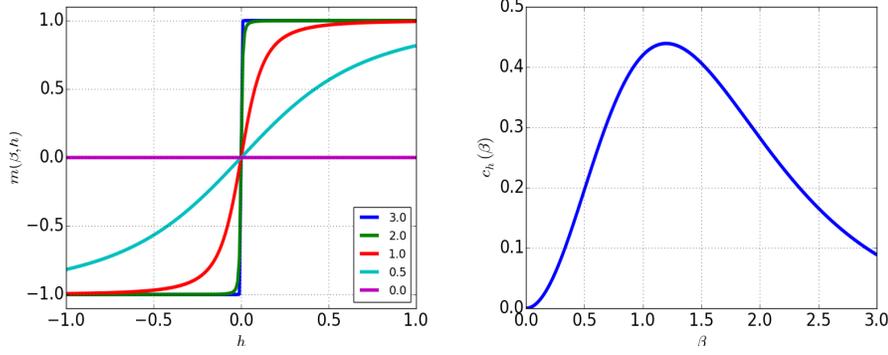


Figura 5.1: Ising 1d. Esquerda: Magnetização $m(\beta, h)$ como função de h para diferentes valores de β . Note que todas as curvas passam pela origem: não há magnetização espontânea. Direita: Calor específico Unidades: $J = 1$).

$$\begin{aligned}\lambda_{1,2} &= e^{\beta J} \cosh(\beta h) \pm \sqrt{e^{2\beta J} \cosh^2(\beta h) - 2 \sinh(2\beta J)} \\ \lambda_{1,2} &= \left(e^{\beta J} \cosh(\beta h) \pm \sqrt{e^{2\beta J} \cosh^2(\beta h) - e^{2\beta J} + e^{-2\beta J}} \right) \\ \lambda_{1,2} &= e^{\beta J} \left(\cosh(\beta h) \pm \sqrt{\sinh^2(\beta h) + e^{-4\beta J}} \right)\end{aligned}\quad (5.15)$$

Exercício Mostre que $\lambda_1 > \lambda_2 > 0$, para qualquer escolha de β e h . Para N grande

$$\begin{aligned}Z(\beta, h, N) &= \lambda_1^N + \lambda_2^N \\ &= \lambda_1^N \left(1 + \left(\frac{\lambda_2}{\lambda_1} \right)^N \right) \\ &\rightarrow \lambda_1^N\end{aligned}\quad (5.16)$$

Segue que a magnetização é dada por

$$\begin{aligned}m(\beta, h) &= \frac{1}{\beta} \frac{\partial \log \lambda_1}{\partial h} \\ &= \frac{\sinh(\beta h) + \frac{\sinh(\beta h) \cosh(\beta h)}{\sqrt{\sinh^2(\beta h) + e^{-4\beta J}}}}{\cosh(\beta h) + \sqrt{\sinh^2(\beta h) + e^{-4\beta J}}} \\ &= \frac{\sinh(\beta h)}{\sqrt{\sinh^2(\beta h) + e^{-4\beta J}}} \xrightarrow{h \rightarrow 0} 0\end{aligned}\quad (5.17)$$

O calor específico pode ser calculado de

$$E(T, N, h) = -\frac{\partial \log Z}{\partial \beta} = -N \frac{\partial \log \lambda_1}{\partial \beta}$$

$$E(T, N, h) = -\frac{N}{\lambda_1} \frac{\partial \lambda_1}{\partial \beta} = -\frac{N}{\lambda_1} \left(J\lambda_1 + e^{\beta J} \left(h \sinh \beta h + \frac{h \sinh \beta h \cosh \beta h - 2e^{-4\beta J}}{\sqrt{\sinh^2(\beta h) + e^{-4\beta J}}} \right) \right)$$

$$E(T, N, h = 0) = E(T, N) = -\frac{N}{\lambda_{1,h=0}} \frac{\partial \lambda_{1,h=0}}{\partial \beta} = -NJ - \frac{Ne^{-\beta J}}{2 \cosh \beta J} \quad (5.18)$$

e, usando $C_h = \frac{\partial E}{\partial T}$ chegamos a

$$c_h(\beta) = \frac{1}{N} C_h(h = 0) = (\beta J)^2 \operatorname{sech}^2(\beta J) \quad (5.19)$$

Note que o pico no calor específico é muito bem comportado e não está associado com uma transição de fase.

5.3 Campo Médio: uma interpretação geométrica

Este método é o começo para métodos mais sofisticados e tem grande aplicabilidade. Os resultados em baixa dimensionalidade perto da transição de fase ferromagnética não são muito precisos. Fora da região crítica, no entanto, apresenta resultados razoáveis e muito simples. Portanto é quase que indicado sempre fazer uma análise investigativa de um modelo usando campo médio. Consideramos um modelo na rede, o leitor pode pensar numa rede cúbica, mas se aplica de forma quase idêntica a outras redes. O desenvolvimento aqui apresentado pode ser encontrado de uma forma ou outra em muitos lugares, aqui seguimos Oppen e Winther.

Sabendo que o valor esperado de H seria informação de grande valor, se a tivéssemos, procedemos por maximização da entropia, como se soubéssemos o vínculo

$$\langle \mathcal{H} \rangle = E$$

e atribuímos a probabilidade

$$P(\{s_i\}) = \frac{e^{-\beta \mathcal{H}}}{Z} \quad (5.20)$$

que é a distribuição de Boltzmann. Como sempre $Z = \sum_{\{s_i\}} \exp(-\beta \mathcal{H})$ é a função de partição. Podemos adicionar um campo magnético externo, ou uma fonte e \mathcal{H} será substituído por $\mathcal{H} = -\sum J s_i s_j - h \sum_i s_i$ e $Z(\beta, h) = \sum_{\{s_i\}} \exp(-\beta \mathcal{H})$. Definimos algum tipo de condição de contorno, por exemplo, fora de uma região Λ de tamanho $|\Lambda| = L^3 = N$ fixamos os spins no estado +1 e olhamos então para a energia livre por spin

$$f(\beta, h) = -\lim_{N \rightarrow \infty} \frac{1}{N\beta} \log Z(\beta, h). \quad (5.21)$$

Dependendo da interpretação do significado de s o modelo pode representar um gás (e.g. na variável $\tau = (s+1)/2$) e portanto a f pode ser chamado de pressão.

Em geral este problema, o cálculo de f não pode ser feito exatamente. Dizemos que o problema é não integrável em Física e que a $P(\{s_i\})$ é não tratável em Ciência de Computação.

Suponha que temos uma classe Q de distribuições $Q(\theta)$ que dependem de parâmetros θ para a qual podemos resolver exatamente o problema equivalente da equação 5.21. Queremos escolher dentre a classe Q o elemento que mais se aproxima da distribuição P , o problema fica simples e as estimativas podem ser comparadas com experimentos ou simulações. O que quer dizer mais próximo? A resposta vem do fato que lidamos com distribuições e entre distribuições infinitesimalmente próximas há uma métrica natural, dada pela matriz de informação de Fisher. Esta é obtida do limite do número de Kullback-Leibler (KL) ou menos a entropia relativa, para $\Delta\theta \rightarrow 0$. Então escolhemos Q pelo mínimo do número de KL

$$KL(Q||P) = \sum_{\{s_i\}} Q \log \frac{Q}{P}. \quad (5.22)$$

Poderíamos usar $KL(P||Q)$, mas aqui isso não resolveria nosso problema de tractabilidade. Usaremos esse número ao olhar para aprendizagem em redes neurais (cap. ??).

A escolha da classe Q determina o tipo de aproximação que será feita.

As opções mais simples consistem em escolher Q na forma de um produto sobre cliques finitos da rede, depois disso ainda tratável, Q pode ser definido sobre estruturas que não tem ciclos fechados, o que inclui árvores de Caley. A solução neste tipo de estrutura, que será generalizada além desse tipo de árvore será abordada ao estudar os métodos de propagação de crenças (*belief propagation*). Aqui simplesmente olharemos para distribuições que fatorizam sobre os sítios individuais da rede: $Q = \prod_i Q_i(s_i)$.

5.3.1 Campo médio: Ising spin $\frac{1}{2}$

O espaço de funções normalizadas de uma variável de dois estados não é muito grande, depende de um parâmetro, portanto $\theta = \{m_1, m_2 \dots m_N\}$ e podem ser escritas como $Q(s_i|m_i) = (1 + s_i m_i)/2$

$$Q = \prod_i Q_i(s_i|m_i) = \prod_i \frac{1 + s_i m_i}{2}. \quad (5.23)$$

O parâmetro m_i tem uma interpretação muito simples: a magnetização ou valor esperado no sítio i , de onde vem a escolha da letra que o denota

$$\langle s_i \rangle = \sum s_i \frac{1 + s_i m_i}{2} = m_i. \quad (5.24)$$

Repetimos: que escolha fazemos para o conjunto dos $\{m_i\}$? Minimizamos a distância de Kullback-Leibler:

$$KL(Q||P) = \log Z + \beta E(Q) - S(Q). \quad (5.25)$$

Dos três termos da equação acima, o primeiro não sabemos calcular, mas isso não é um problema, pois não depende de Q . Os outros dois são o valor esperado da energia na distribuição Q :

$$E(Q) = \sum_{\{s_i\}} Q(s) \mathcal{H}(s) = \langle \mathcal{H}(s) \rangle_Q \quad (5.26)$$

e a entropia de Q :

$$S(Q) = - \sum_{\{s_i\}} Q(s) \log Q(s) = - \langle \log Q \rangle_Q. \quad (5.27)$$

Substituindo Q da equação 5.23 obtém-se

$$E(Q) = - \sum_{\langle ij \rangle} J_{ij} m_i m_j - \sum_i h m_i \quad (5.28)$$

e

$$S(Q) = - \sum_i \left[\frac{1+m_i}{2} \log \frac{1+m_i}{2} + \frac{1-m_i}{2} \log \frac{1-m_i}{2} \right] \quad (5.29)$$

Ao minimizar $KL(Q||P)$ estamos minimizando a energia livre $F(Q) = E(Q) - S(Q)$, note que a temperatura está embutida nos acoplamentos. Isto deve parecer familiar aos físicos. Devido a convexidade de $\log x$ temos que $\log x \leq x-1$, assim colocando $x = P_i/Q_i$ temos

$$\begin{aligned} KL(Q||P) &= - \sum Q_s \log \frac{P_s}{Q_s} \geq \sum Q_s (1 - P_s/Q_s) \\ KL(Q||P) &\geq \sum (Q_s - P_s) = 1 - 1 = 0, \end{aligned} \quad (5.30)$$

preencha os detalhes da demonstração acima. Assim a energia livre exata é limitada por $F(Q)$

$$-\log Z \leq \beta E(Q) - S(Q). \quad (5.31)$$

e vemos de outra maneira que faz sentido encontrar o valor mínimo de F . Esta é a famosa desigualdade de Bogoliubov e aqui temos de um ponto de vista geométrico. A minimização leva a que $\langle s_i \rangle = m_i$ satisfaz

$$m_i = \tanh \left(\beta \left(\sum_j J_{ij} m_j + h_i \right) \right) \quad (5.32)$$

Quando o campo $h_i = h$ e as interações são uniformes, e cada spin interage com ν vizinhos, por exemplo para uma rede quadrada com interações primeiros vizinhos, há $\nu = 4$ vizinhos

$$m = \tanh(\beta(\nu J m + h)) \quad (5.33)$$

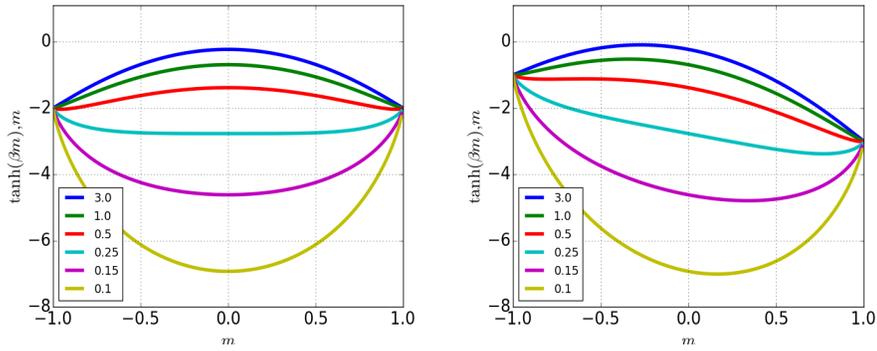


Figura 5.2: Energia livre $f = E - \beta^{-1}S$ na aproximação de Campo Médio para $J_{ij} = J = 1$ e $\nu = 4$. Esquerda: $h = 0$, Direita: $h = 1/2$.

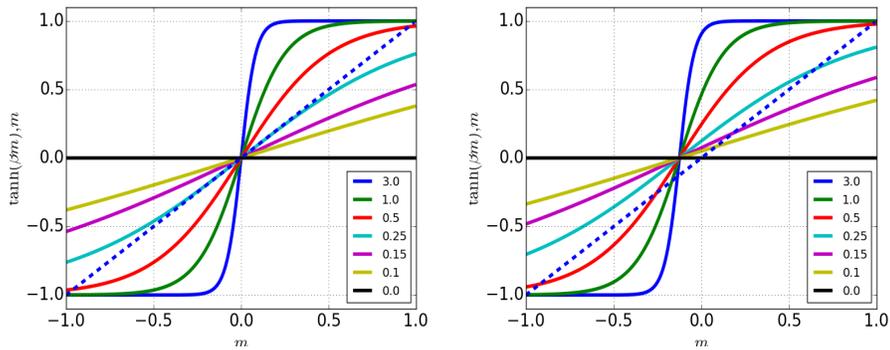


Figura 5.3: Campo Médio para $J_{ij} = J = 1$ e $\nu = 4$. A magnetização $m(\beta, h)$ é determinada pela interseção da tangente hiperbólica (curvas contínuas para diferentes valores de β) e da identidade (curva tracejada), ver equação 5.33. Esquerda: $h = 0$, Direita: $h = 1/2$.

e a energia livre por partícula

$$f = E - \beta^{-1}S = -\frac{\nu J}{2}m^2 - hm + \beta^{-1}\left(\frac{1-m}{2}\log\frac{1-m}{2} + \frac{1+m}{2}\log\frac{1+m}{2}\right) \quad (5.34)$$

Este tipo de equação é chamada de equação de estado de campo médio e remonta em versões de campo médio devidas a Pierre Curie e Pierre-Ernest Weiss no começo do século 20.

Exercício 1

Obtenha a equação 5.32

Exercício 2

Considere a equação de campo médio acima para uma rede de número de coordenação z . Reintroduza a temperatura β^{-1} e faça $J_{ij} = \beta J$ para todos os pares de spin que interagem e o campo h_i independente da posição i . A magnetização deverá também ser independente de i . Portanto $m = \tanh(m/a + h)$, onde $a^{-1} = \beta Jz$. (é útil mostrar que $\tanh^{-1}y = \frac{1}{2}\log\frac{1+y}{1-y}$.)

- Na ausência de um campo externo ($h = 0$) encontre o valor de a crítico, onde a equação de campo médio apresenta uma bifurcação. Acima de a_c , $m^*(\beta, h) = 0$ e abaixo há soluções diferentes de zero. Expanda em série de Taylor, em torno de a_c e descubra a_c e o expoente da magnetização que descreve como m^* vai para zero quando $a \rightarrow a_c$ por valores inferiores.
- Faça o gráfico das soluções $m^*(\beta, h)$ como função da temperatura e encontre a estrutura de bifurcação das soluções.
- Faça o gráfico da energia livre $F(Q) = F(m^*)$ para valores de $a = a_c/2, a_c, 3a_c/2$.

É interessante comparar a equação acima com a chamada equação de Callen, que é exata, e portanto tão insolúvel quanto o problema original:

Exercício 3

Demonstre a equação de Callen:

$$\langle s_i \rangle = \left\langle \tanh \left(\sum_j J_{ij}s_j + h_i \right) \right\rangle, \quad (5.35)$$

seguindo as instruções seguintes para campo nulo. Considere o sistema obtido sem o spin no sítio i : $H_{/i} = H + \sum_j J_{ij} s_i s_j$. Mostre que

$$\begin{aligned} \sum_{s_i} s_i e^{-\beta H} &= 2e^{-\beta H/i} \sinh\left(\sum_j J_{ij}\right) \\ \sum_{s_i} e^{-\beta H} &= 2e^{-\beta H/i} \cosh\left(\sum_j J_{ij}\right) \end{aligned} \quad (5.36)$$

Há uma semelhança óbvia entre a equação de Callen e a aproximação de CM (eq. 5.32). A *única* diferença está na troca entre o valor esperado da \tanh pela \tanh do valor esperado. Isto pode introduzir grandes mudanças nas soluções. Por exemplo, em 1 dimensão, interações de curto alcance, não há transição de fase. O CM prevê sua existência independentemente da dimensão. Para duas e três dimensões, a bifurcação nas soluções, como função da temperatura, tem comportamentos diferentes. As singularidades no comportamento dos valores esperados são diferentes. Soluções exatas ou métodos de aproximação (e.g. Grupo de Renormalização), ou ainda simulações, são necessários para estudar melhor esses fenômenos críticos. Mas há vezes em que o diagrama de fases é muito *maior*¹ que o ponto crítico, e estes métodos aproximados são extremamente úteis para uma visão geral do problema. Em outros casos, porém, as regiões críticas são essencialmente as únicas de interesse, e.g. versões de algumas teorias de campo na rede. Para variáveis não discretas o uso de CM descuidado pode levar a erros grosseiros.

5.3.2 Campo médio: Blume-Capel spin 1

(e.g. M. Blume, Phys. Rev. 141, 517 (1966), H. W. Capel Physica 32 966 (1966)) Neste caso os ions magnéticos modelados tem spin 1 (triplete). O Hamiltoniano é

$$\mathcal{H} = - \sum_{i,j} J_{ij} S_i S_j + D \sum_i S_i^2 - h \sum_i S_i \quad (5.37)$$

onde S_i toma valores $-1, 0, +1$. A interação de *exchange* $J_{ij} > 0$ é de primeiros vizinhos e a geometria da rede é tal que há ν vizinhos, J ainda promove ordem dos estados com $S_i = \pm 1$, mas o termo quadrático favorece $S_i = 0$. Essa competição torna as coisas interessantes. Novamente procuramos a distribuição produto $Q = \prod Q_i$ com Q_i parametrizado pelos valores esperados

$$\begin{aligned} \sum_{S_i=0,\pm 1} Q_i &= 1, \\ \sum_{S_i=0,\pm 1} Q_i S_i &= m_i, \\ \sum_{S_i=0,\pm 1} Q_i S_i^2 &= r_i, \end{aligned} \quad (5.38)$$

¹A região de interesse para um dado problema por ser longe da criticalidade.

que permitem obter $Q_{\pm 1} = (r \pm m)/2$ e $Q_0 = 1 - r$. É óbvio que $r \geq m$. Procuramos maximizar a entropia cruzada ou minimizar a distância de Kullback-Leibler dada pela equação 5.22. A energia livre é $f = E - TS$ com

$$E(Q) = \sum_{\{s_i\}} Q(s) \mathcal{H}(S) = \langle \mathcal{H}(S) \rangle_Q \quad (5.39)$$

e a entropia de Q :

$$S(Q) = - \sum_{\{S_i\}} Q(S) \log Q(S) = - \langle \log Q \rangle_Q . \quad (5.40)$$

Substituindo Q da equação 5.38 obtém-se

$$E(Q) = - \sum_{\langle ij \rangle} J_{ij} m_i m_j + D \sum_i r_i - h \sum_i m_i \quad (5.41)$$

e

$$S(Q) = - \sum_i \left[\frac{r_i + m_i}{2} \log \frac{r_i + m_i}{2} + (1 - r_i) \log(1 - r_i) + \frac{r_i - m_i}{2} \log \frac{r_i - m_i}{2} \right] \quad (5.42)$$

Buscamos soluções espacialmente uniformes, então a energia livre por spin é

$$\frac{F}{N} = f = - \frac{\nu}{2} J m^2 + D r - h m + \beta^{-1} \left[\frac{r + m}{2} \log \frac{r + m}{2} + (1 - r) \log(1 - r) + \frac{r - m}{2} \log \frac{r - m}{2} \right]$$

Agora procuramos as soluções de

$$\frac{\partial f}{\partial m} = 0 \quad (5.43)$$

$$\frac{\partial f}{\partial r} = 0 \quad (5.44)$$

que levam, a primeira a

$$\begin{aligned} \beta(m\nu J + h) &= \frac{1}{2} \log \frac{r + m}{r - m} \\ m &= r \tanh \beta(m\nu J + h) \end{aligned} \quad (5.45)$$

onde usamos novamente que $\tanh^{-1} y = \frac{1}{2} \log \frac{1+y}{1-y}$, e a segunda a

$$e^{-2\beta D} = \frac{r^2 - m^2}{4(1 - r)^2}. \quad (5.46)$$

Esta equação pode ser simplificada substituindo m pela tangente hiperbólica (eq. 5.45), levando a

$$r = \frac{2e^{-\beta D} \cosh(\beta(JM^* + h))}{2e^{-\beta D} \cosh(\beta(JM^* + h)) + 1} \quad (5.47)$$

As duas equações 5.45 e 5.46 podem ser resolvidas numericamente por iteração para cada valor de β, h, D para obter a solução das propriedades termodinâmicas como mostra a figura 5.4. Talvez o mais interessante é que para valores de D maiores que ≈ 1.85 a transição de fase ocorre

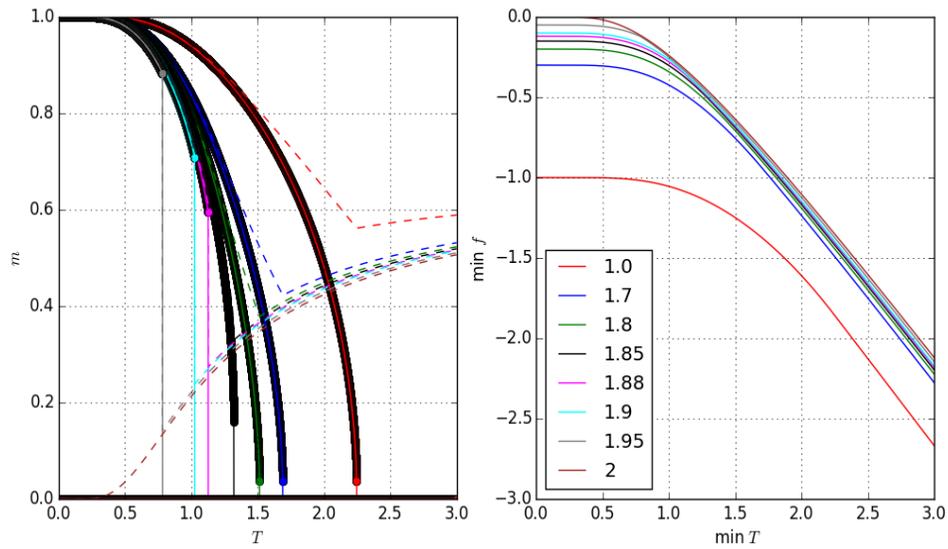


Figura 5.4: Esquerda: m (círculos) e r (tracejada) na aproximação de campo médio para o modelo de Blume Capel com $J_{ij} = J = 1$ e $\nu = 4$, como função de β , para diferentes valores de D . A magnetização $m(\beta, h, D)$ e r são determinados pela solução numérica das equações 5.45 e 5.47 iterando a partir de um par de valores para m e r . É necessário escolher com cuidado o ponto inicial para começar a iteração. A minimização de uma função de uma variável é simples e após encontrar um mínimo razoável o resultado é refinado por iteração das equações de campo médio. Direita: A energia livre obtida em campo médio é uma função com concavidade definida.

5.4 O Modelo de alcance infinito: Ising spin $\frac{1}{2}$

Um modelo que em física é considerado simplesmente um laboratório, em outras áreas pode ser um bom candidato para representar um fenômeno. Este modelo é definido por $J_{ij} = J/N$, para qualquer par $i, j = 1, 2, \dots, N$. Usaremos dois métodos para obter a equação de campo médio, um que usa a aproximação de integração de ponto de sela (PS). O outro, simples e rigoroso, mostrará que a aproximação de PS é no limite $N \rightarrow \infty$ exata. Isso é interessante para atacar outros problemas onde não seja tão fácil a prova rigorosa. No limite termodinâmico o acoplamento vai a zero, mas o número de vizinhos a infinito. A combinação tem propriedades interessantes para essa escolha $J_{ij} = J/N$. Como veremos, neste caso a energia é extensiva e a termodinâmica é não trivial.

5.4.1 Método de Ponto de sela

Este método é simples e informativo. A necessidade de obter um limite termodinâmico aceitável nos permite indicar as quantidades que servem como parâmetro de ordem. Aqui surge a magnetização como a quantidade observável relevante. Quando for aplicado a modelos de spin o método também sugerirá parâmetros de ordem úteis.

Incluimos o acoplamento J em β e o Hamiltoniano é dado por

$$H_N(s, h) = -\frac{1}{N} \sum_{\langle i, j \rangle} s_i s_j - h \sum_i s_i \quad (5.48)$$

Temos $N(N-1)/2$ pares e

$$Z_N(\beta, h) = \sum_{\{s_i\}} e^{-\beta H} \quad (5.49)$$

define a variável

$$m = \frac{1}{N} \sum_i s_i \quad (5.50)$$

em termos da qual podemos escrever o Hamiltoniano

$$H_N(s, h) = -\frac{N}{2} m^2 - hNm \quad (5.51)$$

que obviamente extensivo e a função de partição

$$Z_N(\beta, h) = \sum_{\{s_i\}} e^{\frac{N\beta}{2} m^2 + \beta hNm} \quad (5.52)$$

A relação que será fundamental aqui para obter a solução é o fato simples que

$$1 = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx \quad (5.53)$$

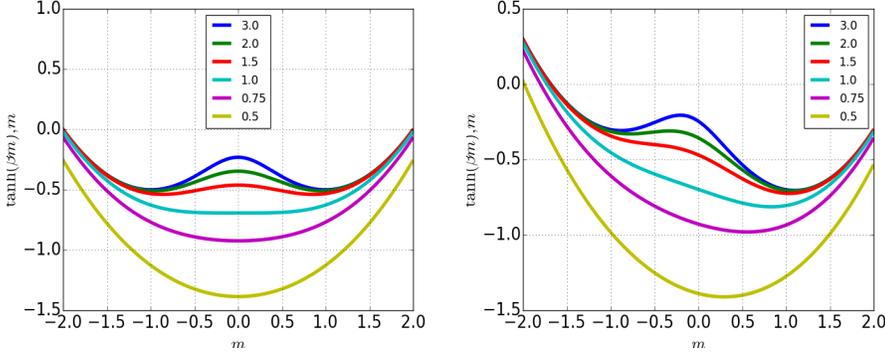


Figura 5.5: f para modelo de alcance infinito, ver equação 5.59. Esquerda: $h = 0$, Direita: $h = .1$.

Podemos mudar variáveis: $x = (M - m)\sqrt{\beta N}$ e assim obter

$$e^{\frac{1}{2}\beta N m^2} = \frac{1}{\sqrt{2\pi/(\beta N)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\beta N M^2 + \beta N m M} dM \quad (5.54)$$

Este truque, descoberto por vários autores ², transforma um termo quadrático em m no expoente, do lado esquerdo, em um linear, do lado direito. O significado é que os spins ficam desacoplados: a exponencial da soma é um produto, portanto vemos que a distribuição é um produto sobre sítios. Isto leva a resultados análogos ao campo médio na seção anterior onde Q era um produto sobre sítios. Por isso é dito que este é um modelo de campo médio. A solução aqui é exata. Se o leitor quiser pode identificar a aproximação na construção do Hamiltoniano, ou pode simplesmente considerá-lo outro modelo.

O preço desta enorme simplificação é pequeno: uma única integral sobre M . A função de partição fica

$$\begin{aligned} Z_N(\beta, h) &= \frac{1}{\sqrt{2\pi/(\beta N)}} \sum_{\{s_i\}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\beta N M^2 + \beta N m(M+h)} dM \\ &= \frac{1}{\sqrt{2\pi/(\beta N)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\beta N M^2} \left(\sum_{s_i} e^{\beta(M+h)s_i} \right)^N dM \\ &= \frac{1}{\sqrt{2\pi/(\beta N)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\beta N M^2} (2 \cosh(\beta(M+h)))^N dM \\ &= \frac{1}{\sqrt{2\pi/(\beta N)}} \int_{-\infty}^{\infty} e^{-\beta N f(M,h)} dM \end{aligned} \quad (5.55)$$

²Acho que o primeiro foi Feynman, mas Hubbard e Stratonovich também costumam ser citados, embora Gauss já devia conhecê-lo.

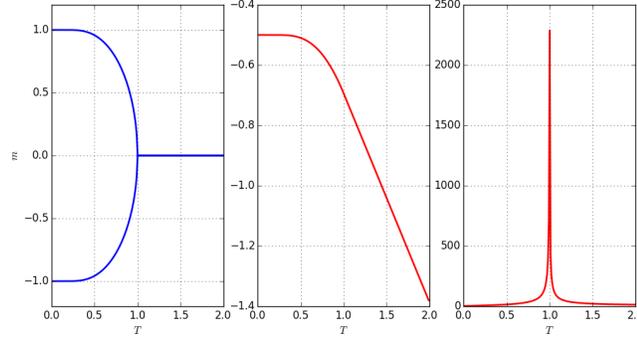


Figura 5.6: Modelo Ising de spin $\frac{1}{2}$ de alcance infinito. Esquerda: magnetização contra T . Centro: $\min\text{-}f$ contra T . Direita: numero de iterações de $m_{t+1} = \tanh \beta m_t$ até que $|m_{t+1} - m_t| < 10^{-6}$. A convergência da dinâmica da iterações fica muito mais lenta no ponto crítico.

onde

$$f(M, h) = \frac{1}{2}M^2 - \frac{1}{\beta} \log(2 \cosh(\beta(M + h))) \quad (5.56)$$

Talvez o preço seja grande se não for possível efetuar a integração. A solução é possível devido ao pré-fator N no expoente da equação 5.55. Estamos interessados em $N \rightarrow \infty$ e a contribuição à integral, à medida que N aumenta será cada vez mais concentrada em torno do máximo de $f(M)$, que chamaremos m^* . Se esse máximo for quadrático, nessa região ³ o integrando fica cada vez mais parecido a uma gaussiana e fica fácil integrar. O interessante é a interpretação e comportamento de m^* . Primeiro a integração. Para determinar m^* olhamos para $\partial f(M)/\partial M = 0$, o que leva à equação de CM

$$m^* = \tanh(\beta(m^* + h)), \quad (5.57)$$

expandindo em série de Taylor em torno do máximo:

$$f(M, h) = f(m^*, h) + \frac{1}{2} \left(\frac{\partial^2 f(M, h)}{\partial M^2} \right)_{|m^*} (M - m^*)^2 + \dots \quad (5.58)$$

onde

$$f(m^*, h) = \frac{1}{2}m^{*2} - \frac{1}{\beta} \log(2 \cosh(\beta(m^* + h))). \quad (5.59)$$

Verifique que a segunda derivada tem o sinal que deve ter. Este método de integração é devido a Laplace, e sua extensão para variáveis complexas que será necessário posteriormente, recebe o nome de ponto de sela. Aqui m^* é simplesmente um máximo do integrando.

³central..novamente, soma de N variáveis s_i , surpresa?

Exercício

Estime o valor relativo entre o termo quadrático e o próximo. Em que condições é razoável desprezá-lo?

A função de partição fica então

$$Z(\beta, h, N) = e^{-N\beta f(m^*)} \frac{1}{\sqrt{2\pi/(\beta N)}} \int_{-\infty}^{\infty} e^{-\frac{\beta N}{2} \frac{\partial^2 f(M, h)}{\partial M^2} |_{m^*} (M - m^*)^2} dM \quad (5.60)$$

A interpretação de m^* é óbvia, novamente é o valor esperado de $\langle \sum_i s_i / N \rangle$ que, pela homogeneidade do sistema, é igual ao valor esperado $\langle s_i \rangle$. Mostramos isso a seguir. Primeiro é fácil ver que

$$\frac{\partial \log Z}{\partial h} = \beta \langle \sum_{i=1}^N s_i \rangle = N\beta m^* \quad (5.61)$$

onde M^* é a magnetização por partícula, ou o valor esperado $\langle s_i \rangle$. Agora tomamos a derivada do log da equação 5.60 vendo que a integral é um fator trivial que não contribui, para obter que

$$M^* = \tanh(\beta m^* + h), \quad (5.62)$$

comparando com a equação 5.57 finalmente obtemos que

$$M^* = m^*, \quad (5.63)$$

ou seja o valor de M que torna $f(M, h)$ um extremo é a magnetização do sistema.

5.4.2 Cálculo rigoroso para o Modelo de Ising com interações de alcance infinito

Qual é a justificativa de jogar fora as derivadas superiores? Podemos controlar a expansão e mostrar que o erro a que isso leva diminui com N e vai a zero no limite. O método que vamos usar é totalmente diferente e extremamente elegante. Seguimos de perto a apresentação de Guerra, que é ponto de partida para os grandes avanços rigorosos recentes dele e de Talagrand. Mostraremos duas coisas, uma de interesse teórico futuro sobre subaditividade da energia livre e a segunda que a aproximação de ponto de sela é exata no limite termodinâmico. Precisamos antes definir convexidade. Sejam $x_1 < x_2$ e $x = px_1 + (1-p)x_2$, $0 \leq p \leq 1$, dizemos que x é uma combinação convexa de x_1 e x_2 . Uma função $f : x \rightarrow f(x)$ é convexa se

$$f(px_1 + (1-p)x_2) \geq pf(x_1) + (1-p)f(x_2). \quad (5.64)$$

Partimos o sistema de N spins em dois grupos $N = N_1 + N_2$ e $m_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} s_i$ e $m_2 = \frac{1}{N_2} \sum_{i=N_1+1}^N s_i$. Obviamente temos

$$\begin{aligned} Nm &= N_1 m_1 + N_2 m_2 \\ m &= pm_1 + (1-p)m_2 \end{aligned} \quad (5.65)$$

de onde vemos que m é uma combinação convexa de m_1 e m_2 e $p = N_1/N$. Dado que a função $m \rightarrow -m^2$ é convexa, para qualquer configuração de spins, temos que

$$m^2 \leq \frac{N_1}{N} m_1^2 + \frac{N_2}{N} m_2^2 \quad (5.66)$$

segue que

$$Z_N(\beta, h) \leq Z_{N_1}(\beta, h) Z_{N_2}(\beta, h) \quad (5.67)$$

e a subaditividade:

$$\frac{1}{N} \log Z_N(\beta, h) \leq \frac{1}{N} \log Z_{N_1}(\beta, h) + \frac{1}{N} \log Z_{N_2}(\beta, h). \quad (5.68)$$

É claro que podemos separar os grupos em subgrupos, até chegar a N grupos de um só spin

$$\frac{1}{N} \log Z_N(\beta, h) \leq \frac{1}{N} \log Z_1(\beta, h) = \log\left(\frac{1}{2}\beta + 2 \cosh(\beta h)\right) \quad (5.69)$$

Temos uma cota uniforme em N e podemos escrever

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log Z_N(\beta, h) \leq \log\left(\frac{1}{2}\beta + 2 \cosh(\beta h)\right) \quad (5.70)$$

uma cota superior para a energia livre no limite termodinâmico.

Para mostrar que o PS é exato começamos pela definição da equação 5.50

$$m = \frac{1}{N} \sum_i s_i$$

e, para qualquer M

$$(M - m)^2 \geq 0 \rightarrow m^2 \geq 2mM - M^2 \quad (5.71)$$

portanto

$$Z_N(\beta, h) \geq \sum_{s_1, s_2, \dots, s_N} e^{(2mM - M^2)\beta N/2 + \beta h N m}. \quad (5.72)$$

Note que essencialmente este é o truque que lineariza o expoente com respeito a m . O preço aqui não é uma integração mas sim uma desigualdade. Isto no levará a buscar uma outra desigualdade no sentido oposto, mas isso após continuar um pouco mais:

$$\begin{aligned} Z_N(\beta, h) &\geq e^{-\frac{M^2 \beta N}{2}} \sum_{s_1, s_2, \dots, s_N} e^{\beta(h+M) \sum_i s_i} \\ &\geq e^{-\frac{M^2 \beta N}{2}} \left(\sum_{s_i} e^{\beta(h+M) s_i} \right)^N \\ &\geq e^{-\frac{M^2 \beta N}{2}} (2 \cosh(\beta(h+M)))^N \end{aligned} \quad (5.73)$$

e

$$\frac{1}{N} \log Z_N(\beta, h) \geq -\frac{M^2\beta}{2} + \log(2 \cosh(\beta(h + M))) \quad (5.74)$$

O lado direito é independente de N e temos uma cota inferior uniforme. Lembremos que o valor de M é arbitrário e podemos escolhê-lo da melhor forma possível, tomamos o supremo. Agora fica facil tomar o limite termodinâmico:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log Z_N(\beta, h) \geq \sup_M \left[-\frac{M^2\beta}{2} + \log(2 \cosh(\beta(h + M))) \right] \quad (5.75)$$

Procuramos agora uma cota inferior. Notamos que m pode ter um de $N + 1$ valores, $-N \leq m \leq N$ com passos $\Delta m = 2/N$. Introduzimos um 1 na função de partição: $1 = \sum_{M'=-N, \dots, N} \delta_{M'm'}$, onde $m' = Nm$ e $M' = NM$, e onde este M é por agora uma nova variável, que veremos será identificada com o M usado anteriormente. Escreveremos a função δ de Kroenecker como δ_{Mm} sem as linhas:

$$Z_N(\beta, h) = \sum_M \sum_{\{s_i\}} \delta_{Mm} e^{\frac{N\beta}{2}m^2 + \beta h Nm} \quad (5.76)$$

Agora usamos a profunda relação $m^2 = 2mM - M^2$ para $M = m$ obtendo

$$Z_N(\beta, h) = \sum_M \sum_{\{s_i\}} \delta_{Mm} e^{\frac{N\beta}{2}(2mM - M^2) + \beta h Nm} \quad (5.77)$$

Novamente linearizamos o expoente com respeito a m . Usamos outro fato trivial: $\delta_{mM} \leq 1$, logo

$$\begin{aligned} Z_N(\beta, h) &\leq \sum_M e^{-\frac{N\beta}{2}M^2} \sum_{s_i} e^{\beta(M+h)Nm} \\ Z_N(\beta, h) &\leq \sum_M e^{-\frac{N\beta}{2}M^2 + N \log(2 \cosh(\beta(M+h)))} \end{aligned} \quad (5.78)$$

Há $N + 1$ termos na soma acima, podemos substituir M pelo argumento do supremo M^* e obter

$$Z_N(\beta, h) \leq (N + 1) e^{N(-\frac{\beta}{2}M^2 + \log(2 \cosh(\beta(M+h))))} \quad (5.79)$$

$$\frac{1}{N} \log Z_N(\beta, h) \leq \frac{1}{N} \log(N + 1) + \sup_M \left(-\frac{\beta}{2}M^2 + \log(2 \cosh(\beta(M + h))) \right) \quad (5.80)$$

Tomamos o limite $N \rightarrow \infty$ e temos uma cota superior. Juntando a equação 5.74 temos o

Teorema

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log Z_N(\beta, h) = \sup_M \left[-\frac{M^2\beta}{2} + \log(2 \cosh(\beta(h + M))) \right] \quad (5.81)$$

mostrando que no limite termodinâmico a aproximação de ponto de sela é exata.

5.4.3 Método de ponto de sela para o modelo de Blume Capel de spin 1

Para o modelo de spin 1, com Hamiltoniano dado pela equação 5.37, e $J_{ij} = J/N$, podemos repetir o que nos levou à equação 5.55

$$\begin{aligned}
Z_N(\beta, h) &= \frac{1}{\sqrt{2\pi/(\beta N)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\beta N J M^2} \left(\sum_{S_i} e^{\beta(JM+h)s_i - \beta D S_i^2} \right)^N dM \\
&= \frac{1}{\sqrt{2\pi/(\beta N)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\beta N J M^2} (2e^{-\beta D} \cosh(\beta(M+h)) + 1)^N dM \\
&= \frac{1}{\sqrt{2\pi/(\beta N)}} \int_{-\infty}^{\infty} e^{-\beta N g(M, h, D)} dM \\
&\propto e^{-\beta N g(M^*, h, D)}
\end{aligned} \tag{5.82}$$

onde

$$g(M, h, D) = \frac{1}{2}\beta N J M^2 - \frac{1}{\beta} \log (2e^{-\beta D} \cosh(\beta(JM+h)) + 1) \tag{5.83}$$

e M^* o valor que a minimiza. É interessante notar que novamente a interpretação do ponto de mínimo é obtido

$$\frac{\partial \log Z}{\partial h} = N\beta M^*$$

e ainda podemos definir r_∞ por

$$\frac{\partial \log Z}{\partial D} = \beta \langle \sum_{i=1}^N S_i^2 \rangle = -N\beta r_\infty \tag{5.84}$$

que neste caso leva a

$$r_\infty = \frac{\partial g}{\partial D} = \frac{2e^{-\beta D} \cosh(\beta(JM^* + h))}{2e^{-\beta D} \cosh(\beta(JM^* + h)) + 1} \tag{5.85}$$

e para a magnetização, usando a expressão para r_∞ a

$$M^* = r_\infty \tanh(\beta(JM^* + h)), \tag{5.86}$$

Mostrando que as equações 5.86 e 5.46 tem exatamente a mesma forma, e portanto as mesmas soluções que para a aproximação de campo médio (equações 5.45 e 5.85).

Mostre que também vale o

Teorema

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log Z_N(\beta, h) = \sup_M \left[-\frac{M^2 \beta}{2} + \log(2e^{-\beta D} \cosh(\beta(M+h)) + 1) \right] \quad (5.87)$$

mostrando novamente que no limite termodinâmico a aproximação de ponto de sela é exata. Volte na prova do teorema para o caso de spin $\frac{1}{2}$ e generalize para o caso com spin 1.

5.5 Modelo de Ising 1d revisitado: O Grupo de Renormalização

Vamos resolver o modelo de Ising 1d novamente, mas agora de uma forma muito mais interessante embora isto talvez não fique aparente a primeira vista. Suponha que estejamos interessados em descrever um sistema e os graus de liberdade, em lugar de ser os momentos magnéticos de uma partícula (fundamental?) são as magnetizações em regiões que podem envolver várias partículas. Podemos descrever o sistema em diferentes escalas de resolução? Em geral as contas envolvidas em encontrar estas descrições não são simples e por isso olhamos para o modelo de Ising em 1 dimensão com interações de primeiros vizinhos porque as contas podem ser feitas de forma exata e isto é muito educativo. Vimos que da função de partição podemos obter qualquer propriedade termodinâmica do sistema, mas a função de partição é uma soma ou integral sobre as configurações dos graus de liberdade. Uma forma de pensar o Grupo de Renormalização é que é uma forma de organizar a realização destas integrais de forma sistemática e iterada. A cada passo os detalhes dos graus de liberdade na escala mais fina são integrados. A medida que estes forem sendo integrados a forma do que resta pode ser descrita como a exponencial de uma função dos graus de liberdade ainda não integrados, chamada de Hamiltoniano renormalizado ou efetivo. Se um Hamiltoniano é caracterizado por constantes de acoplamento entre os graus de liberdade ou com campos externos, o grupo de renormalização pode ser visto como um fluxo no espaço de Hamiltonianos. Em geral este é um espaço de dimensão muito grande (infinita no limite termodinâmico), mas não no Ising 1d. Aqui bastarão duas dimensões, a interação entre primeiros vizinhos e o campo externo efetivos. Novamente a interação entre os graus de liberdade será descrita por constantes do tipo J e h , ou em particular $\kappa = \beta J$ e $g = \beta h$, mas colocaremos um índice para deixar claro em que resolução estamos descrevendo o sistema: κ_m e g_m com m variando de 2 a M . Para ilustrar isso voltamos ao Hamiltoniano com condições de contorno periódicas, onde agora $N = 2^M$ e escrevemos de forma equivalente, mas dando um destaque especial aos sítios com índice par:

$$\beta \mathcal{H}_M = -\kappa_M \sum_{i=1}^{2^{M-1}} (s_{2i-1} + s_{2i+1}) s_{2i} - \frac{g_M}{2} \sum_{i=1}^{2^{M-1}} (s_{2i-1} + 2s_{2i} + s_{2i+1})$$

O fator 2 nos termos de interação dos sítios pares com o campo leva em conta que estes só aparecem uma vez enquanto que os sítios ímpares, duas vezes. Ao reorganizar como faremos as somas sobre configurações, a função de partição toma a forma

$$Z(\kappa_M, g_M) = \sum_{\{s_1, s_3, \dots, s_{N-1}\}} \sum_{s_2, \dots, s_{2i}, s_{2M}} e^{-\mathcal{H}_i^M} = \sum_{\{s_1, s_3, \dots, s_{N-1}\}} \prod_{i=1}^{2^{M-1}} \sum_{s_{2i}} e^{-\mathcal{H}_i^M} \quad (5.88)$$

onde, para qualquer valor de m definimos

$$\mathcal{H}_i^m = -\kappa_m(s_{2i-1} + s_{2i+1})s_{2i} - \frac{g_m}{2}(s_{2i-1} + 2s_{2i} + s_{2i+1}) \quad (5.89)$$

lembrando que os valores iniciais, na resolução mais fina, são $\kappa_M = \beta J$ e $g_M = \beta h$. Mostraremos que

$$Z(\kappa_M, g_M) \propto \sum_{\{s_1, s_3, \dots, s_{N-1}\}} e^{-\mathcal{H}_i^{M-1}} = Z(\kappa_{M-1}, g_{M-1}) \quad (5.90)$$

ou seja, o sistema pode ser descrito em uma resolução mais baixa, mas com novos valores das constantes de acoplamento κ_{M-1} e g_{M-1} . A constante de proporcionalidade que aparece a cada redução de resolução e que não escrevemos acima, é importante, pois no limite em que todos os spins são eliminados só sobram suas contribuições, que trazem informação sobre a energia livre.

A soma sobre os sítios pares leva a quatro termos, dependendo da vizinhança em que s_{2i} se encontra. Para cada uma dessas possibilidades temos

$$\begin{aligned} \sum_{s_{2i}} e^{-\mathcal{H}_i^M} &= e^{\frac{g_M}{2}(s_{2i-1}+s_{2i+1})} (e^{\kappa_M(s_{2i-1}+s_{2i+1})+g_M} + e^{-\kappa_M(s_{2i-1}+s_{2i+1})-g_M}), \\ &= e^{g_M} (e^{2\kappa_M+g_M} + e^{-2\kappa_M-g_M}) \quad \text{para } s_{2i-1} = s_{2i+1} = 1, \\ &= e^{g_M} + e^{-g_M} \quad \text{para } s_{2i-1} = 1, s_{2i+1} = -1, \\ &= e^{g_M} + e^{-g_M} \quad \text{para } s_{2i-1} = -1, s_{2i+1} = 1, \\ &= e^{-g_M} (e^{-2\kappa_M+g_M} + e^{2\kappa_M-g_M}) \quad \text{para } s_{2i-1} = s_{2i+1} = -1. \end{aligned}$$

Queremos escrever esta soma como um fator de Boltzmann para um Hamiltoniano do tipo dado na equação 5.89, e precisamos calcular as novas constantes κ_{M-1} e h_{M-1}

$$\sum_{s_{2i}} e^{-\mathcal{H}_i^M} = e^{\delta f_M} e^{-\mathcal{H}_i^{M-1}} = e^{\delta f_M} e^{\kappa_{M-1}s_{2i-1}s_{2i+1}+g_{M-1}(s_{2i-1}+s_{2i+1})} \quad (5.91)$$

onde δf_M é uma constante independente de s_{2i-1} e s_{2i+1} . Precisamos impor

$$\begin{aligned} e^{\delta f_M} e^{\kappa_{M-1}+2g_{M-1}} &= e^{2\kappa_M+2g_M} + e^{-2\kappa_M} \quad \text{para } s_{2i-1} = s_{2i+1} = 1, \\ e^{\delta f_M} e^{-\kappa_{M-1}} &= e^{g_M} + e^{-g_M} \quad \text{para } s_{2i-1} = 1, s_{2i+1} = -1, \\ e^{\delta f_M} e^{-\kappa_{M-1}} &= e^{g_M} + e^{-g_M} \quad \text{para } s_{2i-1} = -1, s_{2i+1} = 1, \\ e^{\delta f_M} e^{\kappa_{M-1}-2g_{M-1}} &= e^{2\kappa_M-2g_M} + e^{-2\kappa_M} \quad \text{para } s_{2i-1} = s_{2i+1} = -1, \end{aligned}$$

5.5. MODELO DE ISING 1D REVISITADO: O GRUPO DE RENORMALIZAÇÃO 145

a razão entre a primeira e a quarta diz como muda o campo externo efetivo $g_M \rightarrow g_{M-1}$

$$e^{4g_{M-1}} = \frac{e^{2\kappa_M+2g_M} + e^{-2\kappa_M}}{e^{2\kappa_M-2g_M} + e^{-2\kappa_M}} \quad (5.92)$$

o produto

$$\begin{aligned} e^{2\kappa_{M-1}} &= e^{-2\delta f_M} (e^{2\kappa_M+2g_M} + e^{-2\kappa_M})(e^{2\kappa_M-2g_M} + e^{-2\kappa_M}) \\ &= 2e^{-2\delta f_M} (\cosh 4\kappa_M + \cosh 2g_M). \end{aligned} \quad (5.93)$$

Da segunda e da terceira obtemos

$$e^{\delta f_M} = 2 \cosh g_M e^{\kappa_{M-1}} \quad (5.94)$$

Eliminando δf escrevemos o mapa que leva da descrição em uma escala m para uma escala $m-1$, chamado de Grupo de Renormalização:

$$\kappa_{m-1} = \frac{1}{4} \log \left(\frac{\cosh 4\kappa_m + \cosh 2g_m}{4 \cosh^2 g_m} \right) \quad (5.95)$$

$$g_{m-1} = \frac{1}{4} \log \left(\frac{e^{2\kappa_m+2g_m} + e^{-2\kappa_m}}{e^{2\kappa_m-2g_m} + e^{-2\kappa_m}} \right) \quad (5.96)$$

onde escrevemos a transformação de escala de m para $m-1$, pois uma vez que o Hamiltoniano na escala $M-1$ tem a mesma forma que na escala M , podemos iterar a transformação. A evolução do mapa pode ser visto na figura 5.7. Começando de qualquer lugar no espaço (κ, g) as flechas indicam a direção de mudança sob a renormalização. O importante nessa figura é que não importam as condições iniciais, a evolução leva eventualmente a $(\kappa = 0, g = 0)$, ou seja, o único ponto fixo do mapa de GR descreve um sistema totalmente desacoplado.

Isso é consistente com o resultado usando a matriz de transferência, onde se mostrou que o sistema não tem magnetização espontânea. Agora rotulamos os spins que não foram integrados ainda (com índice ímpar), definindo o novo rótulo $i_{\text{novo}} = \frac{i_{\text{velho}}+1}{2}$. Após uma renormalização estamos no mesmo ponto, só que com a metade dos spins. Falta olhar para a contribuição de δf para a função de partição. A cada renormalização há uma contribuição δf_m para cada spin integrado:

$$Z(\kappa_m, g_m) = e^{2^{m-1} \delta f_m} Z(\kappa_{M-1}, g_{M-1}) \quad (5.97)$$

Podemos calcular a energia livre somando as contribuições de δf_m para todas as escalas m . Mas isso não é muito interessante nesta altura.

O resultado acima para o fluxo por um lado é razoável, mas por outro desaponta quem esperava mais. Isso não é culpa do GR mas do modelo em 1 dimensão. Passar para duas ou mais dimensões traz a dificuldade que o espaço em que o Hamiltoniano efetivo é representado não é mais bidimensional, mas começam a aparecer termos com constantes de acoplamento para termos com 3, 4 ou mais spins, dependendo de como se reorganiza a integração parcial. Diferentes aproximações tem sido estudadas. Uma das mais simples é obtida usando

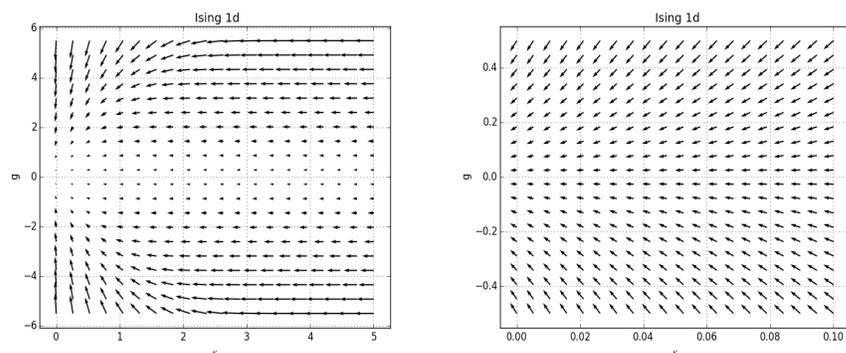


Figura 5.7: Ising 1d. Fluxo no espaço de Hamiltonianos. o lado direito é uma ampliação da figura à esquerda para valores de campo pequeno. A origem ($\kappa = 0, g = 0$) é um ponto fixo do fluxo do grupo de renormalização.

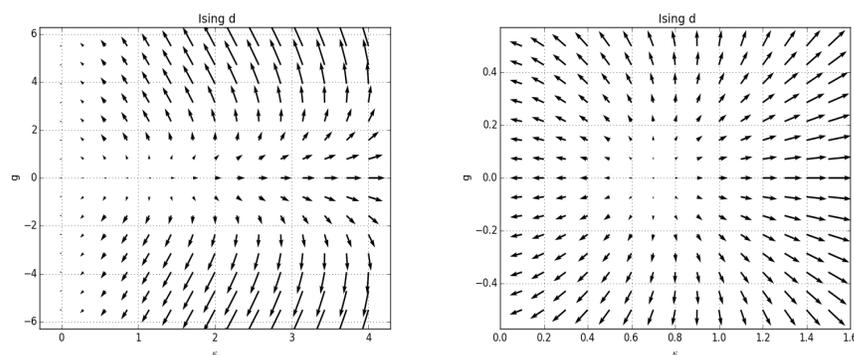


Figura 5.8: Ising 2d. Fluxo no espaço de Hamiltonianos. o lado direito é uma ampliação da figura à esquerda para valores de campo pequeno. A origem ($\kappa = 0, g = 0$) e ($\kappa = \infty, g = 0$) são pontos fixos estáveis do fluxo do grupo de renormalização. Mas há um novo ponto fixo ($\kappa = \kappa_c, g = 0$) que é instável.

o equivalente das equações 5.98 e 5.96 obtidas para 1d, substituindo o numero de vizinhos 2 por 4. Isto é trabalhoso de justificar e não é exato, mas serve para dar uma idéia da estrutura de um fluxo não trivial. O fluxo de GR aparece na figura 5.8 e topologicamente está de acordo com cálculos mais trabalhosos. As equações são

$$\kappa_{m-1} = \frac{1}{4} \log \left(\frac{\cosh 8\kappa_m + \cosh 4g_m}{4 \cosh^2 2g_m} \right) \quad (5.98)$$

$$g_{m-1} = \frac{1}{4} \log \left(\frac{e^{4\kappa_m + 4g_m} + e^{-4\kappa_m}}{e^{4\kappa_m - 4g_m} + e^{-4\kappa_m}} \right) \quad (5.99)$$

e o aluno não deve tentar deduzi-las, pois nesta forma foram simplesmente obtidas aumentando os coeficientes para fazer a mímica de como seria se houvesse um número maior de vizinhos. Mas ao fazer as contas o aluno perceberá que a renormalização leva a outras interações além do campo magnético e primeiros vizinhos. A proliferação de novas interações pode ser tratada ao preço de uma complexidade que não cabe neste ponto. O novo ingrediente é a aparição de dois novos pontos fixos na linha $g = 0$. Agora são três. O antigo na origem continua presente e descreve um sistema desacoplado, para chegar a ele é necessário partir de $g = 0$ e $\kappa < \kappa_c$ um valor crítico (~ 0.7 na figura 5.8). Para $\kappa > \kappa_c$ o fluxo é $\kappa \rightarrow \infty$. Para ficar no ponto crítico precisamos ajustar muito cuidadosamente o valor de κ . A interpretação disto é bem interessante. Vamos discutir um sistema com campo externo nulo: $g = 0$. Para acoplamentos κ pequenos a interação entre os spins é tão pequena que não há ordem, a não ser em regiões pequenas de tamanho característico que chamaremos ξ . Uma vez que a descrição do sistema é feita numa resolução maior que ξ , não há mais ilhas ordenadas e o sistema é descrito por um Hamiltoniano renormalizado com $\kappa \sim 0$. Para $\kappa > \kappa_c$, as ilhas de desordem tipicamente tem tamanho de ordem ξ . Quando a resolução so distingue regiões maiores que ξ , todo o sistema aparece ordenado, como se o valor de κ fosse muito alto. O interessante é quando o sistema é sintonizado exatamente no ponto $\kappa = \kappa_c, g = 0$. As ilhas de ordem e desordem são encontradas em todas as escalas, independentemente da escala de resolução com que analisamos o sistema. A análise do ponto crítico e da simetria não trivial por transformações de escala levou ao estudo de Grupo de Renormalização pois métodos como Campo Médio o séries de potências baseados em teoria de perturbação não conseguem lidar com sistemas em que todas as escalas são simultaneamente relevantes. Há vários pesquisadores envolvidos na história do GR para este tipo de sistema, entre eles citamos Kenneth Wilson, Michael Fisher, Leo Kadanoff. MUITO MAIS.

5.6 Argumentos de Energia- Entropia

5.7 Escalas de Tempo

Provamos que o método de PS dá o resultado exato no modelo de alcance infinito ferromagnético: os termos desprezados vão para zero no limite termodinâmico.

Usaremos nas próximas seções, várias vezes a integração por ponto de sela para estudar vidros de spin e redes neurais.

O primeiro ponto que deve ser entendido é a já mencionada separação das escalas de tempo das duas classes de variáveis: os spins e as interações. Note que são duas classes de graus de liberdade. Os spins de Ising representam os graus de liberdade magnéticos e as interações J_{ij} dependem das posições.

5.8 Vidro de Spin

Ao modelar sistemas físicos onde o número de graus de liberdade N é muito grande é natural considerar o limite $N \rightarrow \infty$, chamado de limite termodinâmico. Há pelo menos duas motivações para considerar isso. Primeiro, é uma boa aproximação. Esperamos que correções devidas a tamanho finito sejam da ordem $\frac{1}{N} \approx 10^{-23}$; segundo é muito mais fácil tecnicamente trabalhar nesse limite. Isso será válido também para cálculos onde $\frac{1}{N}$ não é tão pequeno.

A pergunta fundamental para poder prosequir é qual a informação disponível e como incluí-la. Suponhamos que dada a situação experimental, possamos supor que a energia (eq. 5.1) tem um valor esperado fixo E . Segue, por máxima entropia que a distribuição de probabilidade é uma distribuição de Boltzmann-Gibbs, com função de partição

$$P(s|J) = \frac{e^{-\beta H(\{s_i|\{J_{ij}\})}}{Z_N(\beta|\{J_{ij}\})} \quad (5.100)$$

$$Z_N(\beta|\{J_{ij}\}) = \sum_{(i,j)} e^{-\beta H(\{s_i|\{J_{ij}\})} \quad (5.101)$$

As interações $\{J_{ij}\}$ são fixas e não sabemos calcular nada, nem para um sistema finito nem para um sistema infinito. Precisamos fazer médias sobre a desordem.

Parece razoável estudar a média:

$$\langle\langle Z_N(\beta|\{J_{ij}\}) \rangle\rangle_{J=} = \int \prod_{(i,j)} dJ_{ij} P(\{J_{ij}\}) \sum_{(i,j)} e^{-\beta H(\{s_i|\{J_{ij}\})} \quad (5.102)$$

onde $H(\{s_i|\{J_{ij}\})$ é dado pela eq. 5.1 e $P(\{J_{ij}\})$ pela eq. 5.2. Mas isto não é razoável. Para ver isto, considere um problema totalmente diferente, com dois tipos de variáveis $\{s_i\}$ e $\{J_{ij}\}$ que evoluem na mesma escala de tempo, com hamiltoniano

$$H(\{s_i\}, \{J_{ij}\}) = H(\{s_i|\{J_{ij}\}) - \log P(\{J_{ij}\}). \quad (5.103)$$

A sua função de partição será dada pela eq. 5.102. Esta não pode ser nada mais que uma aproximação ao nosso problema original com escala de tempos separadas ou a função de partição de um outro problema. Como aproximação recebe o nome de *recozida* ou *annealed*. A desordem imposta pelo fato que

as interações não evoluem no tempo é em oposição, chamada de *temperada* ou *quenched*.

Queremos ainda olhar para valores médios, mas quais? A resposta natural⁴ é olhar para o valor esperado da energia livre. Olhe para uma região λ_0 de volume finito de N sítios numa rede. Sua energia livre será

$$F_{\lambda_0}(\beta, \{J_{ij}\}_{\lambda_0}) = -\frac{1}{\beta} \log Z(\{J_{ij}\}_{\lambda_0}). \quad (5.104)$$

onde a função de partição mostra a dependência explícita nos acoplamentos. Olhemos para outras regiões λ_i , que não se sobrepõem, e para a união $\Lambda = \bigcup_i \lambda_i$. Esperamos que a energia livre seja extensiva. Então $F_\Lambda = \sum_i F_{\lambda_i}$. Se para a energia livre o limite termodinâmico ($\Lambda \rightarrow \infty$) for independente da realização particular dos acoplamentos⁵, então teremos

$$\lim_{\Lambda \rightarrow \infty} \frac{1}{|\Lambda|} F_\Lambda(\beta, \{J_{ij}\}_\Lambda) = \lim_{\Lambda \rightarrow \infty} \frac{1}{|\Lambda|} \langle\langle F_\Lambda(\beta, \{J_{ij}\}_\Lambda) \rangle\rangle_J \quad (5.105)$$

Estamos interessados então na energia livre por sítio $f(\beta)$

$$\begin{aligned} f(\beta) &= -\lim_{\Lambda \rightarrow \infty} \frac{1}{|\Lambda|\beta} \int \prod_{(i,j)} dJ_{ij} P(\{J_{ij}\}) \log \left(\sum_{(i,j)} e^{-\beta H(\{s_i\})} \right) \\ &= -\lim_{\Lambda \rightarrow \infty} \frac{1}{|\Lambda|\beta} \langle\langle \log Z_\Lambda(\beta, \{J_{ij}\}) \rangle\rangle_J \end{aligned} \quad (5.106)$$

Mas podemos justificar isto de uma forma mais interessante, recorrendo ao nosso método favorito, que passa por primeiro perguntar qual é a informação relevante, e depois maximizar a entropia.

5.8.1 Máxima Entropia

A distribuição apropriada a considerar é a conjunta $P(s, J)$. O problema se reduz a decidir como introduzir a separação das escalas de tempos. Não sabemos explicitamente as interações $\{J_{ij}\}$, mas sabemos que não evoluem. Portanto pelo menos a distribuição de probabilidades marginal $P(J)$ se for conhecida em um dado instante será a mesma posteriormente. Então impomos os vínculos

$$\begin{aligned} \int ds P(s, J) &= P(J) \\ \int ds P(s|J) &= 1 \\ \int ds P(s|J) H(s|J) &= E \end{aligned} \quad (5.107)$$

⁴depois de pensar bastante ou seguir a indicação de Edwards e Anderson

⁵Isto não é óbvio e muito menos rigoroso

e procuramos maximizar $S[P(s, J)]$. A propriedade de agrupamento da entropia pode ser vista para a distribuição conjunta $P(s, J) = P(J)P(s|J)$

$$\begin{aligned}
S[P(s, J)] &= - \int ds dJ P(s, J) \log P(s, J) = \\
&= - \int ds dJ P(J) P(s|J) \log(P(J)P(s|J)) \\
&= - \int ds dJ P(J) P(s|J) (\log P(J) + \log P(s|J)) \\
&= - \int dJ P(J) \log P(J) - \int dJ P(J) \int ds P(s|J) \log P(s|J) \\
&= S[P(J)] + \langle\langle S[P(s|J)] \rangle\rangle_J \tag{5.108}
\end{aligned}$$

(Aqui usamos uma notação menos carregada onde as integrais representam a soma sobre configurações.) Portanto temos

$$\langle\langle S[P(s|J)] \rangle\rangle_J = S[P(s, J)] - S[P(J)]. \tag{5.109}$$

Note que tanto faz maximizar $S[P(s, J)]$ ou $\langle\langle S[P(s|J)] \rangle\rangle_J$ dado $P(J)$. Mas $P(s|J)$ já foi determinado (eq. 5.101), assim

$$\begin{aligned}
S[P(s, J)] &= - \int \int dJ ds P(s, J) \log P(s, J) = \\
&= - \int \int dJ ds P(s|J) P_0(J) (\log P(s|J) + \log P_0(J)) \\
&= - \int \int dJ ds \frac{e^{-\beta H(s|J)}}{Z(J)} P_0(J) (\log \frac{e^{-\beta H(s|J)}}{Z(J)} + \log P_0(J)) \\
&= - \int \int dJ ds \frac{e^{-\beta H(s|J)}}{Z(J)} P_0(J) (-\beta H(s|J) - \log Z(J) + \log P_0(J)) \\
&= \beta E + \langle\langle \log Z(J) \rangle\rangle_J + S[P_0(J)] \tag{5.110}
\end{aligned}$$

Podemos concluir que a energia livre do sistema $F(\beta) = E - T \langle\langle S[P(s|J)] \rangle\rangle_J$ é, a menos da constante $S[P_0(J)]$

$$F(\beta) = E - T \langle\langle S[P(s, J)] \rangle\rangle_J = -\frac{1}{\beta} \langle\langle \log Z(J) \rangle\rangle_J \tag{5.111}$$

O leitor reconhece neste ponto o resultado da seção anterior. Novamente vemos que a média sobre a desordem não é diretamente feita da função de partição, mas do seu logaritmo. A vantagem disto é que podemos dar dois passos para frente e um para atrás. Para frente: primeiro, se pudermos calcular $\langle\langle \log Z(J) \rangle\rangle_J$ será eliminada a dependência sobre a desordem, e segundo isso permite tomar o limite termodinâmico. Para atrás, não sabemos como fazer a media do log.

Antes de ver como podemos avançar, considere o seguinte problema para pensar.

Exercício

Suponha que a equação 5.107 fosse substituída por

$$\int ds P(s|J) H(s|J) = E(J) \quad (5.112)$$

Qual seria a conclusão a que o método de máxima entropia levaria? Note que isto representa um vínculo para cada realização da desordem. Há resultados de simulações para certos modelos (Kanter et al) de vidro de spin onde a temperatura da transição parece não ser uma constante, depende da realização dos $\{J_{ij}\}$ e mesmo em extrapolações para o limite termodinâmico as flutuações persistem. Pode ser que este seja um ponto interessante de estudo.

5.8.2 Método das Réplicas

De certa forma esta conclusão: olhar para o valor esperado do logaritmo, complica o problema a ponto de parecer insolúvel. No capítulo (??) olharemos para o método da cavidade que dá uma maneira alternativa e interessante do ponto de vista geométrico, de atacar este problema.

Agora introduzimos o famoso método das réplicas ⁶. Este método foi introduzido por Edwards e Anderson em [?] Para manter o problema bem definido olharemos para o modelo de Sherrington e Kirpatrick. Em capítulos futuros olharemos para outros modelos que podem ser estudados pela mesma técnica.

O método usa a relação ⁷ válida para qualquer $Z > 0$

$$\log Z = \lim_{n \rightarrow 0} \frac{Z^n - 1}{n}, \quad (5.113)$$

que nos leva escrever, para a função de partição $Z(J)$

$$\langle\langle \log Z(J) \rangle\rangle_J = \lim_{n \rightarrow 0} \frac{\langle\langle Z^n \rangle\rangle_J - 1}{n}, \quad (5.114)$$

trocando a ordem do limite $n \rightarrow 0$ e a da integração sobre J . Esta operação, ao menos para alguns modelos, não é um problema, tendo sido justificada de forma rigorosa (Van Hemmen) *R: há prova geral?*

O problema não parece mais fácil do que era, pois ainda não sabemos calcular a média sobre J . Mas há um conjunto de valores de n para os quais sabemos calculá-la: os inteiros. Calculamos então $\langle\langle Z^n \rangle\rangle_J$ para n inteiro e uma vez obtida uma expressão que depende de n , fazemos a continuação analítica $n \rightarrow 0$. O nome do método é evidente uma vez que Z^n é a função de partição de n cópias não interagentes de um sistema descrito por Z . Mas, embora não

⁶Para alguns simplesmente um truque. Não parece adequado chamar de truque uma idéia que foi essencial no avanço da Mecânica Estatística nos últimos 30 anos. É fato que matemáticos conseguiram provar algumas das conjecturas feitas a partir desta idéia e que outras permanecem como desafio e inspiração para nova matemática

⁷O leitor é convidado a brincar com outras representações do logaritmo

interagentes, as diferentes réplicas terão correlações entre si, devido a que os grupos independentes de spins interagem via o mesmo conjunto de interações.

Qual é o parâmetro de ordem que deve se usar para descrever o comportamento do sistema? Este é um problema difícil que deixamos para discutir no final. Pode-se dizer que todas as tentativas de adivinhar o parâmetro de ordem correto são essencialmente fúteis. O método de integração por ponto de sela será usado repetidas vezes. Como foi dito antes, no caso do modelo de alcance infinito, o PS mostra qual é parâmetro de ordem correto. Neste caso será de tal forma rico, de fato é uma função para cada temperatura, que não há chance de chutar o que será correto. Como diz o Parisi, a teoria nos levará quase que sozinha até o resultado final⁸.

O modelo de Sherrington Kirpatrick é determinado pelo Hamiltoniano dado na equação (5.1) e pela distribuição (equação 5.2), a função de partição replicada n vezes é

$$\langle\langle Z(J)^n \rangle\rangle_J = \int \prod dJ_{ij} P_0(J) \sum_{\{s_i^a\}} e^{-\beta \sum_{a=1}^n \sum_{i<j} J_{ij} s_i^a s_j^a + \beta h \sum_{i=1}^N \sum_{a=1}^n s_i^a}. \quad (5.115)$$

onde temos o índice de réplica $a = 1 \dots n$ e colocamos um campo externo h . Os acoplamentos entram de forma quadrática no expoente, e podemos integrar a desordem de forma imediata:

$$\langle\langle Z(J)^n \rangle\rangle_J = \sum_{\{s_i^a\}} \exp \left[\frac{1}{N} \sum_{i<j} \left(\frac{1}{2} \beta^2 J^2 \sum_{a,b=1}^n s_i^a s_j^a s_i^b s_j^b + \beta J_0 \sum_{a=1}^n s_i^a s_j^a \right) + \beta h \sum_{i=1}^N \sum_{a=1}^n s_i^a \right]. \quad (5.116)$$

O termo $\sum_{i<j} \sum_{a,b=1}^n s_i^a s_j^a s_i^b s_j^b = \frac{1}{2} \sum_{a,b} (\sum_i s_i^a s_i^b)^2$ pode ser linearizado, usando novamente o truque da equação 5.54. Para isso introduzimos para cada par (a, b) com $a < b$ uma integral sobre a variável q_{ab} .

Da mesma forma o termo $\sum_{i<j} \sum_{a=1}^n s_i^a s_j^a = \frac{1}{2} \sum_{a=1}^n (\sum_i s_i^a)^2$ pode ser linearizado inserindo n integrais sobre variáveis m_a . Assim

$$\begin{aligned} \langle\langle Z(J)^n \rangle\rangle_J &= \exp \frac{N\beta^2 J^2 n}{4} \sum_{\{s_i^a\}} \exp \left[\frac{\beta^2 J^2}{2N} \sum_{a,b} (\sum_i s_i^a s_i^b)^2 \right. \\ &\quad \left. + \frac{\beta J_0 J^2}{2N} \sum_{a=1}^n (\sum_i s_i^a)^2 + \beta h \sum_{i=1}^N \sum_{a=1}^n s_i^a \right]. \quad (5.117) \end{aligned}$$

⁸A teoria leva Parisi, e ele nos guiará

$$\begin{aligned}
\langle\langle Z(J)^n \rangle\rangle_J &= \exp \frac{N\beta^2 J^2 n}{4} \int \prod_{a,b} dq_{ab} \int \prod_a dm_a e^{-\frac{N\beta^2 J^2}{2} \sum_{a<b} q_{ab}^2 - \frac{N\beta J_0}{2} \sum_a m_a^2} \\
&\times \sum_{\{s_i^a\}_{i=1\dots N, a=1\dots n}} \exp \left[\beta^2 J^2 \sum_{a<b} q_{ab} \sum_i s_i^a s_i^b + \beta \sum_a (J_0 m_a + h) \sum_i s_i^a \right],
\end{aligned} \tag{5.118}$$

this factorizes over sites, we drop the i label and obtain

$$\begin{aligned}
\langle\langle Z(J)^n \rangle\rangle_J &= \exp \frac{N\beta^2 J^2 n}{4} \int \prod_{a,b} dq_{ab} \int \prod_a dm_a e^{-\frac{N\beta^2 J^2}{2} \sum_{a<b} q_{ab}^2 - \frac{N\beta J_0}{2} \sum_a m_a^2} \\
&\times \left(\sum_{\{s_i^a\}_{a=1\dots n}} \exp \left[\beta^2 J^2 \sum_{a<b} q_{ab} s^a s^b + \beta \sum_a (J_0 m_a + h) s^a \right] \right)^N,
\end{aligned} \tag{5.119}$$

Defina \mathcal{H} , o hamiltoniano das *interações* ente réplicas

$$\mathcal{H} = -\beta^2 J^2 \sum_{a<b} q_{ab} s^a s^b - \beta \sum_a (J_0 m_a + h) s^a \tag{5.120}$$

e a sua função de partição

$$\mathcal{Z} = \sum_{a<b} \exp -\mathcal{H} \tag{5.121}$$

$$\begin{aligned}
\langle\langle Z(J)^n \rangle\rangle_J &= e^{\frac{N\beta^2 J^2 n}{4}} \int \prod_{a,b} dq_{ab} \int \prod_a dm_a e^{-\frac{N\beta^2 J^2}{2} \sum_{a<b} q_{ab}^2 - \frac{N\beta J_0}{2} \sum_a m_a^2 + N \log \mathcal{Z}} \\
&= \int \prod_{a,b} dq_{ab} \int \prod_a dm_a e^{-N n f_R(\{q_{ab}\}, \{m_a\})}
\end{aligned} \tag{5.122}$$

5.8.3 Integração pelo ponto de sela

Há um prefator N no expoente, que permite a possibilidade de usar o método de ponto de sela, ver equação 5.55. Temos $n(n-1)/2$ integrais sobre os q_{ab} e n integrais sobre as variáveis m_a . É fácil: identificamos o ponto de sela, fazemos as integrais e obtemos um resultado que depende de n e tomamos o limite $n \rightarrow 0$. Não é tão fácil quanto parece.

Para não carregar a notação, a partir deste ponto q_{ab} e m_a se referem aos valores do ponto de sela

$$\langle\langle Z(J)^n \rangle\rangle_J = e^{-N n \beta f_R(\{q_{ab}\}, \{m_a\})} \tag{5.123}$$

e f_R é dado por

$$f_R(\{q_{ab}\}, \{m_a\}) = -\frac{\beta J^2}{4} + \frac{\beta J^2}{2n} \sum_{a<b} q_{ab}^2 + \frac{J_0}{2n} \sum_a m_a^2 - \frac{1}{n\beta} \log \mathcal{Z}. \tag{5.124}$$

Até aqui parece exato, ao menos quando N é grande, exceto pelo problema que não sabemos ainda qual é o ponto de sela. Este é o ponto central do resto deste estudo do modelo SK.

O limite $n \rightarrow 0$

Tomamos o limite de $n \rightarrow 0$ antes de $N \rightarrow \infty$

$$\langle\langle Z(J)^n \rangle\rangle_J = 1 - Nn\beta f_R(\{q_{ab}\}, \{m_a\}) \quad (5.125)$$

$$\frac{\langle\langle Z(J)^n \rangle\rangle_J - 1}{n} = -N\beta f_R(\{q_{ab}\}, \{m_a\}) \quad (5.126)$$

e vemos que f_R , para $n \rightarrow 0$, é a energia livre do sistema temperado (*quenched*).

$$f(\beta) = \lim_{n \rightarrow 0} f_R(\{q_{ab}\}, \{m_a\}) \quad (5.127)$$

Os parâmetros de ordem

Antes de atacar o problema central, vejamos quais são as vantagens deste enfoque. Repetimos que um dos méritos do PS é indicar a natureza dos parâmetros de ordem. Vejamos como. O ponto de sela é determinado por

$$\frac{\partial f_R}{\partial q_{ab}} = 0 \Rightarrow q_{ab} = \frac{1}{\beta^2 J \bar{Z}} \frac{\partial \mathcal{Z}}{\partial q_{ab}} \quad (5.128)$$

$$\frac{\partial f_R}{\partial m_a} = 0 \Rightarrow m_a = \frac{1}{\beta J_0 \bar{Z}} \frac{\partial \mathcal{Z}}{\partial m_a} \quad (5.129)$$

com f_R dado pela equação 5.124. Usando as equações 5.120 e 5.121 obtemos

$$q_{ab} = \frac{1}{\bar{Z}} \sum_{\{s^c\}_{c=1\dots n}} s^a s^b e^{-\mathcal{H}} = \langle s^a s^b \rangle_{\mathcal{H}} \quad (5.130)$$

$$m_a = \frac{1}{\bar{Z}} \sum_{\{s^c\}_{c=1\dots n}} s^a e^{-\mathcal{H}} = \langle s^a \rangle_{\mathcal{H}} \quad (5.131)$$

Exercício

Mostre que estes parâmetros de ordem estão relacionados (são iguais) a

$$q_{ab} = \langle\langle\langle s_i^a s_j^b \rangle_{H_R} \rangle\rangle_J \quad (5.132)$$

$$m_a = \langle\langle\langle s_i \rangle_{H_R} \rangle\rangle_J \quad (5.133)$$

onde H_R é o hamiltoniano original replicado: $H_R = \sum_c H(s_i^c)$. O segundo parâmetro de ordem é simplesmente a magnetização e seu valor servirá para mostrar a existência de ordem ferromagnético. Mas o primeiro é novo e não tem paralelo no sistema ferromagnético simples. Este está relacionado com o parâmetro de ordem de Edwards e Anderson que foi proposto para descrever a transição de fase de vidro de spin vista experimentalmente (*referencias...*) que

descrevia a idéia que o sistema, a baixas temperaturas e para baixas assimetrias na distribuição de interações J_{ij} (baixo J_0) teria uma fase desordenada mas “congelada” dinamicamente. A magnetização de cada sítio seria diferente de zero, mas a magnetização total seria zero. Discutiremos com mais cuidado esta fase mais adiante.

5.8.4 Simetria de réplicas

Como prometido apareceu uma interação entre as réplicas através do termo em $\mathcal{H} \propto \sum_{a<b} q_{ab} s^a s^b$. Os spins de cada réplica são independentes entre si, mas vivem na mesma configuração de desordem. Imagine que tenho um poço de potencial e jogo bolas independentes entre si. Suponha que há dissipação, as bolas independentes acabam todas no fundo do poço, com posições altamente correlacionadas mesmo que uma não interaja com a outra.

Não sabemos lidar com esse termo. Em particular este problema parece tão complicado quanto o inicial. Examine a equação 5.119. Ela parece a mesma de onde partimos (equação 5.116). Temos integrais quadráticas nas interações (antes J , agora q), somas sobre (antes N e agora n) spins. A forma dos Hamiltonianos é a mesma (antes H agora \mathcal{H}). Pior, agora temos integrais sobre m_a , que antes não tínhamos. Pior ainda, antes tínhamos que fazer $N(N-1)/2$ integrais e agora n vai para zero.

Se os n sistemas são independentes podemos embaralhá-los. Chamamos o primeiro de segundo e o segundo de primeiro. Que diferença faz? Que informação física traz o nome ⁹? Portanto o nosso sistema replicado deve ser simétrico ante permutações do índice de réplicas. Fazemos então a seguinte suposição:

- **RS: Simetria de réplicas** O ponto de sela é simétrico ante permutações dos índices de réplicas: $q_{ab}^* = q$ e $m_a^* = m$, para todo a e b .

Agora poderemos fazer as contas. Dependendo dos parâmetros, ou do modelo esta suposição pode levar a resultados corretos, que podem ser deduzidos de outras maneiras ou comparados com simulações. Há casos em que esta suposição faz toda a teoria afundar, levando a entropias $S[P(s|J)]$ negativas, o que não é possível.

Se há um valor único de q para as $n(n-1)/2$ integrais sobre q_{ab} e um único m para as n integrais sobre m_a , então o hamiltoniano \mathcal{H} fica

$$\mathcal{H}_{RS} = -\frac{1}{2}\beta^2 J^2 q \left(\left(\sum_a s_a \right)^2 - n \right) - \beta(J_0 m + h) \sum_a s_a \quad (5.134)$$

e podemos linearizar, ao preço de uma integração sobre a variável z , o termo quadrático nos spins usando uma variante da equação 5.54: $\exp(A^2/2) = \int \exp(-z^2/2 + zA) dz / \sqrt{2\pi}$, para $A = \beta J \sqrt{q} \sum_a s_a$

$$\mathcal{Z}_{RS} = e^{-\frac{n\beta^2 J^2 q}{2}} \sum_{\{s^c\}} e^{\beta(J_0 m + h) \sum_a s_a} \int Dz e^{z \sum_a s_a} \quad (5.135)$$

⁹Não muda o cheiro, mas isso não é importante para modelagem do sistema olfativo

onde usamos daqui em diante a notação, para a medida gaussiana

$$Dz = \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (5.136)$$

$$\mathcal{Z}_{RS} = e^{-\frac{n\beta^2 J^2 q}{2}} \int Dz \left[\sum_{s^a} e^{\beta(zJ\sqrt{q} + J_0 m + h) \sum_a s^a} \right]^n \quad (5.137)$$

$$\mathcal{Z}_{RS} = e^{-\frac{n\beta^2 J^2 q}{2}} \int Dz e^{n \log \{2 \cosh(\beta(zJ\sqrt{q} + J_0 m + h))\}} \quad (5.138)$$

$$\log \mathcal{Z}_{RS} = -\frac{n\beta^2 J^2 q}{2} + \log \int Dz e^{n \log \{2 \cosh(\beta(zJ\sqrt{q} + J_0 m + h))\}} \quad (5.139)$$

Precisamos uma expressão para n indo a 0

$$\log \mathcal{Z}_{RS} = -\frac{n\beta^2 J^2 q}{2} + \log \int Dz [1 + n \log \{2 \cosh(\beta(zJ\sqrt{q} + J_0 m + h))\}] \quad (5.140)$$

$$\log \mathcal{Z}_{RS} = -\frac{n\beta^2 J^2 q}{2} + n \int Dz \log \{2 \cosh(\beta(zJ\sqrt{q} + J_0 m + h))\}. \quad (5.141)$$

Voltando à energia livre réplica simétrica

$$f_{RS}(q, m) = -\frac{\beta J^2}{4} + \frac{\beta J^2}{2n} \frac{n(n-1)}{2} q^2 + \frac{J_0}{2n} n m^2 - \frac{1}{n\beta} \log \mathcal{Z}. \quad (5.142)$$

$$\begin{aligned} f_{RS}(q, m) &= -\frac{\beta J^2}{4} + \frac{\beta J^2}{4} (n-1) q^2 + \frac{J_0}{2} m^2 \\ &- \frac{1}{n\beta} \left[-\frac{n\beta^2 J^2 q}{2} + n \int Dz \log \{2 \cosh(\beta(zJ\sqrt{q} + J_0 m + h))\} \right], \end{aligned} \quad (5.143)$$

O $\log 2 \cosh$ que aparecera anteriormente reaparece de forma um pouco mais complicada. O campo magnético efetivo $J_0 m + h$ aparece agora com uma contribuição aleatória (z). Introduzimos $h(z) = zJ\sqrt{q} + J_0 m + h$ para diminuir a notação

$$\begin{aligned} f_{RS}(q, m) &= -\frac{\beta J^2}{4} + \frac{\beta J^2}{4} (n-1) q^2 + \frac{J_0}{2} m^2 \\ &+ \frac{\beta J^2 q}{2} - \frac{1}{\beta} \int Dz \log \{2 \cosh(\beta h(z))\}. \end{aligned} \quad (5.144)$$

Fica fácil tomar o limite $n \rightarrow 0$, e juntado os termos

$$f_{RS}(q, m) = -\frac{\beta J^2}{4} (1-q)^2 + \frac{J_0}{2} m^2 - \frac{1}{\beta} \int Dz \log \{2 \cosh(\beta h(z))\}. \quad (5.145)$$

Podemos obter as equações de PS

$$\frac{\partial f_{RS}}{\partial q} = 0 \quad (5.146)$$

$$\frac{\partial f_{RS}}{\partial m} = 0 \quad (5.147)$$

Primeiro da eq. 5.147 para m

$$m = \int Dz \tanh(\beta h(z)). \quad (5.148)$$

onde reconhecemos a eq. de CM anterior, só que agora $h(z) = zJ\sqrt{q} + J_0m + h$ ou seja é como se a magnetização fosse determinada na presença de um campo magnético adicional que é aleatório e tem uma amplitude igual a $J\sqrt{q}$ relacionada com a variância das interações e o parâmetro que indicará se o sistema está ou não na fase de vidro de spin. Precisamos obter q , da eq. 5.146:

$$\frac{\beta J^2}{2}(1-q) = \int Dz \tanh(\beta h(z)) \frac{\partial h(z)}{\partial q}, \quad (5.149)$$

e como $\frac{\partial h(z)}{\partial q} = zJ/(2\sqrt{q})$,

$$\beta J(1-q) = \int Dz \frac{z}{\sqrt{q}} \tanh(\beta h(z)) \quad (5.150)$$

Agora integramos por partes com $u = \tanh(h(z))$ e $dv = zDz$ e obtemos

$$q = \int Dz \tanh^2(\beta h(z)). \quad (5.151)$$

5.8.5 Diagrama de fases

Exercício

Dadas as equações de CM obtidas sob a hipótese de que o ponto de sela é simétrico ante permutações das réplicas:

$$\begin{aligned} \frac{\partial f_{RS}}{\partial q} = 0 &\Rightarrow q = \int Dz \tanh^2(\beta h(z)) \\ \frac{\partial f_{RS}}{\partial m} = 0 &\Rightarrow m = \int Dz \tanh(\beta h(z)) \end{aligned} \quad (5.152)$$

desenhe o diagrama de fases RS no espaço $1/(\beta J)$ (nas ordenadas) contra J_0/J nas abscissas.

5.8.6 Entropia negativa

A solução deste problema, que poderia ser considerado resolvido, está errada. O problema não está na hipótese que m_a é independente das réplicas, pois elas são equivalentes. O problema surge do fato que duas réplicas quaisquer não tem o mesmo *overlap* q_{ab} , para qualquer escolha dos índices a e b , ou seja a escolha do ponto de sela está errado. Não podemos saber isso ainda mas começamos a desconfiar ao calcular a entropia para baixas temperaturas.

Não precisamos de grande generalidade para ver que algo está errado. Nos restringimos ao caso $h = 0$ e $J_0 = 0$.

$$f_{RS}(q, m) = -\frac{\beta J^2}{4}(1-q)^2 - \frac{2}{\beta} \int_0^\infty Dz \log\{2 \cosh(\beta z J \sqrt{q})\}. \quad (5.153)$$

e olhamos para $T \rightarrow 0$.

A partir da equação 6.9 podemos ver que $q \rightarrow 1$ para $T \rightarrow 0$, porque a tangente hiperbólica vai para 1. Fazemos então $q = 1 - \alpha T/J$ e depois calcularemos α :

$$f_{RS}(q, m) = -\frac{a^2 T}{4} - \frac{2}{\beta} \int_0^\infty Dz \beta z J \sqrt{1 - \alpha T/J}. \quad (5.154)$$

$$f_{RS}(q, m) = -\frac{a^2 T}{4} - 2J(1 - \alpha \frac{T}{2J}) \int_0^\infty Dz z \quad (5.155)$$

$$f_{RS}(q, m) = -\frac{a^2 T}{4} - 2J(1 - \alpha \frac{T}{2J}) \frac{1}{\sqrt{2\pi}} \quad (5.156)$$

$$f_{RS}(q, m) = -J\sqrt{\frac{2}{\pi}} + T \left(-\frac{\alpha^2}{4} + \frac{\alpha}{\sqrt{2\pi}} \right). \quad (5.157)$$

Precisamos calcular α e começamos pela equação 6.9

$$q = \int Dz \tanh^2(\beta J \sqrt{q} z) = \int Dz \left(1 - \frac{1}{\cosh^2(\beta J \sqrt{q} z)} \right) \quad (5.158)$$

e do lado esquerdo podemos colocar $q = 1$, e

$$\alpha = \beta J(1 - q) = \beta J \int_{-\infty}^\infty Dz \frac{1}{\cosh^2(\beta J z)} \quad (5.159)$$

Usamos

$$\int_{-\infty}^\infty \operatorname{sech}^2(\beta x) \beta dx = 2 \quad (5.160)$$

que permite ver que $\lim_{\beta \rightarrow \infty} \operatorname{sech}^2(\beta x) \beta = 2\delta(x)$

$$\alpha = 2 \int_{-\infty}^\infty Dz \delta(z) = \sqrt{\frac{2}{\pi}} \quad (5.161)$$

$$f_{RS}(q, m) = -J\sqrt{\frac{2}{\pi}} + T\frac{1}{2\pi}. \quad (5.162)$$

A entropia por spin, no limite de baixas temperaturas é negativa:

$$s_{RS} = -\frac{\partial f_{RS}}{\partial T} = -\frac{1}{2\pi}. \quad (5.163)$$

e a energia do estado fundamental

$$e_{RS} = -J\sqrt{\frac{2}{\pi}} \quad (5.164)$$

exercício

Discuta o que significa entropia negativa. Olhe para a entropia de um gás ideal clássico e note que a entropia pode ser negativa. Ficamos preocupados nesse caso? Claro, e apelamos para a Mecânica Quântica.

Capítulo 6

Modelo de Memória associativa: Hopfield

6.1 Memória associativa

Uma memória é um sistema que guarda informação e que permite a sua recuperação. A memória do computador onde estas notas são escritas funciona definindo um endereço para cada ítem guardado. A busca é feita por endereço e não há nenhuma função, durante o processo de estocagem ou de recuperação, do tipo de informação com que se lida. O valor da informação, o contexto, nada disso é importante para buscar a informação. Mas a nossa memória funciona de forma muito diferente. Em primeiro lugar há varias escalas de tempo importantes, que vão desde décadas até alguns milisegundos. Não pretendemos fazer um modelo realista deste sistema tão complexo. Simplesmente iremos olhar um aspecto muito importante que é o fato que nossa memória trabalha por associação. Um pedaço de uma palavra pode ser completado de forma correta. Uma parte de um texto ou de uma imagem podem ser suficientes para recuperar várias informações adicionais, sem necessidade do endereço onde a informação estava guardada e isso o que chamaremos de associação. Não sabemos como funciona exatamente este processo, é muito complexa, não só do ponto de vista de mecanismos como também de implementação neural e bioquímica. O que iremos olhar aqui é uma caricatura, uma simplificação que nem sabemos se guarda alguma relação com memórias neurais biológicas. Mas, não obstante a sua relativa simplicidade por comparação ao que deveria ser, é suficientemente rica para apresentar várias características que permitem encarar esta metáfora da memória como um modelo interessante a ser estudado antes de passar a modelos mais realistas. Nem por um momento pretendemos modelar algum aspecto da implementação da memória usando modelos realistas dos elementos computacionais encontrados num cérebro.

A idéia principal por trás da memória associativa é a construção de um sistema dinâmico formado por N graus de liberdade $\{s_i\}$ que tem diferentes pontos

fixos. Se houver algum tipo de dissipação, a dinâmica levará o sistema até um dos pontos fixos. Estes são atratores e a cada um deles estão associadas as respectivas bacias de atração. A dinâmica depende de um conjunto de parâmetros $\{J_{ij}\}$ e estes definem a localização dos atratores. As condições iniciais definem a bacia de atração e o atrator que será “lembrado”.

As variáveis $\{s_i\}$ e sua dinâmica definem o problema. Iremos estudar o caso em que os $\{s_i\}$ são variáveis de Ising, i.e tomam valores ± 1 . Isso é o mais simples que pode ser feito para modelar um neurônio que pode estar em um de dois estados, o ativo (1) ou o inativo (-1). Esta idéia já estava presente no neurônio de MacCulloch e Pitts. Um neurônio interage com outros e se a interação passa de um certo limiar, o neurônio fica ativo, se não, inativo. A interação depende dos estados dos outros neurônios e das interações sinápticas entre eles. Neste modelo os parâmetros $\{J_{ij}\}$ tomam o papel das conexões sinápticas. Por exemplo, se a dinâmica for discreta, uma escolha popular para uma dinâmica determinista é

$$s_i(t+1) = \operatorname{sgn} \left(\sum_j J_{ij} s_j(t) \right) \quad (6.1)$$

podendo ocorrer as mudanças em paralelo ou em série (Little Hopfield...) Podemos também definir dinâmicas estocásticas:

$$\begin{aligned} s_i(t+1) &= \operatorname{sgn} \left(\sum_j J_{ij} s_j(t) \right) \text{ com prob } p_i \\ &= \operatorname{sgn} \left(\sum_j J_{ij} s_j(t) \right), \text{ prob } 1 - p_i \end{aligned}$$

o que leva, se as conexões forem simétricas $J_{ij} = J_{ji}$, a um hamiltoniano

$$H = - \sum_{\{s_i\}} J_{ij} s_i s_j \quad (6.2)$$

A probabilidade p pode ser associada às interações e a um temperatura,

$$p_i = \frac{1}{1 + \exp \beta h_i} \quad (6.3)$$

onde $h_i = \sum_j J_{ij} s_j(t)$ é o campo “pós-sináptico” local.

A informação é memorizada através das mudanças que devido a sua apresentação, sofrem os parâmetros $\{J_{ij}\}$ e isto chamamos de aprendizagem. Há diferentes maneiras de fazer isso, ou seja, diferentes formas de aprender. Mais adiante estaremos preocupados em estudar os limites, impostos pela teoria, nos diferentes desempenhos que algoritmos de aprendizados possam ter. Por agora estudaremos os limites de um algoritmo simples, chamado de algoritmo de Hebb ou Hebbiano, que apesar das suas limitações, guarda um pouco de semelhança com processos biológicos que o inspiraram.

Utilidade da mecânica estatística para este problema Para padrões sem viés, teremos

$$P(\xi) = \frac{1}{2}\delta_{\xi,1} + \frac{1}{2}\delta_{\xi,-1} \quad (6.4)$$

e para simplificar podemos estudar padrões com independência

$$\langle\langle \xi_i^\mu \xi_j^\nu \rangle\rangle = \delta_{ij}\delta_{\mu\nu} \quad (6.5)$$

6.1.1 Hebb

aprendizagem

6.1.2 Modelo de Hopfield com número finito de padrões:

$$\alpha = 0$$

As interações são uma soma de termos de Hebb sobre as diferentes padrões

$$J_{ij} = \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu \quad (6.6)$$

e o Hamiltoniano tem a forma de Ising convencional

$$H = - \sum_{(i,j)} J_{ij} s_i s_j = - \frac{1}{2} \sum_{i,j} J_{ij} s_i s_j + C \quad (6.7)$$

onde a primeira soma é como sempre, sobre todos os pares (i, j) e a segunda sobre todos os i e todos os j . A constante C é irrelevante e cancela os termos diagonais J_{ii} . A função de partição $Z(\beta, \{\xi\})$

$$Z_N(p) = \sum_{\{s_i\}} e^{-\beta H} \quad (6.8)$$

Esperamos, ao menos a baixas temperaturas, que as configurações de spins sejam próximas de um dos padrões ξ^μ , se não não será uma memória. Para caracterizar a distância é necessário definir algum tipo de distância, escolhemos

$$m_\mu = \frac{1}{N} \sum_i \xi_i^\mu s_i \quad (6.9)$$

que é $m_\mu =$ (número de sítios que concordam - número de sítios que diferem) dividido por N . Da equação 6.7 vemos que Z pode ser escrito como

$$Z_N(p) = \sum_{\{s_i\}} e^{\frac{N\beta}{2} \sum_\mu (m_\mu)^2} \quad (6.10)$$

Nesta altura vemos a chance de usar novamente o método de linearização. Usamos um método análogo ao usado para o modelo de Ising de alcance infinito, introduzido por Guerra. Para cada termo μ usamos a desigualdade

$$m_\mu^2 \geq 2m_\mu M_\mu - M_\mu^2 \quad (6.11)$$

para qualquer M_μ , e temos a cota

$$Z_N(p) \geq \sum_{\{s_i\}} e^{\frac{N\beta}{2} \sum_\mu (2m_\mu M_\mu - M_\mu^2)} \quad (6.12)$$

Colocamos em evidência as variáveis de spin para obter

$$Z_N(p) \geq \sum_{\{s_i\}} \exp \frac{N\beta}{2} \left(\frac{2}{N} \sum_i s_i \sum_\mu (\xi_i^\mu M_\mu) - \sum_\mu M_\mu^2 \right) \quad (6.13)$$

Podemos introduzir a notação $\mathbf{M} \cdot \boldsymbol{\xi}_i = \sum_\mu M_\mu \xi_i^\mu$ e $\mathbf{M}^2 = \sum_\mu M_\mu^2$

$$Z_N(p) \geq \sum_{\{s_i\}} \exp \frac{N\beta}{2} \left(\frac{2}{N} \sum_i s_i \mathbf{M} \cdot \boldsymbol{\xi}_i - \mathbf{M}^2 \right) \quad (6.14)$$

é óbvio que as somas sobre os s_i fatorizam e assim

$$Z_N(p) \geq \exp\left(-\frac{N\beta}{2} \mathbf{M}^2\right) \prod_i \sum_{s_i} \exp(\beta s_i \mathbf{M} \cdot \boldsymbol{\xi}_i) \quad (6.15)$$

$$Z_N(p) \geq \exp\left(-\frac{N\beta}{2} \mathbf{M}^2\right) \prod_i \exp \log (2 \cosh(\beta \mathbf{M} \cdot \boldsymbol{\xi}_i)) \quad (6.16)$$

$$Z_N(p) \geq \exp\left(-\frac{N\beta}{2} \mathbf{M}^2\right) \exp \sum_i \log (2 \cosh(\beta \mathbf{M} \cdot \boldsymbol{\xi}_i)) \quad (6.17)$$

$$\frac{1}{N} \log Z_N(p) \geq -\frac{\beta}{2} \mathbf{M}^2 + \frac{1}{N} \sum_i \log (2 \cosh(\beta \mathbf{M} \cdot \boldsymbol{\xi}_i)) \quad (6.18)$$

e para P finito, ξ aleatórios, independentes e igualmente distribuídos, a soma sobre os sítios passa a estimar a média e então podemos tomar o limite termodinâmico

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log Z_N(p) \geq \sup_{\{M_\mu\}} \left[-\frac{\beta}{2} \mathbf{M}^2 + \langle \langle \log (2 \cosh(\beta \mathbf{M} \cdot \boldsymbol{\xi})) \rangle \rangle \right] \quad (6.19)$$

Dizemos que a soma sobre sítios é autome diante, a simples soma sobre amostras i.i.d., realiza no limite termodinâmico, o cálculo da média. Note que isto é possível porque o número de padrões é finito. Se não o fosse, teríamos problemas em justificar a autome diança, e os problemas seriam impossíveis porque se $P = \alpha N$ também vai para ∞ com N então a autome diança não vale.

Exercício

Demonstre a propriedade de autome diança.

Exercício

Mostre que, seguindo o argumento usado para o modelo de Ising de alcance infinito, podemos obter a cota inferior

$$\frac{1}{N} \log Z_N(p) \leq -\frac{\beta}{2} \mathbf{M}^2 + \frac{1}{N} \sum_i \log(2 \cosh(\beta \mathbf{M} \cdot \boldsymbol{\xi}_i)) + \frac{P}{N} \log(N+1) \quad (6.20)$$

e junto com a equação 6.19 obtemos a energia livre para o modelo de memória associativa

$$f(\beta, p) = - \lim_{N \rightarrow \infty} \frac{1}{N\beta} \log Z_N(p) = \sup_{\{M_\mu\}} \left[-\frac{1}{2} \mathbf{M}^2 + \frac{1}{\beta} \langle \log(2 \cosh(\beta \mathbf{M} \cdot \boldsymbol{\xi})) \rangle \right]. \quad (6.21)$$

Olhando para o último termo da eq. 6.20, fica claro onde falha este argumento se o número de padrões P for proporcional a N .

Exercício

Mostre que a inclusão de um campo magnético que acopla com cada um dos padrões (termo $\sum_\mu \sum_i h_\mu \xi_i^\mu s_i$ no hamiltoniano) é levada em conta se \mathbf{M} for substituído por $\mathbf{M} + \mathbf{h}$ no termo do cosseno hiperbólico e $\mathbf{h} = (h_1, \dots, h_p)$:

$$f(\beta, p) = - \sup_{\{M_\mu\}} \left[-\frac{1}{2} \mathbf{M}^2 + \frac{1}{\beta} \langle \log(2 \cosh(\beta(\mathbf{M} + \mathbf{h}) \cdot \boldsymbol{\xi})) \rangle \right]. \quad (6.22)$$

6.1.3 Estrutura das soluções

Os valores de \mathbf{M} que levam ao supremo podem ser obtidas diferenciando a energia livre:

$$0 = \frac{\partial f}{\partial M_\mu} = -M_\mu + \langle \xi^\mu \tanh(\beta \mathbf{M} \cdot \boldsymbol{\xi}) \rangle. \quad (6.23)$$

$$M_\mu = \langle \xi^\mu \tanh(\beta \mathbf{M} \cdot \boldsymbol{\xi}) \rangle, \quad (6.24)$$

Há muitas soluções para esta equação e como usual a interpretação adequada torna mais fácil a obtenção de soluções. A inclusão de um campo magnético permite interpretar o significado de \mathbf{M}

$$M_\mu = -\frac{\partial f}{\partial h_\mu} = \frac{1}{N\beta} \frac{\partial \log Z_N}{\partial h_\mu} = \frac{1}{N} \sum_i \langle s_i \rangle \xi_i^\mu, \quad (6.25)$$

que permite, comparando com a equação 6.9 ver que \mathbf{M} é o vetor de distâncias médias entre o sistema de spins e os padrões de memória. Já temos experiência com a equação

$$m = \tanh(\beta m), \quad (6.26)$$

que tem soluções estáveis¹ não triviais a temperaturas $T = \beta^{-1} \leq 1$.

¹Estabilidade em relação à energia livre

Exercício

Coloque $\mathbf{M} = (m, 0, 0 \dots 0)$ e encontre a temperatura de transição onde as soluções de memória, aquelas que tem distância pequena com um só padrão, deixam de existir. Há $2p$ soluções deste tipo. o fator 2 aparece pq $\pm m$ são soluções equivalentes.

Exercício

Coloque $\mathbf{M} = (m, m, 0 \dots 0)$ e encontre a temperatura de transição onde as soluções de mistura de dois padrões deixam de existir. Há $2(p-1)(p-2)$ soluções deste tipo. Porque?

Exercício

Coloque $\mathbf{M} = (m, m, \dots, m, 0, \dots, 0)$ com n termos não nulos e encontre a temperatura de transição onde as soluções de mistura de n padrões deixam de existir. Há $2^n \frac{p!}{n!(p-n)!}$ soluções deste tipo. Porque?

Note que estas temperaturas vão decrescendo a medida que o número de padrões misturados vai aumentando.

Exercício

Existem outras soluções?

6.2 O Modelo de Hopfield: Memória Associativa: Número de padrões extensivo

$$J_{ij} = \frac{1}{N} \sum_{\mu} \xi_i^{\mu} \xi_j^{\mu} \quad (6.27)$$

$$H = - \sum_{i < j} J_{ij} s_i s_j \quad (6.28)$$

$$Z(\{J_{ij}\}) = - \sum_{\{s_i\}} e^{-\beta H} \quad (6.29)$$

Novamente a energia livre será obtida tomando a média do logaritmo da função de partição

$$f(\beta) = - \lim_{N \rightarrow \infty} \frac{1}{N\beta} \langle \langle \log Z_N(\beta, \{J_{ij}\}) \rangle \rangle_J \quad (6.30)$$

$$f(\beta) = - \lim_{\Lambda \rightarrow \infty} \frac{1}{|\Lambda|\beta} \int \prod_{(i,j)} dP(\{J_{ij}\}) \log \left(\sum_{(i,j)} e^{-\beta H(\{s_i\})} \right) \quad (6.31)$$

e usamos o método das réplicas para poder tratar o logaritmo.

Os padrões são tais que $\langle \xi_i^\mu \rangle = 0$ e $\langle \xi_i^\mu \xi_i^\nu \rangle = \delta_{ij} \delta_{\mu\nu}$. A função de partição replicada

$$\langle\langle (Z_N(\beta, \{J_{ij}\}))^n \rangle\rangle_J = \langle\langle \sum_{\{s_i^\mu\}} e^{\frac{\beta}{N} \sum_\mu \sum_a \sum_{i<j} \xi_i^\mu \xi_j^\mu s_i^a s_j^a} \rangle\rangle_{\{\xi\}}$$

introduzimos a idéia de overlap, que mede a distância entre a configuração de spins e cada um dos padrões:

$$M_\mu^a = \frac{1}{N} \sum_i \xi_i^\mu s_i^a$$

então

$$\langle\langle (Z_N(\beta, \{J_{ij}\}))^n \rangle\rangle_J = \langle\langle \sum_{\{s_i^\mu\}} e^{\frac{N\beta}{2} \sum_\mu \sum_a (M_\mu^a)^2} \rangle\rangle_{\{\xi\}} \quad (6.32)$$

reconhecendo que podemos linearizar o expoente, usando $\exp(aA^2/2) = \int_{-\infty}^{\infty} \exp(-az^2/2 \pm aAz)/\sqrt{2\pi}$ uma vez para cada termo quadrático, ou seja para cada índice de réplica e cada padrão, com $A = M$ e $a = N\beta$

$$\langle\langle (Z^n) \rangle\rangle_J = \langle\langle \sum_{\{s_i^\mu\}} \prod_\mu \prod_{a=1}^n \int_{-\infty}^{\infty} e^{-\frac{N\beta(m_\mu^a)^2}{2} + N\beta m_\mu^a M_\mu^a} \rangle\rangle \quad (6.33)$$

Agora entra a arte de Amit, Gutfreund e Sompolinski. Queremos estudar este modelo pelo seu funcionamento como uma memória associativa. Devemos estar atentos a um padrão, que será diferente dos outros, no sentido que esse é o padrão lembrado. Qual padrão num caso prático, dependerá das condições iniciais. Talvez, como ocorreu no caso de P finito, poderá haver mistura de um número pequeno (não extensivo) de padrões.

Esse padrão especial será chamado de padrão condensado e por acaso ele é o número $\mu = 1$

$$\langle\langle (Z^n) \rangle\rangle_J = \sum_\xi P(\{\xi\}) \sum_{\{s_i^\mu\}} \int_{-\infty}^{\infty} e^{-\frac{N\beta}{2} \sum_{a=1}^n ((m_1^a)^2 + 2m_1^a M_1^a)} e^{-\frac{N\beta}{2} \sum_{\mu=2}^P \sum_{a=1}^n ((m_\mu^a)^2 + 2m_\mu^a M_\mu^a)} \quad (6.34)$$

Suponha que fizéssemos a integral de ponto de sela, novamente a interpretação do valor de m no ponto de sela será

$$m_\mu^a = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \langle s_i^a \rangle \quad (6.35)$$

para o padrão condensado esperamos que, ao menos a baixas temperaturas, m seja $\mathcal{O}(1)$, e para os padrões não condensados, pelo fato dos padrões serem

independentes esperamos $m_\mu^a \approx \frac{1}{N}\sqrt{N} = \frac{1}{\sqrt{N}}$. A contribuição de P padrões torna o termo quadrático da energia livre devido aos padrões não condensados

$$\sum_{\mu=2}^P (m_\mu^a)^2 \approx \frac{P}{N} \quad (6.36)$$

para P finito isto é desprezível no limite termodinâmico. No caso que estamos interessados esperamos que isto contribua algo de $\mathcal{O}(1)$, (e.g. α).

Mudamos variáveis para os padrões não condensados $\rho_\mu^a = m_\mu^a \sqrt{N\beta}$, agora esperamos que seus valores de ponto de sela sejam $\mathcal{O}(1)$, e obtemos, para a última linha de 6.40, após fazer a média sobre os padrões não condensados

$$\begin{aligned} & \sum_{\xi, \mu=2 \dots P} P(\{\xi\}) e^{-\frac{N\beta}{2} \sum_{\mu=2}^P \sum_{a=1}^n ((m_\mu^a)^2 + 2m_\mu^a M_\mu^a)} \\ = & \sum_{\xi, \mu=2 \dots P} P(\{\xi\}) e^{-\sum_{\mu=2}^P \sum_{a=1}^n \left(\frac{1}{2}(\rho_\mu^a)^2 + \sqrt{\frac{\beta}{N}} \sum_i \rho_\mu^a \xi_i^\mu s_i^a \right)} \\ = & e^{-\sum_{\mu=2}^P \sum_{a=1}^n \frac{1}{2}(\rho_\mu^a)^2} \sum_{\xi, \mu=2 \dots P} P(\{\xi\}) e^{\sqrt{\frac{\beta}{N}} \sum_i \rho_\mu^a \xi_i^\mu s_i^a} \\ = & e^{-\sum_{\mu=2}^P \sum_{a=1}^n \frac{1}{2}(\rho_\mu^a)^2} e^{\sum_{\mu=2}^P \sum_i \log \cosh \left(\sqrt{\frac{\beta}{N}} \sum_{a=1}^n \rho_\mu^a s_i^a \right)} \end{aligned} \quad (6.37)$$

O próximo passo é comum nas provas do teorema central do limite: estamos interessados no limite termodinâmico, e usamos para N grande que $\log \cosh \left(\sqrt{\frac{\beta}{N}} A \right) \approx \log \left(1 + \frac{\beta}{2N} A^2 \right) \approx -\frac{\beta}{2N} A^2$

$$\begin{aligned} & e^{-\sum_{\mu=2}^P \sum_{a=1}^n \frac{1}{2}(\rho_\mu^a)^2} e^{-\frac{\beta}{2N} \sum_{\mu=2}^P \sum_i \sum_{a=1, b=1}^n \rho_\mu^a \rho_\mu^b s_i^a s_i^b} \\ = & e^{-\frac{1}{2} \sum_{\mu=2}^P \sum_{a, b=1}^n \rho_\mu^a \tilde{R}_{ab} \rho_\mu^b} \end{aligned} \quad (6.38)$$

onde

$$\tilde{R}_{ab} = \delta_{ab} - \frac{\beta}{N} \sum_i s_i^a s_i^b \quad (6.39)$$

Voltando isto à equação 6.40

$$\begin{aligned} \langle\langle Z^n \rangle\rangle_J &= \sum_{\xi_i^1} P(\{\xi_i^1\}) \sum_{\{s_i^\mu\}} \int_{-\infty}^{\infty} e^{-\frac{N\beta}{2} \sum_{a=1}^n ((m_1^a)^2 + 2m_1^a M_1^a)} \\ &\times \int_{-\infty}^{\infty} e^{-\frac{1}{2} \sum_{\mu=2}^P \sum_{a, b=1}^n \rho_\mu^a \tilde{R}_{ab} \rho_\mu^b} \end{aligned} \quad (6.40)$$

Fizemos a média sobre os padrões não condensados pagando o preço que introduzir interações entre as réplicas (veja termo com $s_i^a s_i^b$). Isto é natural, diferentes réplicas acabam tendo correlações porque as interações intra-réplicas são correlacionadas, pois memorizam a mesma informação. Não é diferente do que ocorreu no modelo SK.

As integrais sobre os ρ s são relativamente fáceis. Para cada padrão não condensado obtemos $1/\sqrt{\det R}$ e seria suficiente se soubesemos calcular esse determinante, mas ele depende das variáveis de estado de neurônio s_i^a . Podemos começar a atacar este problema pagando o preço de introduzir $\frac{n(n-1)}{2}$ integrais sobre uma nova classe de variáveis q_{ab} que serão interpretadas, nos seus valores de ponto de sela de forma muito semelhante ao overlap entre as réplicas que apareceu no SK. Introduzimos um 1 escrito assim

$$1 = \int \prod_{a < b} dq_{ab} \delta \left(q_{ab} - \frac{1}{N} \sum_i s_i^a s_i^b \right) \quad (6.41)$$

$$\begin{aligned} & \int \prod_{a < b} dq_{ab} \delta \left(q_{ab} - \frac{1}{N} \sum_i s_i^a s_i^b \right) \int_{-\infty}^{\infty} \prod d\rho_\mu^a e^{-\frac{1}{2} \sum_{\mu=2}^P \sum_{a,b=1}^n \rho_\mu^a \tilde{R}_{ab} \rho_\mu^b} \\ &= \int \prod_{a < b} dq_{ab} \delta \left(q_{ab} - \frac{1}{N} \sum_i s_i^a s_i^b \right) \left(\frac{1}{\sqrt{\det R}} \right)^{(P-1)} \\ &= \int \prod_{a < b} dq_{ab} \delta \left(q_{ab} - \frac{1}{N} \sum_i s_i^a s_i^b \right) \left(\frac{1}{\prod \lambda_r} \right)^{\frac{P-1}{2}} \\ &= \int \prod_{a < b} dq_{ab} \delta \left(q_{ab} - \frac{1}{N} \sum_i s_i^a s_i^b \right) e^{-\frac{P-1}{2} \sum_r \log \lambda_r} \end{aligned} \quad (6.42)$$

onde os $\{\lambda_r\}$ são os autovalores da matriz

$$R_{ab} = (1 - \beta)\delta_{ab} - \beta Q_{ab} \quad (6.43)$$

onde $Q_{ab} = q_{ab}$ fora da diagonal e é zero na diagonal. Precisamos lidar com as funções delta : introduzimos ainda mais uma classe de variáveis de integração r_{ab} . É útil usar a representação de Fourier após girar o contorno de integração

$$\begin{aligned} \langle\langle Z^n \rangle\rangle_J &= \int_{-\infty}^{\infty} \prod dm_1^a \prod d\rho_\mu^a \prod dq_{ab} dr_{ab} e^{-\frac{\alpha N \beta^2}{2} \sum_{ab} q_{ab} r_{ab}} \\ &\times \sum_{\{s_i^a\}} e^{\frac{\alpha \beta^2}{2} \sum_{ab,i} s_i^a s_i^b r_{ab}} e^{-\frac{P-1}{2} \sum_r \log \lambda_r} \\ &\times \sum_{\xi} P(\{\xi\}) e^{-\frac{N\beta}{2} \sum_{a=1}^n ((m_1^a)^2 + \frac{2}{N} m_1^a \sum_i \xi_i^1 s_i^a)} \end{aligned} \quad (6.44)$$

olhemos para a parte que depende das variáveis s_i e notemos que fatoriza sobre

os índices de sítio

$$\begin{aligned}
 &= \langle\langle \sum_{s_i^a} e^{\frac{\alpha\beta^2}{2} \sum_{ab,i} s_i^a s_i^b r_{ab}} e^{-\beta \sum_{a=1}^n m_1^a \sum_i \xi_i^1 s_i^a} \rangle\rangle \\
 &= \langle\langle \exp \left\{ \sum_i \log \sum_{s^a} e^{\frac{\alpha\beta^2}{2} \sum_{ab} s^a s^b r_{ab}} e^{-\beta \sum_{a=1}^n m_1^a \xi^1 s^a} \right\} \rangle\rangle \\
 &= \exp N \left\{ \langle\langle \log \sum_{s^a} e^{\frac{\alpha N \beta^2}{2} \sum_{ab} s^a s^b r_{ab}} e^{-\beta \sum_{a=1}^n m_1^a \xi^1 s^a} \rangle\rangle \right\} \quad (6.45)
 \end{aligned}$$

na última linha usamos o mesmo truque que no caso de P finito

$$(6.46)$$

$$(6.47)$$

$$(6.48)$$

$$(6.49)$$

$$(6.50)$$

$$(6.51)$$

$$(6.52)$$

$$(6.53)$$

Capítulo 7

Monte Carlo

7.1 Integração Numérica em espaços de alta dimensão

Considere o método de integração numérica mais simples, chamado método do trapézio (ver de Vries). Aproximamos a integral

$$I = \int_a^b f(x) dx$$

por

$$I_T = \frac{1}{N} \left(\frac{1}{2} f(x_1) + \sum_{i=2}^{N-1} f(x_i) + \frac{1}{2} f(x_N) \right), \quad (7.1)$$

podemos mostrar que o erro cometido é proporcional a h^2 , onde $h = (b - a)/N$, escrevemos então que

$$I = I_T + \vartheta(h^2).$$

Esta estimativa do erro também vale para integrais multidimensionais. Métodos mais sofisticados, baseados neste (e.g. estilo Romberg-Richardson), levam a melhorias no expoente de h , mas como veremos a seguir, não suficientes.

O custo computacional no cálculo de uma integral é proporcional ao número de vezes que a rotina que calcula o integrando é chamada dentro do programa. Na fórmula do trapézio acima este número de chamadas é N . Suponhamos um problema típico de Mecânica Estatística, por exemplo um gás dentro de uma caixa. Temos da ordem de $k = 10^{23}$ moléculas mas digamos que para poder lidar com o problema temos somente $k = 20$. Uma aproximação drástica, mas veremos não suficiente. Neste caso é necessário lidar com integrais do tipo

$$Z = \int g(\{r_{ix}, r_{iy}, r_{iz}\}) dr_1^3 dr_2^3 \dots dr_k^3$$

uma integral em $d = 3k = 60$ dimensões. Suponhamos que o volume da caixa seja $V = L^3$, e dividimos cada uma dos d eixos em intervalos de tamanho h .

Isto significa uma grade com

$$N = \left(\frac{L}{h}\right)^d$$

pontos. Suponhamos que escolhemos um h extremamente grande, tal que $L/h = 10$, ou seja cada eixo será dividido em somente 10 intervalos. Assim temos

$$N = 10^{60}$$

pontos na grade e esperamos ter um erro talvez da ordem de 10^{-2} . O quê significa um número tão grande como 10^{60} ? Suponhamos que a máquina que dispomos é muito veloz, ou que a função que queremos integrar é muito simples, tal que cada chamada à subrotina demore somente 10^{-10} segundos. O tempo que demorará para calcular I_T é 10^{50} s. Para ver que isso é muito basta lembrar que a idade do universo é da ordem de $4 \cdot 10^{17}$ s, portanto nosso algoritmo levará da ordem de 10^{31} idades do universo. Não precisamos muito mais para que nos convençamos a procurar outro método de integração. Variantes do método de trapézio não ajudam muito. Infelizmente o que temos disponível, o Monte Carlo não é muito preciso, mas é muito melhor que isso.

7.2 Monte Carlo

7.2.1 Teorema Central do Limite: revisitado

Considere uma variável X com valores x em um intervalo dado e distribuição $P(x)$. Assumimos que os valores médios \bar{x} e $\overline{x^2}$ existem e são finitos.¹ A variância σ_x é definida por

$$\sigma_x^2 = \overline{x^2} - \bar{x}^2$$

que também é finita.

Considere ainda uma sequência de N amostragens independentes de X : $\{x_i\}_{i=1, \dots, N}$, e outra variável Y com valores y dados por

$$y = \frac{1}{N} \sum_{i=1}^N x_i$$

Assintoticamente, isto é para N grande, a distribuição de y se aproxima de uma distribuição gaussiana, podemos escrever que aproximadamente

$$P(y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(y-\bar{y})^2}{2\sigma_y^2}}$$

A aproximação é boa na região central da gaussiana e melhora quando N cresce. Mais detalhes no futuro (ou em aulas anteriores). O valor médio de y e sua variancia são

$$\bar{y} = \bar{x} \text{ e } \sigma_y = \frac{\sigma_x}{\sqrt{N}}$$

¹Definimos os momentos $\overline{x^n} = \int x^n P(x) dx$

Notem que se o objetivo for encontrar o valor esperado de x , que é \bar{x} , e não for possível realizar a integral, podemos estimar \bar{x} a partir de y (isso pode ser generalizado para o cálculo de $\bar{f} = \int f P(x) dx$.) Qual é vantagem sobre simplesmente fazer uma medida (amostragem) de x ? É que neste último caso o erro seria da ordem de σ_x , enquanto que a estimativa baseada em y terá erro estimado em $\sigma_y = \sigma_x/\sqrt{N}$, portanto **o erro da estimativa é independente da dimensão de x** . Para grandes dimensões isso é uma grande vantagem. O problema é que para reduzir o erro por um fator 2 é necessário trabalhar 4 vezes mais duro. E isso para o caso em que as variáveis são independentes e condicional que sabemos gerar as amostras.... . O erro pode ser diminuído não só aumentando N mas também se mudarmos σ_x . Esse é o objetivo da técnica de amostragem por importância.

Exercício : Considere uma variável aleatória X que toma valores $-\infty < x < \infty$, com probabilidade $P(x)$. é dado que $\sigma_x^2 = \overline{x^2} - \bar{x}^2$ é finito. Dado $y = \frac{1}{N} \sum_{i=1}^N x_i$ mostre, a partir de

$$P(y) = \int \cdots \int dx_1 \cdots dx_N \delta \left(y - \frac{1}{N} \sum_{i=1}^N x_i \right) \prod_{i=1}^N P(x_i)$$

que $P(y)$ é aproximada por uma gaussiana para N grande. Determine a variância de y .

Exercício: Distribuição de Cauchy Considere o problema acima, exceto que $\sigma_x^2 = \overline{x^2} - \bar{x}^2$ é infinito pois $P(x) = \frac{b}{\pi(b^2+x^2)}$. Encontre a distribuição $P(y)$ de y , Note que não é gaussiana para nenhum valor de N . As integrais necessárias são relativamente fáceis de calcular pelo método dos resíduos.

7.2.2 Monte Carlo

A idéia básica é aproximar uma integral I por I_{MC}

$$I = \int_a^b f(x) dx \simeq I_{MC} = \frac{1}{N} \sum_{i=1}^N f(x_i) \quad (7.2)$$

onde os $\{x_i\}$ são escolhidos aleatoriamente de forma independente da distribuição uniforme em $[a, b]$. Se a integral de f^2 existir e for finita, e se as amostras $f(x_i)$ forem estatisticamente independentes - e isto é um grande *se* - então o erro da estimativa MC acima será dado por

$$\sigma_{I_{MC}} = \frac{\sigma_f}{\sqrt{N}}$$

e podemos estimar σ_f a partir dos dados da amostragem

$$\sigma_f^2 \approx \frac{1}{N} \sum f^2(x_i) - \left[\frac{1}{N} \sum f(x_i) \right]^2.$$

Embora eq. (7.2) possa ser usada para o cálculo da integral, em geral é necessário reduzir a variancia da função f . Isso é possível através de uma mudança de variáveis, que nem sempre pode ser implementada analiticamente e será descrita a seguir².

O método que iremos descrever não é útil, em geral, para realizar estimativas de Monte Carlo, mas servirá para motivar e sugerir novos caminhos. Imagine uma integral da forma

$$I = \int f(x)w(x)dx,$$

em geral essa separação do integrando em duas funções é muito natural. Tipicamente x é um vetor em um espaço de muitas dimensões mas $f(x)$ só depende de algumas poucas componentes de x , enquanto que $w(x)$ depende de todas. Suponha que $w(x)$ esteja normalizado. i.e:

$$\int w(x)dx = 1$$

Ilustraremos a separação em uma dimensão, tomemos o intervalo de integração $(0, 1)$ e façamos a seguinte mudança de variáveis

$$y(x) = \int_0^x w(z)dz \quad (7.3)$$

$$y(0) = 0, y(1) = 1$$

então $dy = w(x)dx$ e a integral toma a forma

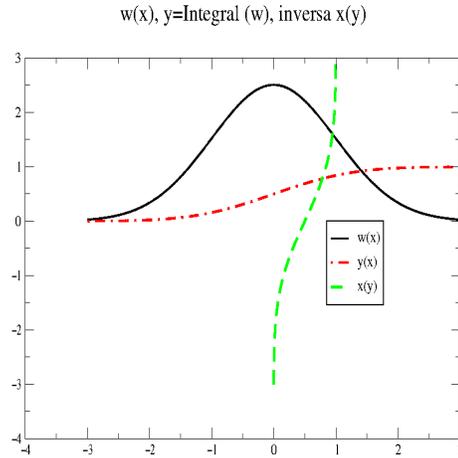
$$I = \int f(x(y))dy$$

e a aproximação Monte Carlo é

$$I = \int_a^b f(x)w(x)dx \simeq I_{MC} = \frac{1}{N} \sum_{i=1}^N f(x(y_i)) \quad (7.4)$$

onde os valores de y_i serão amostrados de uma distribuição uniforme no intervalo $(0, 1)$. Depois basta calcular a função que relaciona y e x (eq. [7.3]). A função inversa permite calcular o valor de x onde deverá ser calculada a função $f(x)$. Este método assume que saibamos fazer a integral da equação 7.3, mas não é em geral possível fazê-lo de forma analítica.

²Uma forma trivial de conseguir a redução de σ_f é considerar variações da identidade $\int_0^1 f(x)dx = \int_0^1 g(x)dx$, onde $g(x) = \frac{1}{2}(f(x) + f(1-x))$. Note que o cálculo de g é duas vezes mais caro que o de f , portanto devemos ter $\frac{\sigma_f^2}{2\sigma_g^2} > 1$ para ter ganho efetivo



7.2.3 Exemplos analíticos.

Ao realizar um cálculo MC teremos, tipicamente, acesso a um gerador de números aleatórios distribuídos uniformemente em $(0, 1)$. O objetivo é, aqui de forma analítica e posteriormente, de forma numérica, mostrar como gerar números aleatórios distribuídos de acordo com uma distribuição dada a partir da distribuição disponível. Apresentaremos dois casos muito úteis que podem ser feitos de forma analítica.

Se duas variáveis (em e.g. R^N) tem uma relação funcional $y = \sigma(x)$, então suas densidades de probabilidade estão relacionadas assim

$$P_Y(y)dy = P_X(x)dx$$

$$P_Y(y)dy = P_X(x) \left| \frac{\partial x}{\partial y} \right| dy \quad (7.5)$$

onde $\left| \frac{\partial x}{\partial y} \right|$ é o jacobiano da transformação e $dy = \prod_i dy_i$. No caso de interesse numérico temos aproximadamente

$$P_Y(y)dy = dy, \quad 0 \leq y_i < 1, i = 1 \dots N$$

e zero fora.

Distribuição Exponencial

Suponha que queremos gerar amostras de uma distribuição exponencial. i.e $P_X(x) = \exp(-x)$. Integrando a eq. (7.5) obtemos

$$y(x) = \int_0^{y(x)} P_X(x) \frac{dx}{dy} dy$$

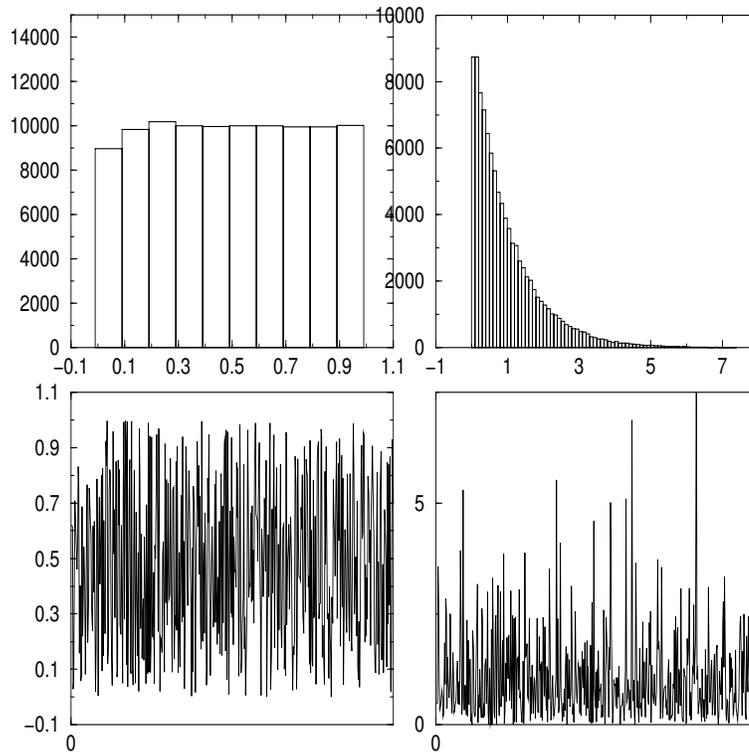
$$y(x) = \int_0^x P_X(x) dx = \int_0^x e^{-z} dz$$

$$y(x) = 1 - \exp(-x)$$

ou $x = -\ln(y)$ terá a distribuição exponencial desejada, pois se y é uniforme em $(0, 1)$ então $1 - y$ também o é. Portanto é suficiente para gerar números distribuídos exponencialmente usar uma função que gera números aleatórios de distribuição uniforme `RAND(SEED)` e somente uma linha de (pseudo-) código

```
x=-log( RAND(SEED))
```

Compare na figura a distribuição uniforme (esquerda) e a a exponencial (direita) (abaixo : série temporal, acima : histogramas)



Distribuição Normal

Para gerar números distribuídos de acordo com a distribuição normal é tentador gerar um número grande de amostras de $P_Y(y)$ e definir $x = \frac{1}{\sqrt{N}} \sum y_i - \frac{\sqrt{N}}{2}$, que terá distribuição gaussiana (aproximadamente). O problema é o custo computacional, pois requer N chamadas da função `RAN` para gerar uma só amostra de x . Portanto nunca gere números aleatórios gaussianos dessa maneira.

Mais fácil, do ponto de vista computacional é partir da equação (7.5). O método de Box-Muller, mostrado a seguir é muito mais eficiente, pois gera dois números gaussianos para duas chamadas da função geradora de uniformes. Dados y_1 e y_2 obtemos x_1 e x_2 a partir da transformação:

$$\begin{aligned}x_1 &= \sqrt{-2 \ln y_1} \cos 2\pi y_2 \\x_2 &= \sqrt{-2 \ln y_2} \sin 2\pi y_2\end{aligned}$$

mostraremos que a sua distribuição conjunta será $P_X(x_1, x_2) = \frac{1}{2\pi} \exp(-(x_1^2 + x_2^2)/2)$. Integrando a eq.(7.5) temos:

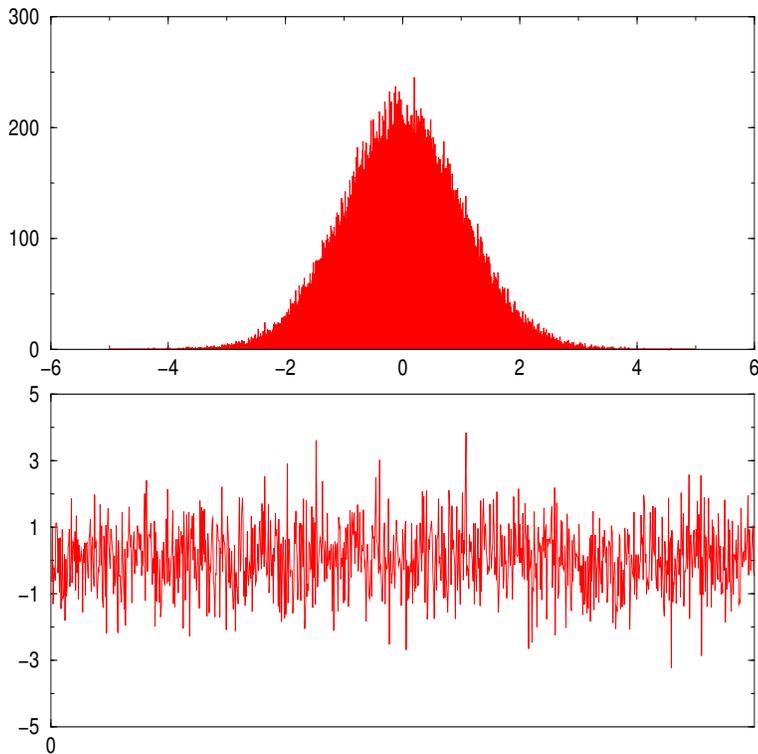
$$\int \int P_Y(y(x_1, x_2)) \left| \frac{\partial y}{\partial x} \right| dy_1 dy_2 = \int \int P_X(x) dx_1 dx_2$$

segue o resultado pois o jacobiano é:

$$J = \left| \frac{\partial y}{\partial x} \right| = \frac{y_1}{2\pi} = \frac{1}{2\pi} e^{-\frac{x_1^2 + x_2^2}{2}}$$

Usando este método obtemos a figura que segue, abaixo temos a série temporal e acima o histograma dos desvios normais:

Estes resultados de muita utilidade na simulação de distribuições gaussianas multivariadas, a ser discutidas posteriormente.



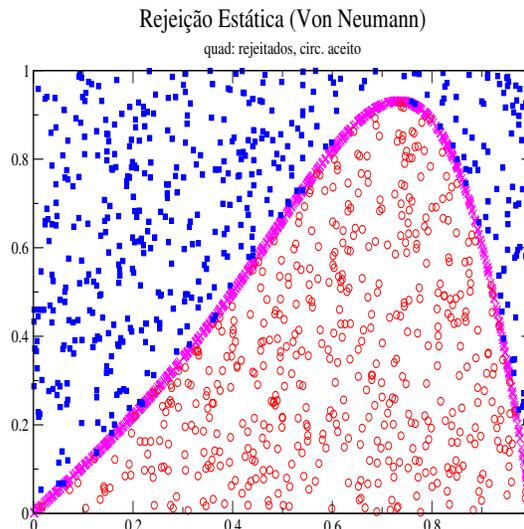
7.2.4 Métodos Estáticos: rejeição

Raramente é possível realizar as integrais que permitem descobrir a transformação exata de variáveis e devemos então encontrar uma forma gerar diretamente os x com a distribuição $w(x)$. Os métodos que apresentaremos podem ser divididos em duas classes, estáticos e dinâmicos. Na primeira os números são gerados independentemente um dos outros³, enquanto que na segunda classe, construiremos um processo dinâmico que usara informação anterior para gerar o próximo número.

Suponhamos que a região onde $w(x) \neq 0$ está contida em (a, b) e que ela é limitada, tal que $w(x) < c$. No método de rejeição estático geramos dois NAU ξ e η e definimos

$$\rho = a + (b - a)\xi, \quad \varphi = c\eta$$

o valor de ρ será aceito como o novo valor de x se $\varphi \leq w(\rho)$ e rejeitado se não.



A seqüência de números aceitos x são as abscissas dos círculos na figura acima.

7.2.5 Círculo

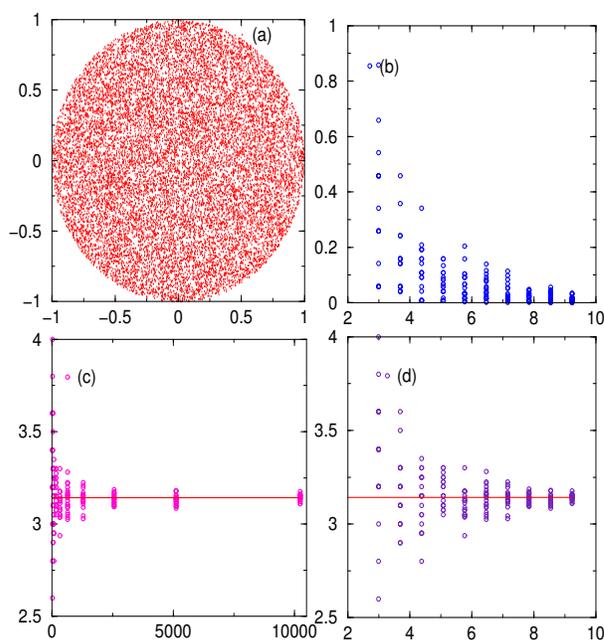
Exemplo: calcule π

A figura mostra os resultados de algumas simulações para estimar π . Foram gerados N_{MC} pares de números aleatórios (x, y) . Se $z = x^2 + y^2 \leq 1$ então o ponto é aceito, de outra forma é rejeitado. Os resultados foram obtidos para $N_{MC} = 10 * 2^m$ passos de Monte Carlo, com $m = 2, 4, \dots, 20$. O resultado (figura (a) abaixo esq. acima) mostra os pares aceitos. Continuando no sentido horário, temos os resultados respectivamente :

³Tão independentemente quanto o gerador de números pseudo-aleatórios o permitir.

- (b) do erro absoluto contra $\log(N_{MC})$
- (d) resultado de $\pi_{MC}=(\text{numero aceito}/\text{numero total})$ contra $\log(N_{MC})$
- (c) resultado de $\pi_{MC}=(\text{numero aceito}/\text{numero total})$ contra N_{MC} , os gráficos mostram os resultados de 20 corridas independentes. A dispersão dos pontos nos dá uma idéia dos erros estatísticos. As barras horizontais mostram o valor 3.14159

i: (a) pontos aceitos, (b) erro abs, (c) pi vs N , (d) pi vs logN



7.2.6 Métodos Dinâmicos

A idéia por trás dos processos de Monte Carlo dinâmicos é a de um processo estocástico em tempo discreto. Um processo determinístico, em oposição, é tal que dado um certo conjunto de informações, é possível -em principio- determinar a evolução futura. Um processo estocástico serve para modelar o caso em que a informação é incompleta e às várias possibilidades de evolução são atribuídas probabilidades. O objetivo é construir um processo estocástico com distribuição de equilíbrio associada igual ao $w(x)$ dado. Note que o processo estocástico é uma caminhada aleatória. Consideremos um grande número de caminhadas independentes. O processo deve ser tal que a fração das caminhadas na vizinhança de x seja proporcional a $w(x)$, pelo menos se aproxime dela

assintoticamente no tempo, e chamaremos de $P(x, t)$ à distribuição no instante t .

O conceito principal para entender o processo de MC dinâmico é a probabilidade de transição, $\Gamma(x|x_n, x_{n-1}, \dots, x_0, \dots)$, que em princípio pode depender de toda a história da evolução. Um processo é chamado Markoviano (de 1 passo) se só depende da estado atual⁴

$$\Gamma(x_{n+1}|x_n, x_{n-1}, \dots, x_0) = \Gamma(x_{n+1}|x_n),$$

ou de forma vaga, para onde o processo vai (o futuro), depende somente de onde está agora (o presente) e não do passado. Chamaremos a sequência $\{x_0, x_1, \dots, x_n, \dots\}$ de cadeia de Markov⁵. Para os nossos objetivos estas cadeias são ferramenta suficiente.

Um ingrediente necessário que o processo deverá satisfazer é convergência para o equilíbrio. A distribuição de equilíbrio ou invariante ou estacionária deve satisfazer a condição de estacionaridade

$$w(x) = \int w(z)\Gamma(x|z)dz, \quad (7.6)$$

mas se não for estacionária teremos a relação entre a probabilidade no instante t e no seguinte $t + 1$ dada por

$$P(x, t + 1) = \int P(z, t)\Gamma(x|z)dz,$$

Dado que as probabilidades de transição são normalizadas $1 = \int \Gamma(z|x)dz$ segue que

$$\begin{aligned} \Delta P(x, t) &= P(x, t + 1) - P(x, t) = \int P(z, t)\Gamma(x|z)dz - P(x, t) \int \Gamma(z|x)dz, \\ \Delta P(x, t) &= P(x, t + 1) - P(x, t) = \int [P(z, t)\Gamma(x|z) - P(x, t)\Gamma(z|x)] dz \quad (7.7) \end{aligned}$$

A interpretação é imediata, a variação da probabilidade, de um instante para o outro, tem duas contribuições, de entrada e saída. O primeiro termo $[P(z, t)\Gamma(x|z)] dz$ representa o número de caminhadas em um volume dz em torno de z no instante t , que fizeram a sua transição para x no instante $t + 1$. O segundo termo representa a saída, isto é os que estavam em x e escapam para z . A integral leva em conta todas as contribuições do espaço. É óbvio a partir das eqs. [7.6, 7.7]

$$\Delta w(x) = \int [w(z)\Gamma(x|z) - w(x)\Gamma(z|x)] dz = 0,$$

o que sugere uma condição

$$w(z)\Gamma(x|z) = w(x)\Gamma(z|x) \quad (7.8)$$

⁴outra notação comum é $\Gamma(x_n \rightarrow x_{n+1})$

⁵A cadeia de Markov é caracterizada pelas probabilidades de transição e pela distribuição inicial de probabilidades de x

que se a matriz de probabilidade de transições satisfizer então $w(x)$ será estacionária. Esta condição, chamada de **balanceamento detalhado**, não é necessária, mas só suficiente. Além de haver motivações físicas para impô-la como condição deve ser ressaltado que é talvez a forma mais fácil de realizar o objetivo para construir a matriz de transição. Com qualquer escolha que satisfaça a condição eq. [7.8] $w(x)$ é um ponto fixo da dinâmica. Mas a pergunta que resta é sobre a estabilidade. É razoável esperar a estabilidade dado que se em t , $P(x, t) > w(x)$, o número de caminhantes que sairão da região de x para z será maior que o que sairiam se a probabilidade fosse $w(x)$. Analogamente, se em t , $P(x, t) < w(x)$ então o número será menor.

Há várias maneiras de satisfazer a equação [7.8]. Embora todas levem a algoritmos corretos, no sentido que

$$I = \int_a^b f(x)w(x)dx \simeq I_{MC} = \frac{1}{N} \sum_{i=1}^N f(x_i) \quad (7.9)$$

é uma aproximação que melhora para maiores valores de N , algumas serão eficientes enquanto outras não. Diferentes escolhas levam a diferentes sequências, e a pergunta relevante é: quanta informação nova é trazida por uma nova amostragem? A função de autocorrelação normalizada, que é fundamental para poder julgar a eficiência do MC, é definida por

$$C(k) \equiv \frac{\langle f_n f_{n+k} \rangle - \langle f \rangle^2}{\langle f_n f_n \rangle - \langle f \rangle^2}$$

onde

$$\langle f \rangle = \int f(x)w(x)dx$$

$$\langle f_n f_{n+k} \rangle = \int \int f(x_n) f(x_{n+k}) w(x_n) \Gamma^k(x_{n+k}|x_n) dx_{n+k} dx_n$$

e

$$\Gamma^k(x_{n+k}|x_n) = \int \dots \int \Gamma(x_{n+k}|x_{n+k-1}) \Gamma(x_{n+k-1}|x_{n+k-2}) \dots \Gamma(x_{n+1}|x_n) dx_{n+k-1} dx_{n+k-2} \dots dx_{n-1}$$

é a probabilidade de transição em k passos. É óbvio que não é, em geral, possível calcular a autocorrelação, mas podemos estimá-la a partir das amostras colhidas:

$$C_{MC}(k) \equiv \frac{\langle f_n f_{n+k} \rangle_{MC} - \langle f \rangle_{MC}^2}{\langle f_n f_n \rangle_{MC} - \langle f \rangle_{MC}^2}$$

onde definimos a média (empírica) sobre a amostra de dados

$$\langle f_n f_{n+k} \rangle_{MC} = \frac{1}{N-k} \sum_{i=1}^{N-k} f(x_i) f(x_{i+k})$$

Tipicamente -mas não sempre - $C(k)$ tem um decaimento exponencial:

$$C(k) = e^{-k/\tau}$$

τ é tempo de correlação exponencial e mede a eficiência do processo em gerar números aleatórios independentes distribuídos de acordo com $w(x)$. Agora podemos escrever

$$I = \int_a^b f(x)w(x)dx \simeq I_{MC} = \frac{1}{N} \sum_{i=1}^N f(x_i) \pm \sigma_f \sqrt{\frac{2\tau}{N}}$$

onde assumimos que depois de um tempo (em unidades de 1 passo MC) aproximadamente 2τ as novas amostras serão estatisticamente independentes e o número efetivo de amostras será reduzido por esse fator.

Outro tempo importante é τ_R , o tempo de relaxação para o equilíbrio. Este mede quanto tempo demora para que o processo estocástico perca memória das condições iniciais e os x sejam efetivamente representativos de $w(x)$. Do ponto de vista de eficiência é razoável não considerar e.g. os primeiros $10\tau_R$ passos gerados pelo processo. Se $C(k)$ efetivamente decair exponencialmente esses dois tempos são iguais, mas há casos em que não, e.g. perto de transições de fase críticas.

7.2.7 Algoritmo de Metropolis

O processo de geração dos números x_n será separado em duas partes. Em primeiro lugar definimos a probabilidade de *tentativa de mudança* $T(x_T|x_n)$, que determina a probabilidade de estando no tempo n em x_n , seja escolhido o ponto x_T como candidato ao próximo passo da sequência. Uma vez gerado x_T passamos à segunda parte, que é onde se decide se é feita a transição $x_n \rightarrow x_{n+1} = x_T$, ou seja x_T é aceito ou se não. Neste caso de rejeição fazemos a transição trivial $x_n \rightarrow x_{n+1} = x_n$, de forma que x_n é incluído novamente na sequência, Isto é feito introduzindo a *matriz de aceitação* $A(x_{n+1}|x_T)$. Ou seja

$$\Gamma(x|z) = A(x|z)T(x|z)$$

e a condição de balanceamento detalhado, para todo par de pontos $x \neq z$ toma a forma

$$A(x|z)T(x|z)w(z) = A(z|x)T(z|x)w(x)$$

que é satisfeita por uma família de escolhas possíveis, em particular se definirmos

$$A(x|z) = F\left(\frac{w(x)T(z|x)}{w(z)T(x|z)}\right)$$

e F tal que

$$\frac{F(a)}{F(1/a)} = a \text{ para todo } a$$

A escolha mais comum, para a probabilidade de tentativa de mudança é tomar

$$T(z|x) = \text{Const dentro de uma bola centrada em } x$$

isso leva a uma taxa de tentativas simétricas ($T(z|x) = T(x|z)$), e portanto basta tomar

$$\frac{A(x|z)}{A(z|x)} = \frac{w(x)}{w(z)}$$

A escolha associada ao nome de Metropolis () é

$$F(a) = \min(1, z)$$

o que leva ao seguinte algoritmo:

1. escolha o valor inicial x_0
2. dado x_n determinaremos x_{n+1} : escolha um valor de tentativa x_T (uniformemente dentro de uma bola de raio d em torno de x_n)
3. verifique se $w(x_T)$ é maior ou menor que $w(x_n)$.
 - Se $w(x_T) \geq w(x_n)$ então **aceita** : $x_{n+1} = x_T$
 - Se $w(x_T) \leq w(x_n)$ então escolhe um número aleatório uniforme $0 \leq \xi < 1$ e

aceita : $x_{n+1} = x_T$ se $w(x_T) \geq w(x_n)\xi$

rejeita : $x_{n+1} = x_n$ se $w(x_T) \leq w(x_n)\xi$

volta ao item 2

Imagine o caso em que a função $w(x)$ pode ser parametrizada da forma

$$w(x) = \frac{e^{-\beta E(x)}}{Z}$$

esse é um dos casos mais interessantes (distribuição d Boltzmann-Gibbs) e a função $E(x)$ é interpretada como a energia de um sistema no estado x ou a função custo de um processo. Z é uma constante em relação a x mas depende do parâmetro β que em física é interpretado como o inverso da temperatura. Este tipo de função ocorre quando a probabilidade que devemos atribuir a uma dada configuração é baseada na informação que temos sobre o valor medio $\langle E(x) \rangle$ e é o resultado de encontrar a distribuição com a máxima entropia consistente com a informação dada.

O algoritmo de Metropolis pode ser redescrito da seguinte forma:

1. escolha o valor inicial x_0

2. dado x_n determinaremos x_{n+1} : escolha um valor de tentativa x_T (uniformemente dentro de uma bola de raio d em torno de x_n)
 3. verifique se $E(x_T)$ é maior ou menor que $E(x_n)$.
 - Se $E(x_T) \leq E(x_n)$ então **aceita** : $x_{n+1} = x_T$
 - Se $E(x_T) \geq E(x_n)$ então escolhe um número aleatório uniforme $0 \leq \xi < 1$ e
 - aceita** : $x_{n+1} = x_T$ se $\exp(-\beta(E(x_T) - E(x_n))) \geq \xi$
 - rejeita** : $x_{n+1} = x_n$ se $\exp(-\beta(E(x_T) - E(x_n))) \leq \xi$
- volta ao item 2

A processo realiza a caminhada aleatória de forma que uma diminuição na energia é sempre aceito, mas se há uma tentativa de escolha de um lugar de energia mais alta, a tentativa não é autoamaticamente rejeitada. Se o aumentode energia for muito grande então sim é rejeitada, mas se não for , então é aceita. A escala de grande ou pequeno é determinada pela razão dos fatores de Boltzmann de cada configuração.