

NESTOR CATICHA

PROBABILIDADES E
MECÂNICA ESTATÍSTICA
4300401

Sumário

1	<i>Teoremas de Regraduação de Cox</i>	5
2	<i>Outras definições de probabilidade</i>	29
3	<i>Uso elementar de Probabilidades</i>	35
4	<i>Frequência e Probabilidade</i>	53
5	<i>Teorema do Limite Central</i>	79
6	<i>Entropia</i>	95
7	<i>Entropia: como usar</i>	111
8	<i>Aplicações Simples</i>	137
9	<i>Aplicações: Modelo de Ising e Campo Médio</i>	171
10	<i>Monte Carlo</i>	197

1

Teoremas de Regraduação de Cox

Alea jacta est

Júlio César

1.1 Introdução: Determinismo Newtoniano ou aleatório?

Júlio César ao cruzar com seu exército o Rio Rubicom quebrou uma regra na República Romana. Não havia volta. Ou conseguia o poder ou perdia tudo. Qual seria o desenlace da sua ação? Nem ele sabia e segundo Suetônio teria dito: *Alea jacta est*. A sorte está lançada. Saber estimar as consequências de uma ação é aconselhável para poder decidir que curso tomar. César talvez tenha procedido da seguinte forma. Primeiro fez uma lista das possibilidades à sua frente. Uma decisão é tomada e uma das possibilidades seguidas. Estas poderiam incluir: (Ação I) Continuar na Gália. (Ação II) Fazer uma aliança com Pompeu, (Ação III) Fugir de Pompeu, (Ação IV) Se aposentar, (Ação V) Voltar a Roma com seu exército e lutar contra Pompeu. Historiadores certamente poderiam incluir outras. Como decidir? Supomos que uma escolha foi feita. Quais as consequências? Para cada curso de ação ele deve ter feito uma lista de possibilidades. Suponha que considere tomada a Ação V. Então as consequências poderiam ser (Consequência 1 da Ação V) Vitória total, com a formação do Império e ele como Imperador. (Consequência 2 da Ação V) Derrota total levando à sua morte. (Consequência 3 da Ação V) Guerra Civil interminável ...etc. Mas não devia acreditar que cada uma das possibilidades teria a mesma chance de ocorrer. A cada consequência de cada Ação, César poderia ter associado um valor numérico indicando sua crença na chance de ocorrer. Veremos que isto será codificado em probabilidades de ocorrência. Mas também poderia ter associado um valor numérico de quão feliz ele seria se efetivamente essa consequência ocorresse. Estes números descrevem o que se chama de utilidade, de cada possibilidade, para o agente Júlio César. Parece óbvio que as utilidades dependem do agente, mas talvez não seja óbvio que as probabilidades também dependam do agente, ou melhor do que este sabe. Resumindo, Júlio César decidiu o seu curso de ação após identificar as possibilidades de ação, das consequências de cada ação, das chances de cada consequência ocorrer, e da utilidade ou felicidade que cada consequência teria. Neste curso

não falaremos sobre decisão a partir das utilidades. Atualmente, em geral, este tópico não cabe em cursos de Física. Faremos um estudo sistemático sobre a chances de algo ocorrer sem importar quão feliz você fique com cada possibilidade. O ponto central será definir com cuidado o que queremos dizer com *chances*, como atribuir números e como mudá-los quando recebemos informação.

Teria Júlio César dúvidas sobre sua sorte ou saberia mais que os outros atores do drama? Se soubesse mais talvez estaria jogando um jogo de cartas marcadas enquanto os outros jogariam a cegas. A frase também implica num certo determinismo. Não há nada a fazer. O curso natural das coisas conduzirá os atores. Como observadores verão simplesmente o desenrolar da história.

Há alguma inconsistência em pensar que as consequências são inevitáveis por um lado, e por outro ficar torcendo para ter sorte? Seria como torcer ao ver a gravação de um jogo de futebol que já foi jogado, mas não sabemos o resultado. Talvez seja um exercício interessante ver grandes jogos do passado sem saber qual jogo é, torcendo para seu time ganhar com direito a ficar tão feliz como quando o jogo é assistido ao vivo.

Todas estas situações são complexas. Começemos por algo mais simples. Uma das maiores revoluções intelectuais da história da humanidade foi a introdução da Mecânica por Newton. Sabemos que caso fosse necessário temos o formalismo da Mecânica para poder calcular a trajetória de uma moeda. O determinismo Newtoniano permite fazer previsões sobre o futuro a partir do estado atual. Por outro lado, os casos mais associados à sorte são o jogo de dados ou um jogo de cara ou coroa com uma moeda. Não é por acaso que a frase de César que teria sido dita em grego menciona $\kappa\nu\beta\omicron\sigma$, o cubo ou dado. Estes jogos deram origem a o estudo matemático das probabilidades.

Como podemos associar a uma moeda simultaneamente as propriedades de ser um sistema determinista, governado pelas leis de Newton e a condição de exemplo mais usado ao falar de sistemas aleatórios? É necessário ter cuidado com as palavras. O que significa *aleatório*? Teremos todo este curso para atribuir-lhe significado. Em geral, ao ser usado coloquialmente, significa que não é totalmente determinado *a priori* por eventos passados.

As possibilidades do estado da moeda são determinados ao especificar 12 números. 3 dizem respeito à sua posição, por exemplo do centro de massa. Sua orientação é determinada por 3 ângulos. Veja, num livro de Mecânica a definição de ângulos de Euler. Ou senão, simplesmente considere 2 eixos no plano da moeda e um terceiro perpendicular ao plano e as rotações em torno deles. Esse número é duplicado ao levar em conta as suas derivadas temporais (velocidades). A dinâmica em 12 dimensões é dada pelas equações de Newton ¹. É óbvio que as equações não são suficientes para determinar como cairá a moeda. Há muitas maneiras de jogar a moeda, mas só um conjunto de equações. As mesmas equações devem ser complementadas com diferentes conjuntos de condições iniciais que

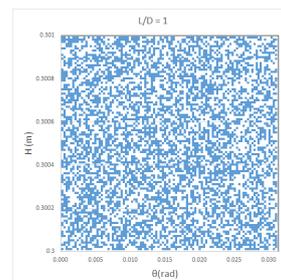


Figura 1.1: Integração das equações de movimento da moeda. Espaço (restrito) de condições iniciais para uma moeda cilíndrica. A moeda é solta de uma altura "grande". Simulações por Guilherme Galanti and Osame Kinouchi, que gentilmente autorizaram o uso da figura, fazendo a ressalva que a simulação é de um sistema mais simples que uma moeda. Na realidade é uma estrutura formada de quatro massas nos vértices ligadas por seis molas nas arestas e diagonais do que seria em repouso um retângulo. O sistema está restrito a duas dimensões e a cada batida mesa há dissipação de energia. É um modelo de uma moeda ou um cubo simplificado

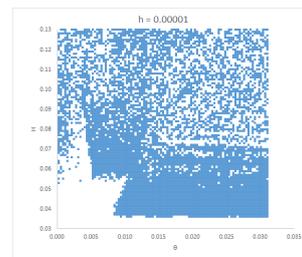


Figura 1.2: Igual à anterior, mas a moeda é solta de uma altura menor, para diferentes ângulos.

¹ Nem a dedução destas equações e muito menos a sua solução, serão necessárias aqui, mas cabem num curso de Mecânica.

parametrizam cada trajetória possível. As figuras 1.1 e 1.2 mostram porque não há incompatibilidade nessas duas caracterizações da moeda. Por simplicidade fixamos 10 parâmetros e olhamos o que ocorre quando dois parâmetros são mudados numa certa região. As figuras foram construídas de forma totalmente determinística. Cada ponto é colorido de acordo com a face mostrada pela moeda. Azul cara, branco coroa. Vemos que a aleatoriedade não está na evolução dinâmica descrita por Newton, mas na ignôncia que poderíamos ter sobre as condições iniciais. Se ao jogarmos a moeda não tivermos conhecimento muito preciso das condições iniciais, não teremos como prever se o ponto final será azul ou branco. Este é um indício que o conhecimento pode influenciar as probabilidades (que ainda não sabemos o que são) de que caia cara ou cora. Dois agentes apostando neste jogo terão chances diferentes de ganhar se tiverem informações diferentes sobre o modo como a moeda será jogada. Note que para alturas muito pequenas, o poder de predição fica mais forte, pois há regiões grandes com a mesma cor. Faça a experiência. Segure uma moeda com os dedos na posição horizontal. Solte a moeda, sem girá-la, de uma altura de 1 metro, 10 cm, 1 cm, 1 mm. Seu poder de prever o que vai ocorrer aumenta. O determinismo é igualmente descrito pelas equações de Newton em todas as condições. A incerteza na previsão tem a ver com a forma como se solta a moeda.

Ainda isto é coloquial e não sabemos o que é probabilidade, informação ou aleatório. O objetivo do que segue é vestir isto com uma estrutura matemática. A história do desenvolvimento das ideias é complexo e não é o interesse destas notas. Porém elas estarão salpicadas de referências a grandes figuras do passado. A história contada, não é certamente como ocorreu, porque isso não sabemos. A seguir discutiremos as ideias que vem de Jakob Bernoulli, Laplace, Maxwell, Kolmogorov, Ramsey, Keynes, Pólya, Jeffreys, Jaynes entre outros. Começaremos a história no meio contando como R. Cox tentou criar uma extensão da lógica Booleana, com origens na Grécia antiga, para situações de informação incompleta. Ele poderia ter suposto de início que a estrutura matemática era a de probabilidades, mas se recusou a isso. Tentou encontrar essa estrutura e ao descobrir que era ou a teoria de probabilidade ou uma regradação monotônica trivial, primeiro se convenceu da impossibilidade de escapar dessa estrutura e segundo forneceu uma sólida interpretação para o que queremos dizer com informação e como molda nossas crenças e para o que queremos dizer com probabilidades.

Há vários exemplos de tentativas de axiomatizar extensões da lógica a situações de informação incompleta. Savage e Lindley são exemplos importantes, mas seu objetivo é descrever o processo de tomada de decisão e isso leva a considerar utilidades. O caminho que escolhemos leva à mesma estrutura de probabilidades deixando claro que decisões é um capítulo a parte. O objetivo de um físico é descrever a natureza fazendo previsões e não tomando decisões.

1.2 *Informação completa ou incompleta*

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 com 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.

Pode parecer estranho para o estudante de 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. Mas a matemática é um tipo de linguagem que tem a vantagem de permitir a quem a usa ser muito cuidadoso com o que diz. 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 ². 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 ferra-

² Embora o formalismo a ser introduzido também possa ser usado nesta direção, mas não agora.

menta 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, 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. A teoria de probabilidades segue porque é a extensão da lógica a situações de informação incompleta. Mas até aqui não sabemos o que é lógica, informação e incompleta.

A análise da lógica remonta a Aristóteles e passa por Boole no século XIX, que contribuiu para que a lógica pudesse ser representada em linguagem matemática.³ Uma lógica envolve (i) um conjunto de proposições supostas verdadeiras, (ii) um método de dedução para estabelecer a validade de argumentos e (iii) um método para estabelecer invalidades.

Um argumento lógico é composto por duas partes. Um conjunto de asserções, chamadas as premissas e uma única asserção chamada de conclusão. Um argumento é válido se a conclusão pode ser obtida aplicando as regras (ii) e (iii).

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 como verdade, 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.

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, novamente no contexto C , ocorre quando é dado como verdade a negação \bar{B} ou seja " B é Falso". Segue \bar{A} , isto é, que " A é Falso" como conclusão inescapável. Note que se A não fosse falso, B não poderia sé-lo.

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, o que quer

³ Veja para uma comparação: Aristotle's Prior Analytics and Boole's Laws of Thought, John Corcoran, History and Philosophy of Logics 2003

que isto seja, e que será discutido mais à frente, será necessária para avançar.⁴

A forma dedutiva da lógica permite somente tres tipos de respostas, *sim*, *não* e *não segue*⁵ A indução nos força ou permite dividir esta última em várias possibilidades e os casos extremos nesse espectro são aqueles onde 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 ou *desiderata*⁶ que a teoria deveria satisfazer, e estes serão então os axiomas da extensão da lógica. Aqui podemos discordar, propor outros desejos ou axiomas, mas uma vez aceitos serão provados os teoremas de reparatização 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*⁷ 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 condicionalidade⁸. 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 ditas 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 *Desiderata à la 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 Bernoulli Frank P. Ramsey e George Pólya. A exposição de Jaynes teve uma grande influência, mas ainda recebeu críticas e complementos⁹. Eu seguirei a apresentação de A. Caticha, que é mais completa e clara, mas farei algumas pequenas mudanças¹⁰.

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

⁴ Segundo 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)

⁵ Nem o leitor nem o autor destas notas deve neste momento ceder à tentação de discutir lógicas de um ponto de vista mais geral. Precisamos um subconjunto de Lógica proposicional, não muito mais que lógica Aristotélica, como exposta por George Boole. Talvez caiba aqui a desculpa "I have not worked out the mathematical logic of this in detail, because this would, I think, be rather like working out to seven places of decimals a result only valid to two. My logic cannot be regarded as giving more than the sort of way it might work". Frank P. Ramsey (1926) "Truth and Probability", in Ramsey, 1931, *The Foundations of Mathematics and other Logical Essays*, Ch. VII, p.156-198, editado por R.B. Braithwaite, 1999 electronic edition.

⁶ *Desiderata*: as coisas desejadas, em Latim. Termo usado em filosofia para denotar um conjunto de propriedades essenciais de alguma estrutura. Alguns ficam tentados a chamar *axiomas*.

⁷ Ao leitor que demande uma definição de racional, podemos dizer que pelo menos não queremos ser manifestamente irracionais. Não acredito que haja uma definição de consenso sobre o que é ser racional. Há consenso porém em apontar alguns casos de irracionalidade.

⁸ A maior fonte de erros será devido a falhas na especificação cuidadosa das asserções condicionantes. Aparentemente a notação $a|b$ com a a a asserção a ser analisada e b a asserção condicionante é devida a John Maynard Keynes, no seu Tratado.

⁹ Tribus, A. C

¹⁰ Notem que há lugar ainda para avanços nestes primeiros passos. Tentem encontrar defeitos, generalizações, melhores argumentos.

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.¹¹ Isto é o que mostraremos a seguir.

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. A *Desiderata* que segue tem cinco desejos denotados $DP_1 \dots DP_5$ e é um bom exercício tentar mostrar que não fazem sentido. Se você conseguir e convencer outros terá feito uma grande contribuição, se não terá mais respeito pelas conclusões que seguem.

1.2.2 DP_1 Representação de crenças e transitividade

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 $A|C \succ A|D$. 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 a transitividade:

- DP_1 : Se $A|B \succ A|C$ e $A|C \succ A|D$ então deve ser o caso que $A|B \succ A|D$

Além disso, dadas duas crenças podemos imaginar que há outra asserção intermediária.

Isto é fácil de satisfazer se impusermos:

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

Podemos satisfazer este tipo de ordenamento representado crenças com números racionais. A escolha de números reais permite usar integrais, o que não é pouco, pois fazer somas é difícil. Nota que sempre usamos integrais em física, mesmo que o espaço tenha uma estrutura subjacente (e.g 10^{-31} m). Não sabemos se tem, mas nos modelos para o *mundo* usados em Mecânica, os pontos do espaço e tempo vivem numa variedade real.

Dados

$$A|B > A|C$$

e

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

¹¹ Este comentário parece trivial, mas o uso que será dado a seguir é totalmente não trivial. 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 que o chamou de indução eliminativa. Usaremos novamente este estilo de fazer teoria ao introduzir o conceito de entropia.

segue imediatamente, uma vez que são números reais, que

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

de acordo com o desejo DP_1 . 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 . Também não diz que a representação das crenças seja única. Uma mudança dos números estritamente monotônica crescente não mudará a ordem. Isto levará a que há famílias de atribuições de números que representam a ordem da mesma forma.

1.2.3 DP_2 Asserções compostas:

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: AB , $A \wedge B$ ou ainda **A e B** .
- A **soma** ou disjunção de duas asserções é uma terceira asserção, que costuma ser denotada por $A + B$ ou $A \vee B$, ou ainda **A ou B** .

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

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

Tabela 1.1

Tabela verdade para a negação e algumas asserções compostas.

Isso significa que $A + B = \overline{\bar{A} \bar{B}}$ portanto o conjunto de operações negação e conjunção permite construir a disjunção de asserções.

Ao falar de silogismos introduzimos a operação \Rightarrow que significa implicação. Se é verdade que $A \Rightarrow B$, significa que se A é verdade segue B . Isto não é um novo operador pois é equivalente dizer que C é verdade para $C = \overline{A \wedge \bar{B}}$.

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 ou negações. 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:

- DP_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 DP_2 acima declara.

1.2.4 DP_3 Informação completa

Um tratamento geral de situações de informação incompleta deve abarcar os casos particulares de informação completa. Então olhemos para casos simples em que há informação completa.

O mais simples é $a|a$ que é a plausibilidade de algo que sabemos ser verdade, para qualquer a .

Se $a|bc$ e $b|ac$ representam a plausibilidade de algo que sabemos ser falso com certeza, chamamos a e b de mutuamente exclusivos na condição c . Poderia ser que hajam falsidades absolutas mais falsas que outras falsidades absolutas; ou verdades absolutas mais verdadeiras que outras verdades absolutas, mas achamos razoável impor

- DP_3 : Existem dois números v_v e v_f tal que 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. Este desejo inclui também a negação de uma asserção, pois a asserção e sua negação são mutuamente exclusivos, e estamos dizendo que $\bar{a}|a = v_f$ para qualquer a .

1.2.5 DP_4 Soma e DP_5 Produto

Como sugerido na tabela 1, todo operador na álgebra Booleana pode ser representado pelas operações conjunção a e b (denotada ab ou $a \wedge b$) e negação de a (denotada por \bar{a})¹², isto é, o produto e a negação lógicas. A soma lógica pode ser obtida usando $a \vee b = \overline{\bar{a}\bar{b}}$. Precisamos então analisar a 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, conjunção e negação. Mas é mais fácil, olhar para a conjunção e a disjunção, e junto com DP_3 obteremos a forma geral de tratar a negação.

Agora olhamos para a disjunção ou soma lógica. Novamente c se refere à informação subjacente e estamos interessados na plausibili-

¹² Este conjunto não é mínimo, mas é útil e claro.

dade $y = a \vee b|c$. Há 4 plausibilidades que serão interessantes para esta análise:

$$x_1 = a|c, x_2 = b|c, x_3 = b|ac, x_4 = a|bc. \quad (1.1)$$

É importante notar que todas estas plausibilidades são condicionadas a c , a informação que por hipótese é suposta verdadeira. Além disso podem ser condicionadas a outras asserções relevantes e as únicas disponíveis são a e b por separado. Não tem sentido considerar ab como parte do condicionante. Deve haver uma dependência entre $a \vee b|c$ e algum subconjunto de $\{x_i\} = \{x_1, x_2, x_3, x_4\}$, então

- DP_4 : Regra da Soma: Deve existir uma função F que relaciona $a \vee b|c$ e algum subconjunto de $\{x_i\}$ e não deve tomar um valor constante, independente dos valores de $\{x_i\}$.

É claro que trocando soma por produto parece razoável desejar:

- DP_5 : Regra do Produto. Deve existir uma função G que relaciona $ab|c$ e algum subconjunto de $\{x_i\}$ e não deve tomar um valor constante, independente dos valores de $\{x_i\}$.

Como F e G representam a plausibilidade de asserções (compostas), também devem tomar valores reais. Além disso não impomos nada além de que dependam em algumas, se não todas, as variáveis $\{x_i\}$. Parece natural exigir que não tenham valores constantes, pois senão a todas as asserções compostas lhes seria atribuído o mesmo número. Para facilitar as deduções também imporemos diferenciabilidade até segunda ordem com respeito a quaisquer dois argumentos. Isto não é necessário, mas as provas ficam mais longas e no fim o resultado vem na forma de funções diferenciáveis.

Porque um subconjunto? Qual subconjunto? Todos? Como decidir? Há 11 subconjuntos de dois ou mais membros: Seis $\binom{4!}{2!2!}$ pares (x_i, x_j) , quatro $\binom{4!}{3!1!}$ triplas (x_i, x_j, x_k) e o conjunto inteiro (x_1, x_2, x_3, x_4) . Analisaremos casos particulares em que é fácil ver que alguns subconjuntos levam a resultados absurdos. Do ponto de vista axiomático poderíamos adicionar estes casos particulares à lista de desejos.

1.3 Consequências da Lista de Desejos

Parece difícil que desta lista $DP_1 \dots DP_5$ surja uma estrutura matemática, quanto mais única. Ou como veremos, essencialmente única a menos de regradações montônicas que não alteram a ordem das crenças. Talvez o que será surpreendente para o leitor, é que seja a teoria de probabilidades. A estrutura matemática aparecerá analisando as restrições nas funções F e G impostas pelos desejos.

1.3.1 A regra da soma

Começamos com a disjunção $a \vee b|c$ e a função F . Primeiro consideramos a e b mutuamente exclusivos, mas depois veremos que isto

permitirá analisar o caso geral. Sob esta restrição $a|bc = b|ac = v_f$ para qualquer c por DP_3 . Logo

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

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

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

Para avançar olhamos para asserções compostas mais complexas, que podem ser analisadas de mais de uma maneira, que pelo desejo DP_2 , devem dar o mesmo resultado. Para três asserções a, b e c mutuamente excludentes nas condições d , duas maneiras equivalentes de escrever a disjunção das três são $(a \vee b) \vee c|d = a \vee (b \vee c)|d$ o que permite usar a função f duas vezes

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

ou em notação óbvia, f satisfaz

$$f(f(x, y), z) = f(x, f(y, z)) \tag{1.2}$$

chamada equação da associatividade, primeiramente estudada por Abel no contexto de teoria de grupos. Pode se provar ¹³ que para toda solução de 1.2, existe um bijeção ϕ , dos reais nos reais, que tomaremos como crescente, e portanto será estritamente monotônica crescente, tal que

$$f(x, y) = \phi^{-1}(\phi(x) + \phi(y)). \tag{1.3}$$

Para o leitor bastará mostrar neste ponto, que a expressão 1.3 é uma solução da equação 1.2.

Agora um ponto central: podemos *regraduar*, usando ϕ , as atribuições de plausibilidade e não mais falar dos números do tipo $a|d$ mas de números $\phi(a|d)$. Por 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, o ordenamento da representação numérica das crenças é o mesmo. É importante ver que a função ϕ é estritamente monotônica: se $x > y$ segue que $\phi(x) > \phi(y)$, sem poder haver igualdade. Isto significa que asserções com crenças diferentes são mapeadas em valores ϕ diferentes. Caso ocorresse a possibilidade de igualdade, antes da regradação teríamos uma separação de preferências e depois da regradação poderíamos ter confusão entre asserções mapeadas no mesmo valor de ϕ . ¹⁴ Continuamos sem saber que números são esses, mas avançamos a ponto de poder dizer que para quaisquer eventos mutuamente exclusivos a crença da disjunção, uma asserção composta pode ser expressa em termos das crenças nas asserções mais simples:

$$\phi(a \vee b|d) = \phi(a|d) + \phi(b|d). \tag{1.4}$$

¹³ Para condições em f ver *Aequationes mathematicae* 1989, Volume 37, Issue 2-3, pp 306-312 *The associativity equation revisited* R. Craigen, Z. Páles, ou o livro Aczél, J. (1966), *Lectures on functional equations and their applications, Mathematics in Science and Engineering* 19, New York: Academic Press,

¹⁴ Veja A. Patriota onde as condições sobre f são relaxadas e as consequências de aceitar soluções não estritamente monotônicas são consideradas.

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

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

$$\phi(b|\bar{a}) = \phi(a|\bar{a}) + \phi(b|\bar{a}) \quad (1.6)$$

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 regrado 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

$$a = (a \wedge b) \vee (a \wedge \bar{b}) \quad \text{e} \quad b = (b \wedge a) \vee (b \wedge \bar{a}). \quad (1.8)$$

O leitor deve mostrar que as relações acima são verdadeiras, no estilo da tabela 1. 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 1.4, que descreve a soma de asserções mutuamente exclusivas, pode ser usada, levando a

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

onde, na última linha adicionamos e subtraímos o mesmo número. Chamamos pela ordem os termos do lado direito da equação acima de 1, 2,...5. Usando novamente a equação 1.4 para asserções mutuamente exclusivas, juntando 1 com 2, e 3 com 4:

$$\begin{aligned} \phi(a \vee b|d) &= \phi((a \wedge b) \vee (a \wedge \bar{b})|d) + \phi((b \wedge \bar{a}) \vee (a \wedge b)|d) - \phi(a \wedge b|d) \\ &= \phi(a|d) + \phi(b|d) - \phi(a \wedge b|d), \end{aligned} \quad (1.9)$$

que segue das relações da equação 1.8. Temos um dos resultados principais para lidar com asserções compostas por somas de asserções

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

Mas ainda não acabamos pois não sabemos o que fazer com $\phi(ab|d)$, que olharemos a seguir.

1.3.2 Regra do produto: quais as variáveis relevantes?

Queremos expressar

$$y = \phi(ab|c) \quad (1.10)$$

em termos da função ainda por determinar G e de algum dos subconjuntos de $\{x_i\}$. Lembramos a notação $x_1 = a|c$, $x_2 = b|c$, $x_3 = b|ac$, $x_4 = a|bc$. Tribus sugeriu a análise das 11 possibilidades para verificar que só há duas que sobrevivem a casos extremos. Seguimos A. Caticha, pois corrige vários erros anteriores. Os dois conjuntos sobreviventes 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. Cox já parte da conclusão de que estes dois subconjuntos são os adequados. O exercício que segue mostra que ele tinha razão, mas retira a arbitrariedade aparente, de fazer a escolha sem analisar outros candidatos.

Vejamos como chegar a esta conclusão (novamente seguimos AC)

Os 11 casos são reduzidos a 7 por simetria:

1. $y = G(\phi(a|I), \phi(b|I))$ (1 possibilidade)
2. $y = G(\phi(a|I), \phi(a|bI))$ (2 possibilidades $a \leftrightarrow b$)
3. $y = G(\phi(a|I), \phi(b|aI))$ (2 possibilidades $a \leftrightarrow b$)
4. $y = G(\phi(a|bI), \phi(b|aI))$ (1 possibilidade)
5. $y = G(\phi(a|I), \phi(b|I), \phi(a|bI))$ (2 possibilidades $a \leftrightarrow b$)
6. $y = G(\phi(a|I), \phi(a|bI), \phi(b|aI))$ (2 possibilidades $a \leftrightarrow b$)
7. $y = G(\phi(a|I), \phi(b|I), \phi(a|bI), \phi(b|aI))$ (1 possibilidade)

Caso 1 Mostraremos que

$y = a \wedge b|I = G(\phi(a|I), \phi(b|c)) = G(x_1, x_2)$ não funciona pois não satisfaz o esperado em um caso simples. Porque não serve o subconjunto mais óbvio (x_1, x_2) ? Primeiro vejamos que não segue o bom senso. 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 seja muito plausível que use um tenis de cada cor ao mesmo tempo. 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 e portanto teria um valor constante, independente das asserções sob consideração. Insistindo, suponha que Bruno joga uma moeda contra o teto, bate no ventilador e cai. A Helena pega outra moeda e faz o mesmo. Temos a mesma crença que saia cara ou coroa nas duas situações. Chamamos c_B a asserção que saiu cara no primeiro experimento e c_H no segundo. Achamos razoável escrever

$$\phi(c_B|I) = \phi(c_H|I) \quad \text{e} \quad \phi(\bar{c}_B|I) = \phi(\bar{c}_H|I)$$

E também achamos impossível que $c_B\bar{c}_B|I$ seja verdade, não pode ser verdade que Bruno obteve cara e coroa nessa única jogada. Mas seríamos levados a pensar que

$$\begin{aligned} \phi(c_B\bar{c}_H|I) &= G(\phi(c_B|I), \phi(\bar{c}_H|I)) \\ &= G(\phi(c_B|I), \phi(\bar{c}_B|I)) = \phi(c_B\bar{c}_B|I) = 0 \end{aligned} \quad (1.11)$$

que significaria que se Bruno obteve cara, Helena não poderia ter obtido coroa.

Caso 2

Para qualquer asserção $b|I$, sob quaisquer condições teríamos

$$\phi(b|I) = \phi(Ib|I) = G(\phi(I|I), \phi(I|bI)) = G(\phi_V, \phi_V) = \text{constante.}$$

Um método que atribui o mesmo valor numérico a todas as asserções não pode ser aceitável.

Caso 3 Para o caso $y = G(\phi(a|I), \phi(b|aI))$ e a alternativa $G(\phi(b|I), \phi(a|bI))$ 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. Não analisaremos as consequências disto agora. Ainda falta eliminar os outros candidatos e posteriormente encontrar a forma específica de G .

Caso 4 Se $y = G(\phi(a|bI), \phi(b|aI))$ somos levados a algo inaceitável considerando que para qualquer asserção b teríamos

$$\phi(b|I) = \phi(bb|I) = G(\phi(b|bI), \phi(b|bI)) = G(\phi_v, \phi_v) = \text{constante}$$

independente de b . Novamente a crença sobre a plausibilidade de uma asserção seria independente da asserção.

Caso 5 $y = G(\phi(a|I), \phi(b|I), \phi(a|bI))$. Este caso é mais complicado de analisar. Mostraremos, no entanto que se reduz a algum dos casos anteriores. Ainda consideraremos a conjunção de mais de duas asserções, $abc|I$, que pode ser escrito de duas formas diferentes $(ab)c|I = a(bc)|I$, portanto, considerando a primeira forma obtemos

$$\begin{aligned} \phi((ab)c|I) &= G(\phi(ab|I), \phi(c|I), \phi(ab|cI)) \\ &= G(G(\phi(a|I), \phi(b|I), \phi(a|bI)), \phi(c|I), G(\phi(a|cI), \phi(b|cI), \phi(a|bcI))) \\ &= G(G(x, y, z), u, G(v, w, s)). \end{aligned} \quad (1.12)$$

Para a segunda, com as mesmas definições das variáveis x, y, \dots , obtemos

$$\begin{aligned}\phi(a|bc|I) &= G(\phi(a|I), \phi(bc|I), \phi(a|bcI)) \\ &= G(\phi(a|I), G(\phi(b|I), \phi(c|I), \phi(b|cI)), \phi(a|bcI)) \\ &= G(x, G(y, u, w), s)\end{aligned}\quad (1.13)$$

Notamos as duas maneiras de escrever a mesma coisa. Repetimos que por DP_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 $z = \phi(a|bI)$ nem de $v = \phi(a|cI)$ explicitamente. Para que essa expressão não dependa de z nem v , podemos impor que G não dependa do terceiro argumento o que levaria a eliminar o que foi riscado na equação abaixo:

$$G(G(x, y, z), u, \cancel{G(v, w, s)}) = G(x, G(y, u, \cancel{w}), s)$$

levando a que G tem só dois argumentos e uma expressão sem z nem v :

$$G(G(x, y), u) = G(x, G(y, u))$$

e portanto somem todas as variáveis exceto x, y e u . Lembrando suas definições

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

que equivale ao **Caso 1** e portanto já foi eliminado.

Mas também podemos dizer que não depende do primeiro argumento, que também elimina z e v :

$$G(\cancel{G(x, y, z)}, u, G(\cancel{v}, w, s)) = G(\cancel{x}, G(y, u, w), s)$$

que leva à expressão

$$G(u, G(w, s)) = G(G(u, w), s)$$

que voltando às variáveis originais toma a forma

$$G(\phi(c|I), G(\phi(b|cI), \phi(a|bcI))) = G(G(\phi(c|I), \phi(b|cI)), \phi(a|bcI))$$

e mostra ser equivalente ao que teríamos obtido se partíssemos do **Caso 3** e portanto aceitável.

Fica como exercício mostrar que

1. o **Caso 6** pode ser reduzido ao **Caso 2**, ao **Caso 3** ou ao **Caso 4**
2. o **Caso 7** pode ser reduzido aos **Caso 5** ou **Caso 6**.

Concluimos portanto que

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

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 1.3.2 refletem a única opção. Outra queixa e que introduzimos casos simples onde os casos diferentes do 3 se mostraram contrários ao bom senso. Isto significa que o DP_5 é mais complexo do que parecia inicialmente.

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 , com a vantagem que pelo menos sabemos seus argumentos.

1.3.3 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:

$$a(b \vee c) = (ab) \vee (ac). \quad (1.14)$$

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

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

$$\begin{aligned} \phi((ab) \vee (ac)|d) &= \phi(ab|d) + \phi(ac|d) \\ &= G(\phi(a|d), \phi(b|ad)) + G(\phi(a|d), \phi(c|ad)) \end{aligned} \quad (1.16)$$

onde a equação 1.15 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 1.16 mostra o resultado de considerar a soma $(ab) \vee (ac)$. Mas devido à equação 1.14 e DP_2 , estas duas formas devem dar o mesmo resultado:

$$G(x, y + z) = G(x, y) + G(x, z). \quad (1.17)$$

Para obter a solução geral desta equação notemos que o primeiro argumento é o mesmo nos três termos, é portanto um parâmetro

que podemos manter fixo em qualquer valor arbitrário. Não é necessário supor diferenciabilidade, mas requerindo que G seja duas vezes diferenciável, e definindo $w = y + z$ obtemos a equação diferencial

$$\frac{\partial^2 G(x, w)}{\partial w^2} = 0 \quad (1.18)$$

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 1.17 obtemos

$$A(x)(y + z) + B(x) = A(x)y + B(x) + A(x)z + B(x), \quad (1.19)$$

portanto $B(x) = 0$, ou seja $G(x, w) = A(x)w = G(x, 1)w$ ¹⁵. Agora olhamos para $a|d$ e usamos $a|d = ad|d$ para a e d quaisquer.

$$\begin{aligned} \phi(a|d) &= \phi(ad|d) = G(\phi(a|d), \phi(d|ad)) \\ &= G(\phi(a|d), \phi_v) = A(\phi(a|d))\phi_v \end{aligned} \quad (1.20)$$

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

$$G(x, w) = \frac{xw}{\phi_v} \quad (1.21)$$

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

$$\phi(ae|d) = \frac{\phi(a|d)\phi(e|ad)}{\phi_v}. \quad (1.22)$$

Mas resta um problema: e se retirarmos a restrição de b e c mutuamente exclusivos? Precisamos usar a equação 1.9 para obter:

$$\phi(a \vee b|d) = \phi(a|d) + \phi(b|d) - \phi(ab|d) \quad (1.23)$$

$$\begin{aligned} \phi(a(b \vee c)|d) &= G(\phi(a|d), \phi(b \vee c|ad)) \\ &= G(\phi(a|d), \phi(b|ad) + \phi(c|ad) - \phi(bc|ad)) \end{aligned} \quad (1.24)$$

$$\begin{aligned} \phi((ab) \vee (ac)|d) &= \phi(ab|d) + \phi(ac|d) - \phi(abc|d) \\ &= G(\phi(a|d), \phi(b|ad)) + G(\phi(a|d), \phi(c|ad)) - G(\phi(a|d), \phi(bc|ad)) \\ &= G(\phi(a|d), \phi(b|ad)) + G(\phi(a|d), \phi(c|ad)) - G(\phi(a|d), G(\phi(b|ad), \phi(c|abd))) \end{aligned} \quad (1.25)$$

igualando os lados direitos das equações 1.24 e 1.25, obtemos uma nova equação funcional. Chamando $x = \phi(a|d)$, $y = \phi(b|ad)$, $z = \phi(c|ad)$, $w = \phi(c|abd)$ para simplificar, obtemos

$$G(x, y + z - G(y, w)) = G(x, y) + G(x, z) - G(x, G(x, w))$$

e é simples mostrar que a forma produto (eq. 1.21) é solução da equação funcional. É mais difícil mostrar que esta é a única forma se G for diferenciável duas vezes em cada argumento. O leitor interessado deverá consultar Áczel.

¹⁵ Suponha a equação $h(x + y) = h(x) + h(y)$, para qualquer x, y . Em particular, para $n \neq 0$ e m inteiros, vale que $h(nx) = h((n-1)x + x) = h((n-1)x) + h(x) = h((n-2)x + x) + h(x) = h((n-2)x) + 2h(x) = \dots = nh(x)$. Considere $x = x'/n$. Segue que $h(x') = nh(x'/n)$. Tome $x' = m$, portanto $h(x') = h(m) = mh(1) = nh(m/n)$. Logo $h(m/n) = (m/n)f(1)$. Basta supor continuidade que podemos passar dos racionais para os reais e obter $h(x) = xh(1)$.

Da equação 1.22, dividindo por ϕ_v obtemos

$$\frac{\phi(ae|d)}{\phi_v} = \frac{\phi(a|d)\phi(e|ad)}{\phi_v\phi_v} \quad (1.26)$$

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

$$p(a|b) = \frac{\phi(a|b)}{\phi_v} \quad (1.27)$$

que serão daqui a pouco chamados de probabilidade de a dado b . E a regra do produto em termos destes novos números regrados é

$$\boxed{p(ab|c) = p(b|c)p(a|bc) = p(a|c)p(b|ac)}$$

Temos uma regra para o produto e para soma lógicas de asserções. Como fica a negação? Apesar de não ter introduzido nada específico sobre ela veremos que com os desejos impostos podemos deduzir a plausibilidade regrada ou probabilidade da negação de uma asserção a partir da probabilidade de sua afirmação.

A regra do produto e a consistência permitem escrever

$$p(a|bc) = \frac{p(a|c)p(b|ac)}{p(b|c)} \quad (1.28)$$

que é chamado de Teorema de Bayes, mas que foi escrito pela primeira vez por Laplace. A contribuição de Bayes foi apontar a relação chamada de inversão

$$p(a|bc) \propto p(b|ac) \quad (1.29)$$

onde a probabilidade de uma asserção a condicionada a outra b é proporcional à probabilidade de b condicionada a a . Não podemos exagerar a importância desta afirmação que ficara clara à luz da variedade de aplicações tanto teóricas quanto experimentais que veremos adiante.

1.3.4 Negação

A lista de desejos inclui a menção de algo sobre a negação. A crença em asserções condicionadas à sua negação constituem casos de informação completa: $\phi(a|\bar{a}) = p(a|\bar{a}) = 0$. Também sabemos que $a \vee \bar{a}$ deve ser verdade, pois não resta alternativa. Portanto

$$\begin{aligned} \phi(a|\bar{a}d) &= p(a|\bar{a}d) = 0 \\ p(a \vee \bar{a}|d) &= \frac{\phi(a \vee \bar{a}|d)}{\phi_v} = 1 \end{aligned} \quad (1.30)$$

$$\begin{aligned}
 1 &= p(a \vee \bar{a}|d) \\
 &= p(a|d) + p(\bar{a}|d) - p(a\bar{a}|d) \\
 &= p(a|d) + p(\bar{a}|d) - p(\bar{a}|d)p(a|\bar{a}d) \\
 &= p(a|d) + p(\bar{a}|d), \tag{1.31}
 \end{aligned}$$

$$p(\bar{a}|d) = 1 - p(a|d)$$

ou a soma das crenças regraduadas de uma asserção e da sua negação é 1. Essencialmente chegamos ao fim do começo.

1.3.5 Estrutura matemática sobrevivente

Em termos destes números, reescrevemos os resultados até aqui obtidos:

$p(a a)$	$= p_v = 1$	Certeza da veracidade
$p(a \bar{a})$	$= p_f = 0$	Certeza da falsidade
$p(a \vee b c)$	$= p(a c) + p(b c) - p(ab c)$	regra da soma
$p(ab c)$	$= p(a c)p(b ac)$	regra do produto
$p(ab c)$	$= p(b c)p(a bc)$	regra do produto
$p(\bar{a} d)$	$= 1 - p(a d)$	regra da negação

Tabela 1.2
Probabilidades

Não falaremos mais em números $a|b$, nem na sua regradação $\phi(a|b)$ mas somente na última transformação $p(a|b)$ que chamaremos a probabilidade de a dado b , ou a probabilidade de a condicionada à informação que b é verdadeira. O motivo disto é que ao longo de séculos estas regras foram destiladas pelo bom senso de vários matemáticos e cientistas. Por volta de 1930, Kolmogorov formalizou, sem incluir a regra do produto nem condicionantes, usando linguagem de teoria de medida ou integração, mas já eram conhecidas desde Laplace. O que não estava claro é porque essas e não outras. Está completa a identificação das crenças ou plausibilidade regraduadas 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 leitor, caso deseje usar outras regras para manipular informação deverá responder quais dos desejos considerados acima não é razoável e portanto ao ser evitado, justificar essas outras regras.

O que foi obtido vai ser comparado com os axiomas de Kolmogorov na próxima secção. 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, razoável mas *ad hoc*, a probabilidade condicional definida a partir da regra do produto e portanto colocando com a mão o teorema de Bayes, que Cox obteve como uma consequência direta da consistência em particular e dos outros membros da desiderata.

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.

1.4 Exercícios

Mostre, construindo a tabela verdade as seguintes propriedades da Álgebra Booleana a partir da tabela verdade para a soma, produto e negação

- Idempotência do produto $AA = A$
- Idempotência da soma $A + A = A$
- Comutatividade do produto $AB = BA$
- Comutatividade da soma $A + B = B + A$
- Associatividade do produto $A(BC) = (AB)C$
- Associatividade da soma $(A + B) + C = A + (B + C)$
- Distributividade $A(B + C) = AB + AC$
- Dualidade $C = AB \Rightarrow \bar{C} = \bar{A} + \bar{B}$ e
 $C = A + B \Rightarrow \bar{C} = \bar{A}\bar{B}$

Mostre que $(A + B)A = A$ e portanto $A + BC = (A + B)(A + C)$

1.5 Exercícios Propostos

- Mostre que a conjunção e a disjunção não formam um conjunto de operadores completo para a álgebra booleana. Por exemplo mostre que não há combinação de estes operadores que permitam obter a negação. Mas nos propusemos uma função F e uma G e obtivemos uma forma de lidar com a negação. Como isso é possível? A resposta será achada ao ver que o desejo DP_3 sobre informação completa introduz a noção de negação mas só parcialmente ao dizer que a e sua negação são mutuamente exclusivos e que $a|\bar{a} = v_f$ como o mesmo v_f para todo a . Outra forma de proceder poderia ser introduzir um desejo do tipo: Deve existir uma função H , desconhecida tal que $a|c = H(\bar{a}|c)$. Isto codifica o desejo de encontrar uma teoria em que conhecimento sobre a implica conhecimento sobre \bar{a} . Claro que nesta altura sabemos que $H(x) = 1 - x$. Tente deduzir a as consequências ao trocar disjunção F por negação H no Desiderata para lidar com informação incompleta.
- Mostre a relação da equação 1.8. Desenhe o diagrama de Venn.
- A equação 1.9 relaciona a crença da disjunção às crenças nas asserções primitivas, mas inclui a subtração da crença na conjunção. Desenhe o diagrama de Venn adequado a esta situação. Discuta a origem do term subtraído.
- Voltemos ao **Caso 5** e suponhamos que G seja diferenciável com respeito a qualquer argumento. As derivadas parciais com respeito a z ou v devem dar zero. Use a regra da cadeia para mostrar que

$$\begin{aligned} 0 &= \frac{\partial}{\partial z} G(G(x, y, z), u, G(v, w, s)) \\ &= \frac{\partial}{\partial r} G(r, u, G(v, w, s))_{r=G(x, y, z)} \frac{\partial}{\partial z} G(x, y, z) \quad (1.32) \end{aligned}$$

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 mostre que voltamos ao **Caso 3**. Se não depende do terceiro mostre que voltamos ao **Caso 1**.

- **2** Para a função G da regra do produto 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**.
- Mostre que a forma produto (eq. 1.21) é solução da equação funcional. Mostre que esta é a única forma se G for diferenciável duas vezes em cada argumento.

- Escreva a regra do produto $P(AB|I)$, da soma $P(A + B|I)$ e da negação de $A|I$, de A no contexto I em termos das Chances, percentagem e Logprob definidos abaixo. Mostre que cada uma dessas é uma transformação monotónica ϕ das probabilidades e portanto uma regradação possível da representação numérica das crenças.

1. Chances: Defina as chances (odds em inglês) como $O(A|I) = \frac{P(A|I)}{P(\bar{A})}$.
2. Percentagem é o que chamariamos a probabilidade se em lugar de estar confinada ao intervalo $[0, 1]$ estivesse no intervalo $[0, 100]$.
3. Logprob $L_P(A|I) = \log P(A|I)$.
4. Logit ou log-odds: $\text{Logit}(P(A|I)) = \log\left(\frac{P(A|I)}{P(\bar{A})}\right)$.
5. Exprob $\text{Exp}_P(A|I) = \exp P(A|I)$ (Essa acabei de inventar).
6. Sineprob $\text{Sen}_P(A|I) = \sin \frac{P(A|I)\pi}{2}$ (Posso continuar.)

Em algum caso as regras escritas em termos das regradações são mais simples do que a regradação que leva às probabilidades? ¹⁶

Exercício Problema de Linda 1. Amos Tversky and Daniel Kahneman colocaram a questão a seguir, chamada de Problema de Linda, sobre probabilidades. Considere as asserções a seguir:

- I : Linda tem 31 anos, é solteira, assertiva, e muito inteligente. Ela se formou em filosofia. Quando estudante, estava profundamente preocupada com questões de discriminação e justiça social, e também participou de manifestações anti-nucleares.
- A : Linda é bancária .
- B : Linda é bancária e participa do movimento feminista .

Responda rapidamente qual das duas asserções é mais provável?

Exercício Problema de Linda 2. Não continue lendo até ter respondido à pergunta anterior.

Responda após pensar. O problema é atribuir números a $P(A|I)$ e $P(B|I)$. Qual é maior? Responda usando a regra do produto e use o fato que qualquer probabilidade tem uma cota superior 1. Este problema também é chamado de Falácia da conjunção. ¹⁷ Introduza a asserção

- C : Linda é bancária e não participa do movimento feminista .

Qual seria o ordenamento das três probabilidades $P(A|I)$, $P(B|I)$ e $P(C|I)$? Procure alguém feminista e faça a pergunta, faça o mesmo com alguém machista. Divirta-se com a percepção que as pessoas são irracionais. O que você acha que as pessoas acham que respondem quando tem que ser rápidas? Note que muitas vezes ao

¹⁶ Não é verdade que neste caso "What's in a name? that which we call a rose By any other name would smell as sweet"

¹⁷ These long-term studies have provided our finest insight into "natural reasoning" and its curious departure from logical truth. Stephen Jay Gould, sobre Tversky and Kahneman

fazer uma pergunta, quem responde está respondendo a uma pergunta parecida mas não exatamente aquela demandada.

Exercício Problema de Linda 3. Mostre usando a regra do produto que $P(A|I) \geq P(B|I)$. Tente inferir o que as pessoas fazem quando acham que está certo que $P(A|I) \leq P(B|I)$. Encontre asserções $A'|I'$ e $B'|I'$ parecidas com $A|I$ e $B|I$ tal que seja razoável supor mais provável supor o ordenamento contrário.

Exercício

- I : O preço do petróleo cai a 10 dolares o barril
- A : A Rússia invade a Ucrânia
- B : A Rússia invade a Ucrânia e os Estados Unidos corta relações diplomáticas com a Rússia

Dado I qual é mais provável, A ou B ? Note que as pessoas que cometem o erro de Falácia da Conjunção agem aparentemente como se estivessem comparando $P(A|I)$ com $P(C|AI)$, onde $B = AC$. Se você fosse presidente, manteria como assessor em política internacional alguém que ache $A|I$ menos provável que $B|I$?

Exercício

- I : Sou estudante da USP;
- A : Não estudei probabilidades
- B : Não estudei probabilidades e cometo a Falácia da conjunção

Dado I qual é mais provável, A ou B ?

2

Outras definições de probabilidade

2.1 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. Já descreveremos alguns dos motivos que nos levam a achar a posição freqüentista, incompleta e até, como mostraremos abaixo, insuficiente e errada. Pelo contrário, os axiomas de Kolmogorov, que codificam o bom senso da área já existente no trabalho de Laplace, podem ser vistos como não antagonicos aos resultados obtidos no capítulo 1. Interessante ler Kolmogorov. Ele não tem outro objetivo que

"... colocar no seu lugar natural, entre as noções gerais de matemática moderna, os conceitos básicos da teoria de probabilidade - conceitos que até recentemente eram considerados bastante peculiares.

Esta tarefa não teria tido esperança de sucesso antes da introdução das teorias de medida e integração de Lebesgue..."

A. N. Kolmogorov

Ele está organizando uma área após ficar claro como fazê-lo graças ao trabalho de Lebesgue e também Fréchet e admite que este ponto de vista era comum entre certos matemáticos mas merecia uma exposição concisa e livre de algumas complicações desnecessárias. Kolmogorov começa por considerar E uma coleção de elementos A, B, C, \dots que são eventos elementares ² e em nossa discussão anterior chamamos de asserções. \mathcal{F} é o conjunto de subconjuntos de E . Um tal sistema de conjuntos é chamado um campo se a soma, produto, interseção de dois elementos quaisquer pertencem ao sistema. Os axiomas de Kolmogorov para a teoria de Probabilidades são

- AK1) \mathcal{F} é um campo de conjuntos fechado ante um número de uniões (disjunções) e interseções (conjunções) 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,

¹ Foundations of the Theory of Probability
<http://www.mathematik.com/Kolmogorov/index.html>

² Em física E é conhecido como espaço de fases

- 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 denotado por $P(A)$.
- AK₄) $P(E) = 1$
- AK₅) Se $A \cap B = \emptyset$, então $P(A \cup B) = P(A) + P(B)$

Vejam se estes axiomas estão de acordo com os resultados da seção anterior. Em primeiro lugar uma definição que não será necessária neste curso, a de σ -campo. É uma coleção de subconjuntos fechado ante um número contável de operações de conjunto, tais como disjunção, conjunção, negação. Esta noção só é necessária ao falar de conjuntos com infinitos elementos. Vimos que a coleção de asserções também permite tais operações. Portanto estamos lidando com o mesmo tipo de coleção de eventos que Kolmogorov.³ Um exemplo de um σ -campo é o conjunto de conjuntos abertos nos reais. Neste curso usaremos asserções do tipo: "A variável X tem valor no aberto $(x, x + dx)$ " e extensões a \mathbb{R}^N . A ideia de σ -campo é essencial na teoria de integração de Lebesgue e aparecerá em tratamentos matematicamente mais sofisticados de probabilidade. Neste curso não iremos além de integrais de Riemann e somas infinitas.

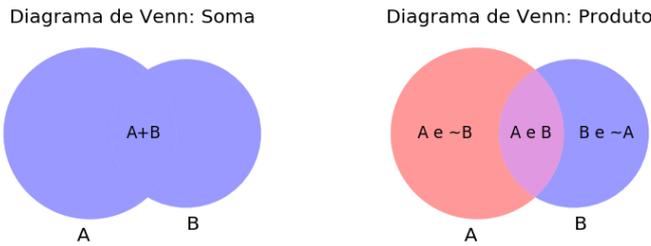
A probabilidade da *certeza* é 1 por AK₄; a probabilidade está entre zero e um e a probabilidade da disjunção de asserções que não tem elementos em comum é a soma das probabilidades. Notamos, na introdução aos axiomas no livro de Kolmogorov, porém a falta de uma regra para o produto. Kolmogorov não a introduziu inicialmente e só em trabalhos posteriores foi incluída por sugestão de Rao. No livro (página 6) ele introduz, como um adendo aos axiomas, 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, usando a comutação de A e B , portanto $P(AB) = P(BA)$ e a simetria ante troca $A \leftrightarrow B$.

Se uma vez estabelecidos os axiomas e dados valores numéricos para as probabilidades partirmos para as aplicações matemáticas, não haverá nenhuma diferença de resultados pois será a mesma estrutura matemática. Enfatizamos que as diferenças que temos são sobre a motivação dos axiomas e com a interpretação da idéia de probabilidades. Isso tem importância em inferência e portanto em aplicações. Em muitos livros o estudante encontrará uma diferença entre probabilidades e probabilidades condicionais. Deve ficar claro que no ponto de vista destas notas, não há probabilidade que não seja condicional.

³ Talvez a queixa é que as provas do capítulo 1 são para número finito de conjunções e disjunções. Isto porém não deve ser motivo de preocupação agora pois não é um empecilho irremovível.



2.1.1 Ainda outras definições de Probabilidade

Outra proposta de definição de probabilidades é a frequentista, que tem mais chances de ser a que o leitor já viu. A definição parece muito simples: é o limite da razão entre o número de vezes que um evento é verdade e o número de tentativas, quando este último vai para infinito.

Esta definição veio no esteio de uma colocada por Jacob Bernoulli e Laplace. Para eles é às vezes conveniente definir a teoria de chances pela redução de eventos do mesmo tipo a certo número de casos, igualmente possíveis e a

"...probabilidade, que é então simplesmente a fração cujo numerador é o número de casos favoráveis e cujo denominador é o número de todos os casos possíveis."⁴

O que significa "do mesmo tipo"? O físico verá aqui a uso da ideia de simetria. Se diferentes estados são tais que somos indiferentes ou incapazes de distingui-los então os colocamos na mesma categoria. Idéias de simetria são extremamente frutíferas. Mas quando não há simetria ou simplesmente não temos informação sobre ela é preciso estender a definição. Na época de Laplace as coisas não estavam muito claras, embora este tipo de regra seja útil e como veremos adiante não é uma regra nova a ser adicionada à *Desiderata* mas a ser deduzida do que já obtivemos. Além disso Laplace e Bernoulli deixaram claro em outros lugares que a probabilidade era uma manifestação numérica de crenças a partir de informação, portanto foram predecessores do exposto aqui. Considere, como Laplace há mais de duzentos anos, M_S a massa de Saturno. Ele fez asserções do tipo: "A probabilidade que $M_S < M_0$ ou $M_S > M_0 + \Delta m$ é menor que 10^{-4} ", que ele colocou em linguagem de apostas. Em linguagem atual é algo como $P(M_0 < M_S < M_0 + \Delta m | I) > 1 - 10^{-4}$. A informação de fundo condicionante I representa a teoria de Newton e os dados experimentais ⁵. Ele não está dizendo que a massa de Saturno é uma grandeza que apresenta variações e se for medida exatamente apresentará diferentes valores. Esqueça meteoritos, que poderiam mudar sua massa. Por exemplo, ao jogar um dado, se medirmos qual é o número de pontos na face que está para cima,

⁴ The theory of chances consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all te cases possible

A Philosophical Essay on Probabilities, Pierre Simon, Marquis de Laplace. 6a ed. F.W.Truscott e F.L. Emory trans.

⁵ A incerteza Δm que Laplace tem é da ordem de 1% de M_0 . O erro da estimativa de Laplace em relação ao valor estimado moderno é de aproximadamente 0.6%. Ou seja, ele teria ganho a aposta. O valor numérico $P(M_0 < M_S < M_0 + \Delta m | I)$ representa a crença que M_S esteja dentro do intervalo $(M_0, M_0 + \Delta m)$

este terá variações para diferentes jogadas. Alguns autores acham que só este tipo de variável merece ser descrito por probabilidades. Mas não a massa de Saturno à qual se atribui a propriedade de ter um valor *real*⁶. O que Laplace quer dizer é sobre o valor que atribuímos, com base nos dados e teoria, à crença que a massa está em um ou em outro intervalo. Quem acredita na definição de probabilidades como frequência, não pode falar da massa de Saturno em termos de probabilidade, pois não há um conjunto de Saturnos com diferentes massas. Falamos em lugar disto, da probabilidade de que o conjunto de medidas seja observado para o caso em que a massa seja M_0 . Em alguns casos isto dará o mesmo resultado, mas em outros não. Se você for acudado a definir a maior diferença entre alguém que define probabilidades através de frequências e quem as usa para expressar graus de crença, poderá responder de forma simplificada que este último não hesita em falar da distribuição de probabilidades de um parâmetro, como a massa de Saturno, enquanto o primeiro não admite tal linguagem⁷.

⁶ Real no sentido de ter existência independente do observador. Procure o significado de ontológico e de epistêmico.

⁷ Em linguagem mais técnica, ao espaço de parâmetros também é atribuído um σ -campo.

2.2 Algumas definições

Nesta altura podemos identificar os elementos formais principais para falar de probabilidades na linguagem de Kolmogorov. Primeiro é necessário deixar claro sobre o que se está falando:

- E a coleção de elementos A, B, C, \dots eventos elementares ou de asserções. Em alguns meios é chamado de espaço amostral.
- \mathcal{F} o campo: o sistema de conjuntos de asserções. Espaço de eventos
- $P(A)$ a atribuição de um número positivo a cada elemento de \mathcal{F} ;

Desta forma é costumeiro chamar a trinca (Espaço amostral, Espaço de eventos, Probabilidade de cada evento).

$$(E, \mathcal{F}, \mathcal{P})$$

de Espaço de probabilidades.

A apresentação do capítulo 1 não discorda disto, a não ser pelo ponto essencial que as probabilidades serão sempre condicionais e esquecer isso será a maior fonte de erros nas aplicações. Quando alguém se refere a uma probabilidade tipicamente tem em mente detalhes que se recusa a deixar explícitos pois esse exercício pode parecer cansativo. Outras vezes, e isso é mais perigoso, age como se tivesse em mente certos detalhes de informação, mas ao não perceber pode achar que não há alternativas. Além disso quando a teoria tem parâmetros, como será discutido em mais detalhes no próximo capítulo, queremos poder falar das probabilidades de que os parâmetros tenham valores em uma dada região. Isto não está em desacordo com a posição de quem adota os axiomas de

Kolmogorov. Basta aumentar o espaço amostral e o σ -campo e atribuir probabilidades aos elementos do novo campo. Isto porém não está de acordo com uma visão frequentista pois a massa de Saturno, ou qualquer outro parâmetro de uma teoria tem uma natureza ontológica que não lhes permite ser descrito em termos de frequência.

3

Uso elementar de Probabilidades

Este capítulo é muito mais simples que os anteriores, pois agora é uma questão de começar a desenvolver a estrutura matemática para poder lidar com aplicações simples.

3.1 Distribuições de probabilidade e densidade de probabilidade

A escolha das variáveis e identificação de suas características para descrever um problema é o passo mais importante em todo o processo que iremos descrever. Em geral estamos interessados em identificar antes de tudo os graus de liberdade relevantes do problema e o espaço em que essas variáveis vivem.

Agora introduziremos alguns exemplos de sistemas que permitirão justificar o interesse do estudante no desenvolvimento futuro da teoria:

Moeda: Vamos apostar num jogo que envolve jogar uma moeda a vários metros de altura e deixar cair no chão. Uma moeda é feita de níquel e ferro, tem propriedades magnéticas. No desenho na parte central, de um lado, aparece em relevo o 1 real e o ano que foi cunhada. Do outro, a imagem do rosto da República. Na parte externa um disco de bronze. A massa é aproximadamente 7g. A espessura 1.95 mm... Posso continuar dando informação irrelevante. Neste caso é fácil reconhecer que é irrelevante. O que você quer saber é que posso descrever o estado final por uma variável que toma um de dois valores: $s = +1$ ou $s = -1$. Uma asserção sobre a que podemos pensar é "A moeda caiu com a cara para cima". É claro que neste caso foi muito fácil identificar a irrelevância da maior parte do que foi dito, mas isso nem sempre é óbvio e devemos ter cuidado.

Radioatividade: Um contador Geiger detecta partículas ionizantes. A asserção sobre a qual não temos informação completa é: "O intervalo de tempo entre detecções T ", que pode tomar valores t , com $0 < t < \infty$. Ou no mesmo problema: "Qual é o número n de partículas detectadas num intervalo Δt ."

Partícula: As coordenadas de uma partícula $\mathbf{R} = (X, Y, Z)$ tomam valores $\mathbf{r} = (x, y, z)$ dentro de uma caixa cúbica de lado L . Assim e.g. $0 < x < L$. Podemos atribuir probabilidade a asserções

do tipo "A partícula tem coordenadas \mathbf{R} dentro uma caixa de volume dV centrada em \mathbf{r} ". Por preguiça diremos a mesma frase de forma simplificada "A partícula tem coordenadas $\mathbf{R} = \mathbf{r}$ ". Ou "A velocidade tem valores numa vizinhança de \mathbf{v} ", onde a vizinhança tem tamanho dado $v_x \in (v_x, v_x + \delta v_x)$, com expressões similares para a v_y e v_z . Mais sobre isto daqui a pouco.

Isto será interessante para descrever um gás de moléculas numa caixa:

Dois partículas. Na caixa descrita acima temos duas partículas idênticas mas distinguíveis. As coordenadas de cada uma são respectivamente \mathbf{R}_1 e \mathbf{R}_2 . O espaço de fases é o produto cartesiano dos dois espaços. Como exemplos de asserções em que podemos estar interessados: "(A partícula 1 tem coordenadas $\mathbf{R}_1 = \mathbf{r}_1$) e (A partícula 2 tem coordenadas $\mathbf{R}_2 = \mathbf{r}_2$). Note que ao falar de $P(\mathbf{R}_1 = \mathbf{r}_1 \text{ e } \mathbf{R}_2 = \mathbf{r}_2 | I)$ estamos falando do produto lógico das asserções individuais. Em geral e por preguiça a escreveremos $P(\mathbf{r}_1, \mathbf{r}_2 | I)$ ou simplesmente $P(\mathbf{r}_1, \mathbf{r}_2)$

N partículas. Igual que acima mas agora N partículas. Falaremos da probabilidade $P(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N | I)$. Isto e variações sobre tema serão os tópicos principais do curso de Mecânica Estatística. O significado de I é de extrema importância, pois as probabilidades dependerão de que tipo de partícula estamos falando, das suas interações e das condições experimentais do sistema. A influência das partículas vizinhas sobre a partícula 1 pode ser descrita por probabilidades $P(\mathbf{r}_1 | \mathbf{r}_2, \dots, \mathbf{r}_N, I)$.

Medida da carga do elétron: Um conjunto $D = (d_1, d_2, \dots, d_K)$ de medidas é feito no laboratório. A teoria nos fornece um modelo para a experiência que relaciona o parâmetro de interesse, neste caso a carga do elétron e , com a quantidade que medimos: $d = F(e)$. Mas sabemos que o dado d_i não é livre de erro de medida, ou seja não temos informação completa sobre d . Podemos então tentar codificar o que sabemos sobre d através de uma distribuição $P(d|D)$. Finalmente podemos falar sobre o conhecimento incompleto que temos sobre a carga e através de $P(e|D, I)$. Este tipo de análise é básico para a extração de informação a partir de medidas experimentais.

Cognição Um modelo de cognição de um animal pode ser feito considerando as variáveis relevantes. Os estados de neurônios de um sistema sensorial são descritos conjuntamente por uma variável X que toma valores x em algum espaço bastante complicado que não vem ao caso agora. Os estados de outras partes do cérebro são descritos por uma variável Z que toma valores z . O meio ambiente onde se encontra o animal é modelado por um conjunto de variáveis Y que tomam valores y , que certamente é um subconjunto das variáveis que poderiam ser usadas para descrever o *mundo lá fora*. O problema de cognição pode ser atacado considerando probabilidades $P(y|x, z, I)$. Neste caso I representa o conhecimento de Neurociência que tenhamos incluindo anatomia, dinâmica dos neurônios e dinâmica das sinápses. O mundo está em algum

estado, mas o modelo só pode atribuir probabilidades às diferentes possibilidades, pois o animal tem informação incompleta. Pense sobre a modelagem de ilusões visuais, onde algo *parece mas não é verdade*. Substitua a palavra animal por máquina nesta modelagem e teremos a possibilidade de descrever modelos artificiais de cognição que são básicos na área de aprendizagem de máquinas (*machine learning*).

Agentes Econômicos e Sociais: Daremos alguns exemplos no decorrer das aulas, mas é interessante notar que o uso de estatística em ciências sociais precede o seu uso em física.

Esportes Um jogador de basquete arremessa com uma probabilidade $P(C|I)$ de converter uma cesta. Há dias em que tem uma mão quente?

Como vemos, tanto o teórico quanto o experimental poderão usar as ferramentas da teoria de probabilidades para tratar situações de informação incompleta.

Continue, olhe em volta e identifique sistemas que possam ser interessantes e descreva as variáveis de interesse. Exemplos: Um dado cúbico, jogo de Bingo, condições de vida em um planeta, epidemia de Zika, bolsa de valores, uma amostra de ferro, e muito mais.

A partir de agora introduziremos alguns resultados matemáticos que serão úteis no desenrolar do curso.

3.2 Tipos de Variáveis

3.2.1 Variáveis discretas

Uma variável S toma valores no conjunto $E = (s_1, s_2, \dots, s_N)$. Por exemplo para um dado de cúbico $E_{\text{dado}} = (1, 2, 3, 4, 5, 6)$. Mas pode ser muito mais rico que isto. As asserções que faremos serão do tipo $A_i = "S \text{ vale } s_i"$. Ou talvez $B_{13} = "S \text{ toma valores no conjunto } (s_1, s_3)."$

Por preguiça, ou melhor para simplificar a notação, confundiremos as notações de tal forma que sob condições I a probabilidade $P(A_i|I)$ pode ser escrita simplesmente por $P(s_i|I)$. Ainda cometeremos a notação $P(s_i)$ sem especificar que há um condicionante I , talvez tacitamente suposto presente, mas às vezes esquecido de forma a levar a confusão e até a erros grosseiros. I será chamado de informação de fundo e envolve tudo o que sabemos sobre o problema. Chamaremos o conjunto de valores $P(s_i|I)$ de distribuição de probabilidades da variável S .

As asserções A_i são mutuamente exclusivas se o valor de S não pode ter simultaneamente dois valores quaisquer. Neste caso $A_i \wedge A_j = \emptyset$, para $i \neq j$ e portanto $P(A_i \wedge A_j|I) = 0$. Também são exaustivos de forma que não há possibilidade de que S tenha valores fora desse conjunto. Assim temos que

$$A_1 \vee A_2 \vee \dots \vee A_N = E$$

e temos certeza que E é verdadeiro. Segue que

$$\begin{aligned} 1 &= P(A_1 \vee A_2 \vee \dots \vee A_N | I) \\ 1 &= \sum_{i=1}^N P(A_i | I). \end{aligned} \quad (3.1)$$

Esta última expressão indica que a soma sobre todas os valores possíveis de S é um e será satisfeita por toda distribuição de probabilidades. É chamada condição de *normalização*.

3.2.2 *Váriáveis reais: densidades de probabilidades*

Em particular estamos interessados em grandezas físicas descritas por variáveis que tomam valores em intervalos dos reais, que chamaremos L .

No que segue lidaremos com asserções do tipo “a variável X toma valores entre x e $x + dx$ ”. Não sabemos ainda como, mas suponha que atribuímos um número a esta probabilidade. Como seria se lidássemos com a probabilidade de “ X toma valor x ”? Escolha um número entre 0 e 1. Se todos os números forem igualmente prováveis, a probabilidade de cada um deles seria zero, pois a soma deve dar um. Vemos que rapidamente chegamos a bobagens. Em geral e porque ainda não temos a matemática para lidar como esse tipo de problema, iremos falar somente de probabilidade de intervalos. Isso nos permite introduzir a densidade $P(x|I)$ tal que a probabilidade de que “a variável X toma valores entre x e $x + dx$ ” é dada por $P(x|I)dx$. $P(x|I)$ não é uma probabilidade mas é chamada de densidade de probabilidade¹. Teremos então que

- $P(x|I) \geq 0$
- $\int_L P(x|I)dx = 1$

Aqui reconhecemos a generalização da condição de normalização da equação 3.1, pois o intervalo L engloba todas as possibilidades de valores de X . Mas para qualquer intervalo $D : \{x|x \in [x_1, x_2]\}$, a probabilidade de X estar em D ou $x_1 \leq x \leq x_2$ é

$$P(x \in D | I) = \int_D P(x|I)dx$$

3.2.3 *Distribuição cumulativa de probabilidade*

Se uma variável X toma valores x no eixo real, e é descrita por uma densidade $P(x|I)$, a distribuição cumulativa é definida por

$$\Phi(x|I) = \int_{-\infty}^x P(x'|I)dx'. \quad (3.2)$$

segue que $\Phi(x|I)$ é a probabilidade de X tomar valores menores que x e a densidade de probabilidade é

$$P(x|I) = \frac{d}{dx}\Phi(x) \quad (3.3)$$

¹ Usamos a letra P por motivos históricos e eventualmente a chamaremos de probabilidade, por preguiça. Também esqueceremos de apontar os condicionantes e escreveremos muitas vezes simplesmente $P(x)$.

3.3 Caracterização de distribuições e densidades de Probabilidade

A informação disponível ao falar de X será equivalente à densidade de probabilidade para todo x . Mas isto talvez seja muito. É comum que seja necessário caracterizar, pelo menos parcialmente, o valor de X com um número, isto é um estimador ou estimativa de X . Há várias possibilidades e cada uma tem utilidade

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

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

A moda é o valor mais provável. Não quer dizer que se fizermos uma medida de X o obteremos, mas é o valor que terá mais probabilidade de ser encontrado. Podem haver vários valores que satisfazem o critério. A média leva em consideração todos os valores possíveis, cada um com voto proporcional à sua probabilidade. A mediana é o valor tal que a probabilidade de ser menor ou maior é igual. Cada uma é útil ou não em diferentes circunstâncias. Veja os exercícios. Cada uma resume a informação de forma a contar uma história. Devemos ter cuidado pois o contador da história pode ter um motivo para contar a história de forma resumida da maneira que é mais ou menos favorável a uma idéia que quer ver defendida. Podemos pensar em outras formas generalizando as idéias acima.

O valor esperado ou esperança de uma função $f(x)$, denotado por $E_x(f)$ ou $E(f)$, ou ainda alternativamente por $\langle f(x) \rangle$, é definido por

$$E_x(f) = \langle f(x) \rangle = \int_L f(x)P(x)dx \quad (3.4)$$

Usaremos tanto a notação $E_x(f)$ ou $E(f)$, preferida em textos de Matemática quanto $\langle f(x) \rangle$ mais usada em textos de Física. A notação que usamos de alguma forma deixa esquecida a idéia que a probabilidade depende da informação disponível. Quando for necessário deixar explícita a informação condicionante usaremos $E_x(f|C)$ ou $\langle f(x) \rangle|_C$.²

Pode 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 idéia de flutuação não seja útil, só porque por construção a sua média é nula. A média do seu quadrado é muito útil e recebe o nome de *variância*:

$$\text{Var}(X) := E((x - E(x))^2) = \langle (x - \langle x \rangle)^2 \rangle. \quad (3.5)$$

Obviamente $\text{Var}(X) \geq 0$. É fácil mostrar que $\text{Var}(X) = \langle x^2 \rangle - \langle x \rangle^2$. Algumas vezes nos referiremos à variância por σ_x^2 , por preguiça que veremos justificada algumas vezes, mas outras não.

² Usaremos esta notação às vezes, pois usaremos o direito de ser inconsistentes na notação, esperando que isso não confunda o leitor, mas o torne imune às várias notações na literatura. Isso

O valor esperado será muito usado no que segue, podemos generalizar a ideia e introduzir os momentos de uma distribuição:

- $m_n := \langle x^n \rangle = E[x^n] = \int_L x^n P(x) dx$

para valores inteiros de n (claro que caso a integral exista). Em notação mais carregada

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

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 sua média seja nula:

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

e note que $\text{Var}(X) = M_2$.

3.4 Marginais e Independência

As idéias de Marginalização e independência são de grande importância em toda a teoria e as aplicações que seguem.

Marginalização

Considere as asserções a, b, \bar{b}, c e os produtos $ab|c$ e $a\bar{b}|c$. Um resultado extremamente útil é

$$p(a|c) = p(ab|c) + p(a\bar{b}|c)$$

A prova é simples e a intuição também. Por exemplo $a =$ 'uma pessoa tem altura entre h e $h + \Delta h$ ', $b =$ 'uma pessoa tem peso maior que w '. Assim temos que a probabilidade de a , ter altura no intervalo é a soma das probabilidades de ter altura nesse intervalo e ter peso menor que w somada à probabilidade de ter altura nesse intervalo e ter peso maior ou igual a w .

A prova usa a regra do produto duas vezes, e a da negação uma:

$$\begin{aligned} p(ab|c) + p(a\bar{b}|c) &= p(a|c)p(b|ac) + p(a|c)p(\bar{b}|ac) \\ &= p(a|c) \left(p(b|ac) + p(\bar{b}|ac) \right) \\ &= p(a|c) \end{aligned} \quad (3.6)$$

Claro que se tivermos b que toma valores sobre um conjunto de asserções $\{b_i\}_{i=1, \dots, N}$ mutuamente exclusivas e exaustivo teremos

$$p(a|c) = \sum_{i=1}^N p(ab_i|c)$$

e dizemos ao marginalizar $p(ab|c)$ sobre a variável b obtemos a distribuição $p(a|c)$.

Voltando às alturas e pesos olhe uma tabela das probabilidades conjuntas onde cada entrada descreve o conhecimento para uma certa faixa de peso e de altura. Somamos as entradas para cada linha e as escrevemos na margem direita. Estas são simplesmente as

probabilidades para a faixa de altura sem levar em conta o peso. Essa é a origem do termo marginal, pois era anotado à margem da tabela conjunta quando o papel era o meio usado para aumentar a memória do usuário. Somando as entradas ao longo das colunas temos a probabilidade do peso independente de altura.

	w_1	w_2	...	w_N	$\sum_{i=1...N} P(h_j, w_i)$
h_1	$P(h_1, w_1)$	$P(h_1, w_2)$...	$P(h_1, w_N)$	$P(h_1)$
h_2	$P(h_2, w_1)$	$P(h_2, w_2)$...	$P(h_2, w_N)$	$P(h_2)$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
h_M	$P(h_M, w_1)$	$P(h_M, w_2)$...	$P(h_M, w_N)$	$P(h_M)$
$\sum_{j=1...M} P(h_j, w_i)$	$P(w_1)$	$P(w_2)$...	$P(w_N)$	

Tabela 1.3

As marginais são escritas na margem!

O outro conceito de extrema importância é o de

Independência

A regra do produto em geral

$$p(ab|c) = p(a|c)p(b|ac)$$

se reduz ao produto das marginais

$$p(ab|c) = p(a|c)p(b|c) \tag{3.7}$$

quando $p(b|ac)$ não depende de a . Se informação da veracidade de a não altera crenças sobre b , dizemos que nas condições de que c seja verdadeiro, b é independente de a . É óbvio que a independência é reflexiva, pois também podemos escrever

$$p(ab|c) = p(b|c)p(a|bc)$$

o que significa, dada a equação 6.35 que $p(a|bc) = p(a|c)$. Assim temos que distribuições conjuntas de variáveis independentes se reduzem a produtos. As interações físicas entre partículas serão descritas por distribuições que não se fatorizam nas marginais, i.e nas probabilidades das variáveis de cada partícula.

3.4.1 *Independência aos pares , mútua e condicional*

Suponha que tenhamos um conjunto de asserções sob consideração $S = \{A_1, A_2, \dots, A_K\}$. Dizemos que os A_i são independentes aos pares na condição C , se para todo i, j com $1 \leq i \leq K$ e $1 \leq j \leq K$ tivermos

$$P(A_i|A_jC) = P(A_i|C).$$

Dizemos que os membros de S são mutuamente independentes na condição C se

$$P(A_1, A_2, A_3 \dots A_K | C) = P(A_1 | C)P(A_2 | C) \dots P(A_K | C).$$

Mas é claro que em geral, a distribuição conjunta pode ser manipulada usando a regra do produto. Para $K = 3$, supondo independência aos pares

$$\begin{aligned} P(A_1, A_2, A_3 | C) &= P(A_1 | C)P(A_2 A_3 | A_1 C) = P(A_1 | C)P(A_2 | A_1 C)P(A_3 | A_1 A_2 C) \\ &= P(A_1 | C)P(A_2 | C)P(A_3 | A_1 A_2 C) \end{aligned} \quad (3.8)$$

e para a chegar ao produto $\prod_{i=1,2,3} P(A_i | C)$, deveríamos ainda impor que $P(A_3 | A_1 A_2 C) = P(A_3 | C)$ que é mais restritiva que independência aos pares. Mas isto é sutil e merece um exemplo específico para ficar mais claro.

Vamos imaginar uma moeda sendo jogada. A_1 : "cara na primeira jogada", A_2 : "cara na segunda jogada", A_3 : "as duas jogadas tiveram o mesmo resultado", que é equivalente a escrever $A_3 = A_1 A_2 + \overline{A_1} \overline{A_2}$. Dada a independência das duas jogadas temos

$$\begin{aligned} P(A_1 | A_2 C) &= P(A_1 | C), & P(A_2 | A_1 C) &= P(A_2 | C) \\ P(A_2 | A_3 C) &= P(A_2 | C), & P(A_3 | A_2 C) &= P(A_3 | C) \\ P(A_3 | A_1 C) &= P(A_3 | C), & P(A_1 | A_3 C) &= P(A_1 | C) \end{aligned} \quad (3.9)$$

mas

$$P(A_1 | A_2 A_3 C) = 1 \neq P(A_1 | C)$$

Completamos a definição de mutuamente independente se $P(A_i | B_i C) = P(A_i | C)$ onde B_i em um subconjunto qualquer de S que não inclua A_i .

Como toda probabilidade é condicional, a independência também depende do contexto. Podemos ter $P(X | Y Z_1) = P(X | Z_1)$ mas $P(X | Y Z_2) \neq P(X | Z_2)$. Por exemplo no caso das moedas Z_1 e Z_2 poderiam diferir nas condições iniciais do lançamento e.g. altura, energia, velocidade angular, etc. Vamos supor que X , Y e Z tomem valores reais. Se X e Y forem independentes na condição Z , ou seja $P(XY | Z) = P(X | Z)P(Y | Z)$, então a como funções dos valores destas variáveis teremos

$$P(XY | Z) = P(X = x, Y = y | Z = z) = P(x, y | z) \text{ deve satisfazer}$$

$$P(x, y | z) = f(x, z)g(y, z).$$

Por outro lado se $P(x, y | z) = f(x, z)g(y, z)$ é possível mostrar que X e Y são independentes na condição Z .

Para concluir o estudante deve notar que a idéia de independência não deve ser confundida com a de mutuamente exclusivo. Independência leva a que a regra do produto é $P(ab | c) = P(a | c)P(b | ac) = P(a | c)P(b | c)$. Mutuamente exclusivo implica em $P(a + b | c) = P(a | c) + P(b | c) - P(ab | c) = P(a | c) + P(b | c) - P(a | c)P(b | ac) = P(a | c) + P(b | c)$.

3.5 Exemplos de Famílias de Distribuições de probabilidade

No contexto deste curso, uma variável aleatória é simplesmente alguma variável para a qual não temos informação completa e portanto, o que soubermos será usado para construir uma distribuição de probabilidades. É comum que a distribuição seja escolhida dentro de uma família. Uma função de pelo menos duas variáveis $f(x; \Theta)$, não negativas e integráveis, pode ser considerada uma família paramétrica de funções de x com Θ como parâmetro. Tanto x quanto Θ podem ser multidimensionais. Apresentaremos a seguir exemplos de famílias onde x pode ser discreto ou contínuo, unidimensional ou multidimensional. Algumas famílias das distribuições aparecem de forma recorrente em muitas aplicações e vale a pena ter certa familiaridade. Podemos ter diferentes motivos que levem ao uso de uma família. Por exemplo, desde o mais simples como informação sobre o domínio de valores de uma variável, a motivos teóricos sobre a dependência entre as variáveis relevantes. Os motivos teóricos podem ser toda a área da Mecânica Estatística e as dependências terem relação com forças entre partículas.

O que segue não pode ser considerado uma exposição completa das propriedades das distribuições. Algumas, como a binomial e a gaussiana, serão tratadas com muito mais detalhe em capítulos posteriores. A notação usual em estatística ao dizer que a variável X tem distribuição do tipo Blablabla com parâmetros Θ é

$$X \sim Bla(\Theta)$$

usando algumas letras do nome da distribuição que pode ser o nome de alguma pessoa, indicando também os valores ou nomes dos parâmetros.

A utilidade varia de motivações teóricas que forcem um dado tipo de modelo a simplesmente a possibilidade de fazer algum avanço analítico. De qualquer forma é sempre útil ter um poste onde possamos procurar a chave perdida.

3.5.1 Bernoulli

Esta distribuição é uma das mais simples. Se uma variável está distribuída de acordo com a distribuição (ou equivalentemente é uma variável) de Bernoulli escrevemos $S \sim \text{Ber}(p)$. Neste caso, S tem dois valores possíveis. Por exemplo o espaço de valores possíveis de S é $E = \{-1, +1\}$ ou $\{\text{cara}, \text{coroa}\}$, ou $\{0, 1\}$. A distribuição de Bernoulli é em termos de um parâmetro p , $0 \leq p \leq 1$

$$P(S|p) = \begin{cases} p & \text{se } S = +1 \\ 1 - p & \text{se } S = -1. \end{cases}$$

Também pode ser escrita, usando o parâmetro $m = 2p - 1$ como

$$P(S|m) = \begin{cases} \frac{1+m}{2} & \text{se } S = +1 \\ \frac{1-m}{2} & \text{se } S = -1. \end{cases}$$

A interpretação de p é simples: é a probabilidade de S ter valor 1. Obviamente a de ter valor -1 é $1 - p$. O valor esperado de S é

$$\langle S \rangle = \sum_{s=-1,1} sP(S=s) = m = 2p - 1,$$

que dá a interpretação de m e o motivo por que é interessante usá-lo como parâmetro da distribuição. O segundo momento é simples pois $S^2 = 1$ portanto

$$\langle S^2 \rangle = 1$$

a para a variância σ_S temos

$$\sigma_S^2 = \langle S^2 \rangle - \langle S \rangle^2 = 1 - m^2 = 4p(1 - p)$$

Sob o risco de ser maçante introduzimos a variável $R = \frac{1+S}{2}$ e agora temos

$$P(R|p) = \begin{cases} p & \text{se } R = 1 \\ 1 - p & \text{se } R = 0. \end{cases}$$

portanto o valor esperado $\langle R \rangle = p$ e a variância $\sigma_R^2 = \langle R^2 \rangle - \langle R \rangle^2 = p(1 - p)$. Estas variáveis sozinhas podem parecer muito simples, mas ao juntar várias partículas cujos estados são descritos por variáveis deste tipo vamos poder modelar fenômenos bem complexos. Por exemplo S pode representar classicamente o spin de um íon numa rede cristalina ou R pode indicar a presença ou ausência de uma partícula num modelo do que se chama um gás de rede.

A variância, o valor esperado do quadrado da flutuação, vai a zero quando $p = 0$ ou $p = 1$ que são os casos em que a informação é completa: $S = -1$ sempre no primeiro caso e $S = 1$ sempre no segundo. A variância traz informação sobre a largura da distribuição e isso não se restringe a esta distribuição.

3.5.2 Uniforme

Uma variável $X \sim U(0, L)$ toma valores no intervalo do eixo real $\mathcal{L} : 0 < x < L$ e sua probabilidade é uma constante dentro do intervalo e zero fora:

$$P(X|L) = \begin{cases} \frac{1}{L} & \text{se } X \in \mathcal{L} \\ 0 & \text{se não.} \end{cases}$$

Os valores esperados e variância são

$$\begin{aligned} \langle X \rangle &= \int_{\mathcal{L}} xP(x)dx = \frac{L}{2} \\ \langle X^2 \rangle &= \int_{\mathcal{L}} x^2P(x)dx = \frac{L^2}{3} \\ \sigma_X &= \frac{L}{2\sqrt{3}} \end{aligned}$$

Obviamente podemos fazer translações $Y = aX + B$ e teremos $Y \sim U(B, aL + B)$ com probabilidade $1/aL$ dentro e 0 fora do intervalo.

3.5.3 Binomial

Uma variável de Bernoulli toma valores $s = +1$ ou $s = -1$ e é amostrada N vezes. Ou seja temos um conjunto de dados escritos como uma lista (s_1, s_2, \dots, s_N) . A variável binomial é m o número de vezes que aparece o $+1$ nessa lista. Assim $m \sim \text{Bin}(p; N)$. Obviamente a distribuição de Bernoulli é $\text{Ber}(p) = \text{Bin}(p; 1)$. Mostraremos no próximo capítulo que

$$\begin{aligned} P(m|pN) &= \binom{N}{m} p^m (1-p)^{N-m} \\ &= \frac{N!}{m!(N-m)!} p^m (1-p)^{N-m} \end{aligned} \quad (3.10)$$

Você encontrará frequentemente que isto é descrito como a probabilidade de n sucessos em N tentativas, quando p é a probabilidade de sucesso em cada tentativa. Voltaremos a falar desta distribuição várias vezes. Em particular fica faltando aqui definir independência entre cada uma das tentativas.

3.5.4 Binomial Negativa

Esta é uma variação sutil sobre o tema de Bernoulli com respeito à distribuição anterior. Se obter $s_1 = 1$ foi chamado de sucesso, então é natural chamar $s_i = -1$ de fracasso. Agora fixamos o número de fracassos k e pedimos a probabilidade do número de sucessos n até obter k fracassos.

$$P(n|pk) = \binom{n+k-1}{n} p^n (1-p)^k$$

Nas primeiras $n+k-1$ tentativas a ordem pode ser qualquer e o número destas seqüências é $\binom{n+k-1}{n}$. A última tentativa, a $n+k$ deve ser um fracasso. A média é $E(n) = pk/(1-p)$ e a variância $pk/(1-p)^2$

Para verificar que a normalização é correta precisamos alguns truques. Primeiro usamos a soma da progressão geométrica

$$\frac{1}{1-p} = \sum_{s=0}^{\infty} p^s = 1 + p + p^2 \dots + p^{k-1} + p^k + \dots$$

e a derivada de ordem k

$$\frac{d^k}{dp^k} \left(\frac{1}{1-p} \right) = \frac{(k-1)!}{(1-p)^k}$$

que elimina os primeiros $k-1$ termos da soma da PG. Deixamos os detalhes para o leitor.

3.5.5 Poisson

Para descrever a estatística de contagens de um detector é útil introduzir a distribuição de Poisson. Veremos adiante que esta distribuição está relacionada com a binomial. A probabilidade de n ,

número de contagens em um certo intervalo de tempo, dado o parâmetro λ que caracteriza o processo, é

$$P(n|\lambda) = \frac{\lambda^n}{n!} e^{-\lambda}$$

O valor médio

$$\begin{aligned} \langle n \rangle &= \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} \\ &= e^{-\lambda} \lambda \frac{d}{d\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \\ &= e^{-\lambda} \lambda \frac{de^\lambda}{d\lambda} \\ &= \lambda, \end{aligned} \quad (3.11)$$

e o segundo momento

$$\begin{aligned} \langle n^2 \rangle &= \sum_{n=0}^{\infty} n^2 \frac{\lambda^n}{n!} e^{-\lambda} \\ &= e^{-\lambda} \left(\lambda \frac{d}{d\lambda} \right) \left(\lambda \frac{d}{d\lambda} \right) e^\lambda \\ &= \lambda + \lambda^2. \end{aligned} \quad (3.12)$$

Portanto a variância

$$\sigma_{Poisson}^2 = \lambda \quad (3.13)$$

3.5.6 Beta

Uma variável X toma valores x no intervalo $0 \leq x \leq 1$ e tem dois parâmetros

$$P(x|a; b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \quad (3.14)$$

Note que se a e b forem números inteiros, podemos escrever

$$\begin{aligned} P(x|a; b) &= \frac{(a+b-1)!}{((a-1)!(b-1)!)} x^{a-1} (1-x)^{b-1} \\ P(x|n = a-1; N = b+m-1) &= \frac{(N+1)!}{n!(N-n)!} x^n (1-x)^{N-n} \end{aligned} \quad (3.15)$$

onde a parametrização da última linha mostra uma certa semelhança com a binomial. Uma pequena diferença é que em lugar de N temos $N+1$ no numerador. A diferença fundamental é que na binomial falamos da probabilidade de n e aqui de x . As duas distribuições estão relacionadas pelo resultado de Bayes: $P(n|x) \propto P(x|n)$. Voltaremos a falar nesta relação ao falar de distribuições conjugadas.

Para a normalização usamos um resultado devido a Euler

$$E_{N-n}^n = \int_0^1 p^n (1-p)^{N-n} dp = \frac{n!(N-n)!}{(N+1)!} \quad (3.16)$$

Suponha que $n < N - n$. Integramos por partes com $dv = (1 - p)^k dp$ e $u = p^r$ que leva a $v = -\frac{1}{k+1}(1 - p)^{k+1}$ e $du = rp^{r-1}$, assim

$$\begin{aligned} E_k^r &= \int_0^1 p^r (1 - p)^k dp \\ &= \frac{r}{k+1} \int_0^1 p^{r-1} (1 - p)^{k+1} dp \\ &= \frac{r}{k+1} E_{k+1}^{r-1}. \end{aligned}$$

Começando com $r = n$, Após n passos temos uma integral $\propto \int_0^1 (1 - p)^N dp = 1/(N + 1)$. Iterando

$$E_{N-n}^n = \frac{n}{N - n + 1} \frac{n - 1}{N - n + 2} \cdots \frac{n - (n - 1)}{N - n + n} \frac{1}{N + 1} \quad (3.17)$$

Multiplicando e dividindo por $(N - n)!$ obtemos o resultado 3.16. Se $n > N - n$, mude variáveis de integração $p \rightarrow 1 - p$ e proceda da mesma forma. Podemos calcular momentos da Beta da mesma forma, pois $E(p^r | \text{Beta}(n, N)) \propto E_{N-n}^{n+r}$

3.5.7 Gamma

O nome desta distribuição é devido a que a função Gama é definida pela integral

$$\Gamma(u) = \int_0^\infty e^{-t} t^{u-1} dt, \quad (3.18)$$

que voltaremos a ver várias vezes, em particular no capítulo ??.

Uma variável X toma valores x no intervalo $0 \leq x < \infty$ e tem dois parâmetros a , conhecido com o parâmetro de escala e b o parâmetro de forma:

$$P(x|a; b) = \frac{1}{a\Gamma(b)} \left(\frac{x}{a}\right)^{b-1} e^{-\frac{x}{a}} \quad (3.19)$$

O valor esperado $\langle x \rangle = E(x) = ab$ e a variância $E(x^2) - E(x)^2 = a^2 b$.

3.5.8 Gaussiana ou Normal

Dedicaremos o capítulo ?? ao estudo desta distribuição. Uma variável X toma valores x no intervalo $-\infty < x < \infty$ e tem dois parâmetros μ a média e σ^2 a variância:

$$P(x|\mu; \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3.20)$$

uma variável que tem esta distribuição é dita normal ou gaussiana e tipicamente na literatura estatística se escreve

$$X \sim \mathcal{N}(\mu, \sigma)$$

3.5.9 Distribuição Exponencial

X toma valores reais não negativos, $x \geq 0$. Um único parâmetro, $a > 0$ dá a escala e a variância. A distribuição é

$$P(x|a) = \frac{1}{a} e^{-\frac{x}{a}} \quad (3.21)$$

O valor médio e a variância são respectivamente

$$E(x) = a, \quad E(x^2) - E(x)^2 = a^2$$

3.5.10 Laplace

Semelhante à exponencial, mas com x podendo ser qualquer valor real, portanto também conhecida como dupla exponencial,

$$P(x|\mu a) = \frac{1}{2a} e^{-\frac{|x-\mu|}{a}} \quad (3.22)$$

onde μ é um parâmetro de localização e a de escala. Note o fator 2 para garantir a normalização.

3.5.11 Cauchy

a distribuição de Cauchy tem vários nomes associados, Lorentz, Cauchy-Lorentz, Breit-Wigner.

$$P(x|x_0, a) = \frac{1}{a\pi} \frac{1}{1 + \frac{(x-x_0)^2}{a^2}} \quad (3.23)$$

O valor médio não é definido da forma convencional, mas usando uma definição da integração em intervalos infinitos devida a Cauchy, o valor principal de Cauchy

$$E(x - x_0) = E(x) - x_0 = \mathcal{P} \int_{-\infty}^{\infty} (x - x_0) P(x|x_0, a) dx = \lim_{L \rightarrow \infty} \int_{-L}^L (x - x_0) P(x|x_0, a) dx = 0$$

por simetria, logo

$$E(x) = x_0,$$

que coincide com a moda e a mediana. O interessante é que

$E((x - x_0)^2) = \infty$ e portanto a variância é infinita.

Assintoticamente as contribuições para a integral vão como $x^2/x^2 \sim$ constante. Mas ainda podemos definir a largura a meia altura, que é $2a$, a a separação entre os pontos $x_0 + a$ e $x_0 - a$, onde a probabilidade é $1/2\pi a$.

3.6 Mudança de Variáveis

Ao analisar um sistema em física, o problema mais importante e imediato é o de identificar as variáveis relevantes para representar seus estados. Estudantes inexperientes podem achar que essa parte é fácil. O motivo é que foi dito que o espaço tem esta e aquela característica, que o tempo é esse parâmetro t que todos sabem o que é (menos eu). O que talvez não fique claro é que milhares de

anos de tentativas levaram a atribuir certos modelos matemáticos a sistemas físicos e ficam escondidas as várias tentativas que acabaram em becos sem saída, ou que se verificou posteriormente, podiam ser significativamente simplificados.

Suponha que voce tenha informação I sobre uma variável X que toma valores x reais e codifique esse conhecimento numa densidade de probabilidade $P(x|I)$. Por algum motivo, fica claro que seria útil introduzir Y que esta relacionada com X por uma função f conhecida

$$y = f(x).$$

A pergunta que se coloca é o que podemos dizer sobre a densidade de Y sob as mesmas condições de informação I ?

A resposta é fácil se pensarmos sobre o significado de densidade de probabilidade. Vamos começar com $f(x)$ uma função monotónica, que permite uma inversão $x = f^{-1}(y)$. Consideremos $y_i = f(x_i)$ para $i = 1, 2$ e f crescente. A asserção

"O valor de X toma valores x , tal que $x_1 < x < x_2$ "

deve ser equivalente à asserção

"O valor de Y toma valores y , tal que $y_1 < y < y_2$ "

Equivalente no sentido de que a mesma probabilidade deve ser atribuida a cada uma delas se o contexto for o mesmo

$$Prob(y_1 < y < y_2|I) = Prob(x_1 < x < x_2|I)$$

A relação entre as densidades de probabilidades deve ser

$$\int_{y_1}^{y_2} P(y|I)dy = \int_{x_1}^{x_2} P(x|I)dx$$

Se os intervalos de integração forem suficientemente pequenos podemos escrever

$$P(y|I)\Delta y = P(x|I)\Delta x$$

e no limite

$$P(y|I) = P(x|I) \frac{dx}{dy}$$

isto não é mais do que simplesmente tomar a derivada com respeito ao limite superior (no ponto $y_2 = y$) e usar a regra da cadeia. As regras de mudança de variáveis não são mais que as regras de mudança de variável na teoria de integração ou de medida. É difícil exagerar a importância deste resultado.

O leitor poderá agora estender os resultados para o caso em que f for decrescente. Agora $dx/dy = df^{-1}(y)/dy$ deverá ser substituida por $-dx/dy$. Também deve poder encontrar as regras quando f não for monotónica, ou ainda quando x e y forem generalizadas para mais dimensões.

Se a função $f(x)$ não for monotónica precisamos ter cuidado. Olhemos para um exemplo simples. Seja $U = X^2$, portanto um valor u de U está associado a um valor x de X por $u = x^2$. A

asserção que U é menor que um dado valor u , $U < u$ é idêntica à asserção que $-\sqrt{u} < X < \sqrt{u}$, portanto, em termos da cumulativa

$$\begin{aligned}\Phi(u|I) = \text{Prob}(U < u|I) &= \text{Prob}(-\sqrt{u} < X < \sqrt{u}|I) \\ &= \text{Prob}(X < \sqrt{u}|I) - \text{Prob}(X < -\sqrt{u}|I) \\ &= \text{Prob}(X < x|I) - \text{Prob}(X < -x|I) \\ &= \Phi(x|I) - \Phi(-x|I)\end{aligned}\quad (3.24)$$

derivando com respeito a u temos a densidade de probabilidade

$$\begin{aligned}P(u|I) &= \frac{d}{du}\text{Prob}(U < u|I) \\ &= \left(\frac{d}{dx}\text{Prob}(X < x|I) - \frac{d}{dx}\text{Prob}(X < -x|I)\right)\frac{dx}{du} \text{ onde } x = \sqrt{u} \\ &= (P(X = \sqrt{u}|I) + P(X = -\sqrt{u}|I))\frac{1}{2\sqrt{u}}.\end{aligned}\quad (3.25)$$

A transformação neste caso não é invertível e precisamos levar em conta os dois ramos da inversa, tanto $+\sqrt{u}$ quanto $-\sqrt{u}$.

A integração especialmente em espaços de alta dimensionalidade é uma das tarefas mais comuns nas aplicações e consumirá a maior parte dos esforços computacionais. No capítulo sobre integração Monte Carlo veremos como mudanças de variáveis serão elevadas a uma forma de arte.

3.7 Covariância e correlações

Introduziremos de forma rápida mas voltaremos a usar muitas vezes a idéia de correlações que é central nas aplicações. Duas variáveis X e Y tem distribuição conjunta $P(x, y|I)$ sob informação I . O valor esperado do produto é

$$E(xy) = \langle xy \rangle = \int xyP(xy|I)dxdy.$$

e o valor esperado do produto das variáveis truncado, isto é, subtraído o valor médio de cada variável, é a covariância

$$\begin{aligned}\text{Cov}_{xy} = E((x - E(x))(y - E(y))) &= \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle \\ &= \langle xy \rangle - \langle x \rangle \langle y \rangle\end{aligned}\quad (3.26)$$

que é o valor esperado do produto das flutuações em torno da média. Dadas as propriedades de X , o maior valor que a covariância pode ter é quando X e Y são iguais, pois o integral só tem contribuições positivas. Nesse caso $\text{Cov}_{xx} = \text{Var}(x)$. Isso sugere introduzir a correlação r , que aparentemente foi introduzida por Pearson

$$r = \frac{\text{Cov}_{xy}}{\sqrt{\text{Var}(x)\text{Var}(y)}}\quad (3.27)$$

e que satisfaz $-1 \leq r \leq 1$.

No caso de n variáveis $X_i, i = 1 \dots n$, a matriz de correlações C_{ij} tem elementos $C_{ij} = \text{Cov}_{x_i x_j}$. Quando i é um índice temporal o estudo das correlações temporais é de grande utilidade em Física para caracterizar a dinâmica de um sistema.

Exercício Pense no significado de cada um dos estimadores de X e da variância Var_X e proponha outros estimadores. Mostre casos em que a moda, a média e a mediana não são iguais.

Exercício Um herói luta com inimigos iguais e sempre com as mesmas armas. Cada luta é independente de todas as outras e o herói tem probabilidade $q = 1 - p$ de ganhar cada luta, Observamos que ele se aposenta após lutar N vezes, quando é derrotado pela n -ésima vez. Ou seja temos $n = N - k$ derrotas e k vitórias. O problema é estimar p supondo

- (1) Ele pode lutar um número indefinido de lutas, mas só pode perder n vezes até sua aposentadoria. Portanto N é uma variável aleatória.
- (2) Ele só pode lutar um número N de lutas e o número de derrotas n é aleatório.

4

Frequência e Probabilidade

Professors of probability have been often and justly derided for arguing as if nature were an urn containing black and white balls in fixed proportions. Quetelet once declared in so many words—“l’urne que nous interrogeons, c’est la nature.” John Maynard Keynes, *Treatise on Probability*

Considere as duas frases abaixo

- 1) Acredito que o estudante que chega a este ponto já estudou algo sobre probabilidade.
- 2) Amiúde o estudante que chega a este ponto já estudou algo sobre probabilidade.

Parece que ambas dizem essencialmente a mesma coisa. Uma expressa uma crença sobre a história dos estudantes, a outra revela que se verifica algo para os alunos que aqui chegaram. Mas não dizem exatamente a mesma coisa. Poderia ser que o conhecimento da primeira deriva de ter estudado o currículo do secundário e mesmo sem nunca ter visto um estudante, nem uma aproximação, poderíamos ter informação sobre o que estudou. A segunda revela que é frequente encontrar estudantes que já fizeram algo.

A linguagem comum pode ser muito rigorosa e sutil. No entanto outras interpretações poderiam ser dadas às frases. Como essencialmente as frases acima não são verdadeiras tentaremos, dentro do formalismo descrito nos capítulos anteriores, deixar mais claro de que forma a intuição de que são equivalentes é justificada e de que forma não o é.

Até agora nos preocupamos com as regras de manipular probabilidades, mas não lhe atribuímos valores numéricos. Vamos começar por estudar de que forma a informação sobre simetria permite essa atribuição.

4.1 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.”

Suponha que você, o jogador J_1 , jogue contra o jogador J_2 . Esta pessoa, por exemplo a Linda, não fala muito bem português e chama os resultados de Karra e Korroa. Consideremos o seguinte jogo, se $\sigma = 1$ você ganha e ela perde. Do contrário, ela ganha. Ela aposta um feijão. Quanto você estaria disposta a apostar?¹ A resposta tem relação, para pessoas racionais, que não dependem do feijão para sobreviver, com as probabilidades $P(\sigma = 1|I_{J_1})$ e $P(\sigma = -1|I_{J_1})$ que você atribui com base na informação I que inclui todo o que se sabe sobre a moeda e a forma como foi jogada².

É natural supor que vocês concordem que

$$\begin{aligned} P(\sigma = 1|I_{J_2}) &= P(\sigma = 1|I_{J_1}) \\ P(\sigma = -1|I_{J_2}) &= P(\sigma = -1|I_{J_1}). \end{aligned} \quad (4.1)$$

Mas agora descobrimos uma falha enorme de comunicação, o que Linda chama de Karra, você chama de coroa. Vocês pensam um pouco e atribuem probabilidades

$$\begin{aligned} P(\sigma = 1|I'_{J_2}) &= P(\sigma = -1|I'_{J_1}) \\ P(\sigma = -1|I'_{J_2}) &= P(\sigma = 1|I'_{J_1}). \end{aligned} \quad (4.2)$$

onde I' descreve o novo estado de informação. Se os jogadores acharem que a nova informação não leva a mudar suas expectativas com respeito à atribuição de probabilidades, ou seja são indiferentes, dirão que os conjuntos de equações 4.1 e 4.2 continuam válidos, mas agora podem ser escritos

$$\begin{aligned} P(\sigma = 1|I''_{J_2}) &= P(\sigma = 1|I''_{J_1}) \\ P(\sigma = -1|I''_{J_2}) &= P(\sigma = -1|I''_{J_1}) \\ P(\sigma = 1|I''_{J_2}) &= P(\sigma = -1|I''_{J_1}) \\ P(\sigma = -1|I''_{J_2}) &= P(\sigma = 1|I''_{J_1}). \end{aligned} \quad (4.3)$$

onde I'' declara que I e I' são equivalentes.

Dado que $P(\sigma = 1|I''_{J_1}) + P(\sigma = -1|I''_{J_1}) = 1$ e que ambos termos são iguais a $P(\sigma = -1|I''_{J_2})$, devemos concluir que $P(\sigma = 1|I''_{J_1}) = 1/2$ e $P(\sigma = -1|I''_{J_1}) = 1/2$.

Porque tantas voltas para chegar ao óbvio? Por vários motivos. Em primeiro lugar notamos que este não é o único exemplo onde usaremos simetria ou indiferença. 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, com σ tomando 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*.

¹ Jaynes 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 amíúde não são por dinheiro mas sim pela própria vida.

² Este 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, mesmo que isso acabe com objetivos de longo prazo. Definir racionalidade deve passar por estipular uma escala de tempo em que o agente deve maximizar algo que pode ser chamado de *utilidade* ou *felicidade*, mas às vezes na ausência de boas definições, são comumente substituídas por *dinheiro*. Em ciência e em geral nas atividades humanas, perguntas difíceis costumam ser substituídas por outras mais simples, à primeira vista parecidas, mas que não necessariamente o são. Veja o livro de D. Kahneman, *Thinking fast and slow*.

Não tem a ver com a simetria do dado. Representar o dado através de um modelo matemático para o cubo perfeito, de densidade uniforme, não passa de uma aproximação. Não é que será difícil, mas é impossível de aproximar na prática. Portanto $1/n$ é devido à simetria de informação e não a simetria física do cubo.

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

4.2 Moedas, Dados, Baralhos, Urnas

Ao longo dos estudos o estudante encontrará sistemas que são simples e portanto estudados muitas vezes. Em dinâmica estudará a partícula livre e o oscilador harmônico, posteriormente o átomo de hidrogênio e o spin de Ising. Em termodinâmica usará caixas rígidas de paredes termicamente isolantes. Nada será tão simples na vida real. Uma partícula nunca está isolada. Nem mesmo o átomo de hidrogênio é um próton e um elétron e nada mais. E mesmo assim é desta forma que aprendemos. Aqui a urna, estudada por Bernoulli e Laplace é o sistema simples. Um baralho de cartas ou uma moeda também são sistemas simples e recorrentes, embora nunca sejam de interesse final nas aplicações que nos motivam a estes estudos. Não obstante Quetelet, a urna ideal não tem nada a ver com a natureza. Isto é um exercício e se não soubermos como agir em condições simples não teremos nenhuma chance contra os problemas reais. É um erro grosseiro olhar para um recorte do mundo, achar que é uma urna e depois criticar a teoria de probabilidades por resultados que contradigam o bom senso.

Uma urna ideal é uma bolsa opaca com bolas iguais ao tato. Alguém com uma luva de box fará a extração de uma bola por vez. Há vários jogos que podem ser jogados. O conjunto de bolas pode ter número conhecido ou não. As bolas podem ter cores diferentes e poderemos saber ou não quantas bolas de cada cor estão dentro. Podemos retirar bolas e repó-las ou não, podemos tirar uma bola sem ver que tipo é e proceder a retirar outras. Você pode retirar a bola de uma urna que eu preparei, ou você pode ver um mago retirar a bola de uma urna que você viu enquanto ele a preparava. Há uma fauna enorme de jogos que podem ser feitos e essencialmente em todos, o objetivo é fazer previsões sobre o que pode ocorrer a seguir, ou o que pode ter ocorrido antes.³

4.3 Urnas

O caso mais simples talvez seja $I_1 = \text{"uma urna com } N \text{ bolas numeradas de } i = 1 \dots N \text{"}$. Qual é a probabilidade de extrair a bola j ? Por simetria de informação é natural associar a mesma probabilidade a cada uma delas. Como são exclusivas e mutuamente exaustivas além de iguais, temos que $P(\text{bola}$

³ *Predictions are risky, specially about the future.* Vários autores, alguns sérios outros não. Já a vi atribuída a Bertrand Russell e Niels Bohr mas também a Dan Quayle e Yogi Berra. Não sei se estas atribuições são verdadeiras. O significado de uma frase é condicionada a quem a enunciou.

$= j|I) = P(j|I) = 1/N$. Isso é óbvio. Parece até uma imposição da qual não podemos escapar. Mais ainda, uma lei da física. Mas certamente não é.

Suponha que voce jogue contra um mafioso e a bola será extraída por um mágico profissional cuja filha foi raptada pelo mafioso. É claro que você deve suspeitar que as probabilidades das diferentes bolas não devem ser iguais para o mágico nem para o mafioso. Mas e para você? A simetria de sua informação não permite distinguir entre as bolas e não pode ir além de atribuir a mesma probabilidade. Agora você escuta que o mágico sugeriu ao mafioso apostar na bola 17. A informação não é mais simétrica. Tudo isso ocorreu antes de extrair uma bola sequer. A frequência ainda não pode ser definida.

Voltemos ao caso simétrico. $I_2 =$ "Das N bolas M são vermelhas (V) e $K = N - M$ são azuis (A)". Por simplicidade para $1 \leq i \leq M$ as bolas são vermelhas e para $M + 1 \leq i \leq N$ são azuis. Esqueça o mágico, agora acreditamos que a pessoa que realiza extração não é influenciada pela cor da bola. Portanto a probabilidade de extração de cada bola é igual a $1/N$. Qual é a probabilidade que a bola extraída seja vermelha? O evento "a bola é V " é verdadeiro se a bola extraída tem o número i com $1 \leq i \leq M$. Os eventos "a bola é i " são mutuamente exclusivos, portanto $V = (i = 1) \vee (i = 2) \vee \dots \vee (i = M)$ que a bola seja V é a união ou soma de que tenha índice $1 \leq i \leq M$. A regra da soma nos dá

$$\begin{aligned} P(V|I_2) &= \sum_{i=1}^M P(i|I_2) = M \times \frac{1}{N} \\ &= \frac{M}{N} \end{aligned} \quad (4.4)$$

Este é um resultado obtido a partir da regra da soma e da simetria de informação sobre as bolas antes de extrair uma única bola. Na seção 2.1.1 vimos que em algum ponto da história isto foi usado como definição de probabilidade por Bernoulli e Laplace ⁴. A probabilidade de extração de uma bola V é simplesmente a razão entre os casos "favoráveis" ou vermelhos e o total de casos. O estudante pode achar que já sabia isto e portanto é uma perda de tempo. Deve entender que o objetivo aqui era o de identificar as hipóteses por trás deste resultado trivial e intuitivo. Deve ficar claro que isto não é nenhuma frequência porque ainda não foi retirada uma única bola da urna. Aprender a identificar as hipóteses subjacentes é um dos objetivos do curso. Quando é fácil, quando é intuitivo, quando lembramos de ter escutado falar deste problema no curso primário, parece desnecessário percorrer um caminho longo. Quando o estudante tiver que resolver problemas nunca antes vistos, ou mais interessante ainda, formular novos problemas, o exercício de identificar as hipóteses subjacentes será amiúde a única ferramenta disponível. Vemos que a regra M/N é muito retritiva pois se aplica ao caso I_2 e não permite levar em conta a existência de mafiosos nem outras variantes que podem ocorrer na

⁴ Repetimos: "...probabilidade, que é então simplesmente a fração cujo numerador é o número de casos favoráveis e cujo denominador é o número de todos os casos possíveis." No contexto:

"The theory of chances consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible."

A Philosophical Essay on Probabilities, Pierre Simon, Marquis de Laplace. 6a ed. F.W.Truscott e F.L. Emory trads.

natureza. Portanto não deveria ser tomada como a definição de probabilidades mas simplesmente um resultado obtido a partir das regras de manipulação dos números que representam nossas crenças, obtidas no capítulo 1, para uma experiência realizada sob um conjunto de restrições determinado.

4.3.1 Urnas: extrações repetidas com reposição.

Extraímos uma bola, que chamamos a primeira, anotamos sua cor e chamamos x_1 que pode ser V ou A . Colocamos a bola novamente, isto é chamado de Reposição. Chacoalhamos a urna. Fazemos isso R vezes e obtemos assim a série $D_R = \{x_1, x_2, \dots, x_R\}$, que chamamos os Dados (dado=*datum* e dados=*data* em inglês.) Chamaremos R de tamanho da sequência.

Pense e discuta o que significa chacoalhar a urna. Para cada extração estamos nas condições do caso anterior: M , K e N tem o mesmo significado que antes. O resultado de uma extração independe de quais foram as bolas extraídas antes:

$$P(x_n = V | x_1, x_2, \dots, x_{n-1} I_2) = P(x_n = V | I_2) = \frac{M}{N} \quad (4.5)$$

Para uma dada sequência usamos a regra do produto

$$\begin{aligned} P(x_1, x_2, \dots, x_n | I_2) &= P(x_n | x_1, x_2, \dots, x_{n-1} I_2) P(x_1, x_2, \dots, x_{n-1} | I_2) \\ &= P(x_n | I_2) P(x_{n-1} | x_1, x_2, \dots, x_{n-2} I_2) P(x_1, x_2, \dots, x_{n-2} | I_2) \\ &= \dots \\ &= P(x_n | I_2) P(x_{n-1} | I_2) \dots P(x_1 | I_2) \\ &= \prod_{i=1}^n P(x_i | I_2) \end{aligned} \quad (4.6)$$

Se a sequência for e.g. $VVAAAVV$ teremos

$$P(VVAAAVV | I_2) = ppqqqpp = p^4 q^3 \quad (4.7)$$

onde usamos a notação $p = M/N$ e $q = K/N$, com $p + q = 1$. Devido à independência entre os resultados de cada extração, a ordem temporal das ocorrências de vermelho e azul é irrelevante, portanto a única coisa que importa é o número de vezes que na sequência apareceu o vermelho ou que apareceu o azul.

4.4 A distribuição binomial

Agora fazemos outra pergunta: independentemente da ordem, qual é a probabilidade de ter m vermelhas e $k = R - m$ azuis (numa extração com reposição de R repetições da extração de uma bola, quando M e K são os números conhecidos de bolas vermelhas e azuis, respectivamente)? É comum dizer de forma equivalente que queremos a distribuição de m sucessos em R tentativas, quando a probabilidade de sucesso é $p = M/N$.

Novamente usaremos as regras da probabilidade. Primeiro as sequências diferentes de R extrações são eventos mutuamente

exclusivos. Ou aconteceu uma, ou aconteceu outra, alguma aconteceu e não podem ser duas simultaneamente verdadeiras. Dado R , a probabilidade de obter m bolas vermelhas (e portanto obrigatoriamente k azuis) é obtida da regra da soma, como a soma das probabilidades sobre todas as sequências com m, k . Mas cada sequência tem a mesma probabilidade $p^m q^k$, buscamos portanto o número de sequências com m e k .

O resultado deve ser familiar. Chame o número de sequências de tamanho R com m, k de C_R^m . Considere que já resolvemos o problema para sequências de tamanho $R - 1$, para qualquer $0 \leq m \leq R - 1$. Portanto C_{R-1}^{m-1} e C_{R-1}^m são consideradas conhecidas. Suponha que extrairmos $R - 1$ bolas. Há somente duas formas de obter m e k após a retirada da última bola. Isto só pode ocorrer se após $R - 1$ extrações

- (i) tivermos obtido $m - 1$ vermelhas após $R - 1$ extrações e na R -ésima, for extraída uma bola vermelha, o que pode ter ocorrido de C_{R-1}^{m-1} formas,
- (ii) tivermos obtido m vermelhas após $R - 1$ extrações e na R -ésima, for extraída uma bola azul que pode ter ocorrido de C_{R-1}^m formas diferentes.

Portanto, temos, para $R > 1$ a relação de recorrência para o número de sequências

$$C_R^m = C_{R-1}^{m-1} + C_{R-1}^m, \quad (4.8)$$

que é a famosa relação de recorrência devida a Pascal. Isto é uma máquina de gerar os coeficientes binomiais, que precisa ser alimentada com valores iniciais. Para $R = 1$ é óbvio que $C_1^0 = C_1^1 = 1$, pois se olharmos sequências de tamanho 1, só há duas possibilidades, a primeira bola foi azul ($C_1^0 = 1$), ou alternativamente foi vermelha ($C_1^1 = 1$).

Usando a notação do fatorial, que é definida pela recursão $n! = n(n-1)!$, para $n = 1, 2, \dots$ inteiros positivos, com condições iniciais $0! = 1$ e portanto $n! = 1.2.3 \dots n$, os coeficientes são dados por

$$C_R^m = \frac{R!}{m!(R-m)!}, \quad (4.9)$$

pois satisfazem as relações de recorrência e às condições iniciais. Basta provar unicidade da solução, que é fácil. Note que simplesmente, usando o resultado 4.9, e

$$\frac{R!}{m!(R-m)!} = \frac{(R-1)!}{(m-1)!(R-1-m+1)!} + \frac{(R-1)!}{m!(R-1-m)!} \quad (4.10)$$

temos que a relação 4.8 é satisfeita. Estes coeficientes são chamados os coeficientes binomiais. O motivo disto é que

$$(a+b)^R = \sum_{m=0}^R C_R^m a^m b^{R-m}, \quad (4.11)$$

que é amplamente conhecida desde Newton. Mas é instrutivo provar este resultado, supondo-o válido para $R - 1$, calculando

$(a + b)^{R-1}(a + b)$ e usando a relação de recorrência. A notação $C_R^m = \binom{R}{m}$ também é muito popular e é dito que representa o número de maneiras de escolher m elementos de um total de R ou o número de combinações de R , m a m .

Temos o resultado desejado,

$$P(m|p, R, I_2) = \binom{R}{m} p^m q^{R-m} \quad (4.12)$$

que é a distribuição binomial. Também poderíamos ter escrito $P(m|M, R, I_2)$. Obviamente a distribuição está normalizada, pois

$$\sum_{m=0}^R P(m|p, R, I_2) = \sum_{m=0}^R \binom{R}{m} p^m q^{R-m} = (p + q)^R = 1 \quad (4.13)$$

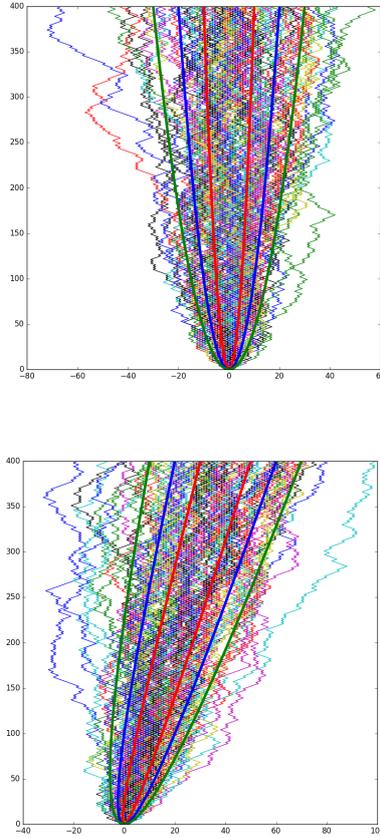


Figura 4.1: Modelo de difusão I: Caminhos aleatórios binomiais, $N = 100$ corridas de $T_{\max} = 400$ passos cada uma. Acima: A cada instante de uma dinâmica discreta, um caminhante dá um passo à direita com $p = 1/2$ ou à esquerda com $q = 1/2$, independentemente de qualquer outra coisa. As parábolas mostram os valores σ , 2σ e 2σ respectivamente, como função do tempo, onde $\sigma = \sqrt{Tp(1-p)}$. Abaixo: $p = 0.55$.

4.4.1 Momentos da Binomial

É interessante calcular os valores esperados da distribuição binomial. A expressão da expansão binomial 4.11 escrita com p e q arbitrário é útil para calcular os valores esperados $\langle m \rangle$, $\langle m^2 \rangle$.

Usamos a expansão binomial para valores p e q quaisquer, derivamos com respeito a p e multiplicamos por p para obter, usando o truque que $p \frac{\partial}{\partial p} p^m = m p^m$:

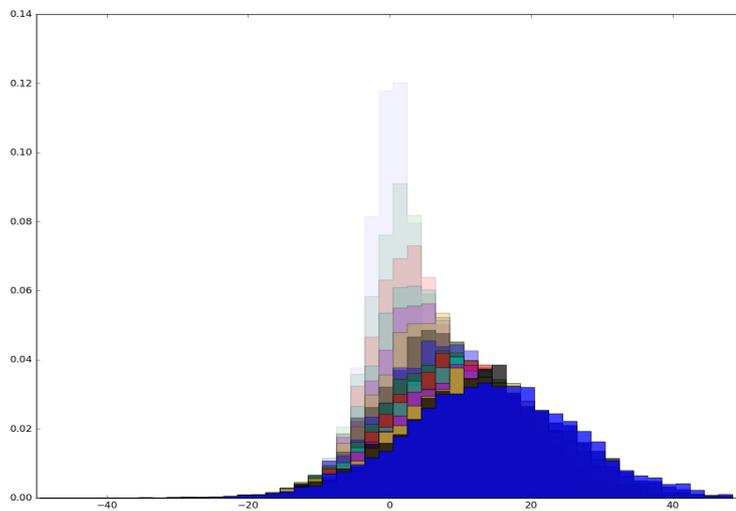
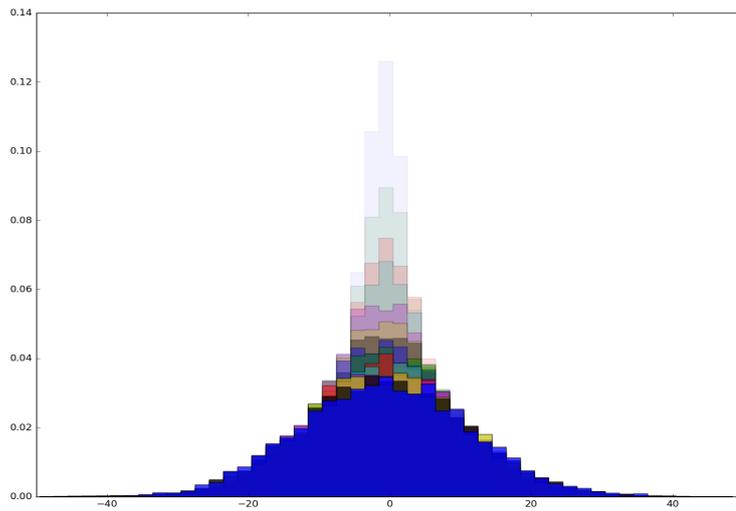
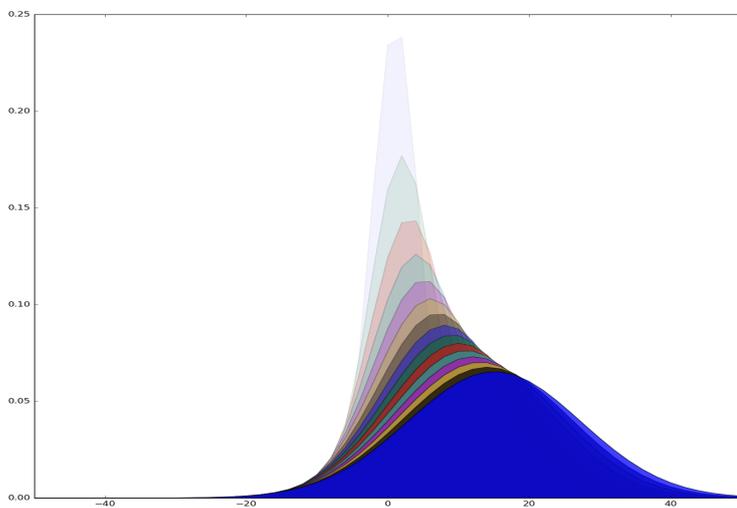
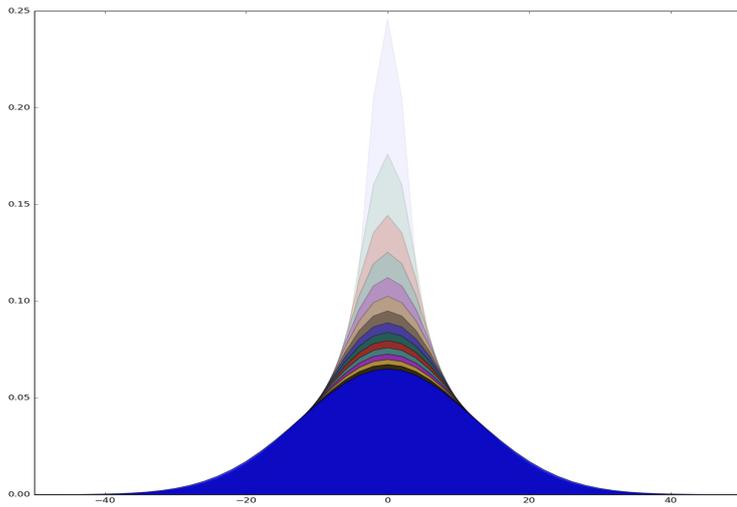


Figura 4.2: Difusão II: Histogramas obtidos dos caminhos simuladas da figura para valores $N = 10, 20, \dots, 160$ com (Acima) $p = 1/2$, (Abaixo) $p = .55$.

Figura 4.3: Difusão III: A distribuição binomial para valores $N = 10, 20, \dots, 160$ com (Acima) $p = 1/2$, (Abaixo) $p = .55$.



$$\begin{aligned}
\langle m \rangle &= \sum_{m=0}^R m P(m|p, R, I_2) = \sum_{m=0}^R \binom{R}{m} m p^m (1-p)^{R-m} \\
&= \left(\sum_{m=0}^R \binom{R}{m} m p^m q^{R-m} \right)_{q=1-p} \\
&= \left(\sum_{m=0}^R \binom{R}{m} \left(p \frac{\partial}{\partial p} p^m \right) q^{R-m} \right)_{q=1-p} \\
&= \left(p \frac{\partial}{\partial p} \sum_{m=0}^R \binom{R}{m} p^m q^{R-m} \right)_{q=1-p} \\
&= \left(p \frac{\partial}{\partial p} (p+q)^R \right)_{q=1-p} \\
&= pR(p+1-p)^{R-1} = pR \tag{4.14}
\end{aligned}$$

O truque vale somente se colocarmos $q = 1 - p$ no final⁵. Para calcular $\langle m^2 \rangle$ vemos que dentro da soma aparece $m^2 p^m$ que podemos escrever como

$$p \frac{\partial}{\partial p} \left(p \frac{\partial}{\partial p} p^m \right) = m^2 p^m$$

que permite escrever

$$\langle m^2 \rangle = \left[\left(p \frac{\partial}{\partial p} \left(p \frac{\partial}{\partial p} (p+q)^R \right) \right) \right]_{q=1-p}$$

que leva a $\langle m^2 \rangle = R^2 p^2 + Rp(1-p)$

A variância é comumente denotada $\text{var}(m)$ ou σ^2 ou ainda σ_m^2 e definida por

$$\sigma_m^2 = \langle m^2 \rangle - \langle m \rangle^2$$

e portanto para a distribuição binomial de m sucessos em R tentativas a raiz variância ou o desvio padrão é

$$\sigma_m = \sqrt{Rp(1-p)}. \tag{4.15}$$

Olhe as figuras 5.8 e 4.3. Na primeira são mostradas trajetórias individuais e na segunda as distribuições binomiais para $p = 0.5$ e para $p = 0.55$, para valores de R cada vez maiores. Para $p \neq 0.5$ há deriva. O deslocamento, após R passos dos quais m são para a direita e $R - m$ são para a esquerda é

$$X = m - (R - m) = 2m - R$$

e o valor médio do deslocamento é

$$\mathbb{E}(X) = 2\langle m \rangle - R = (2p - 1)R \tag{4.16}$$

que é positivo para $p > 1/2$.

A comparação entre estas figuras das trajetórias e da distribuição permitirá começar a entender o processo de simulação conhecido como Monte Carlo, onde um processo individual, gerado muitas

⁵ É importante notar que a derivada parcial $(\partial f(p, q) / \partial p)_q$ é definida reduzindo a função de duas variáveis a uma função de uma só variável, que é feito ao declarar que q é mantido constante. Se pensarmos na superfície $z = f(p, q)$, notamos que em um dado ponto (p_1, q_1) podemos tomar a derivada em qualquer direção, em particular mantendo $q = q_1$ fixo, ou mantendo $p = p_1$ fixo que dá $(\partial f(p, q) / \partial q)_p$ ou ainda ao longo de qualquer direção, e.g $p = 1 - q$, mas os resultados não são os mesmos.

vezes permite estimar valores esperados de funções de uma variáveis estocásticas cuja distribuição pode ser muito difícil de tratar analiticamente. A raiz quadrada que aparece na equação 4.15 é extremamente importante. Não ocorre por acaso e de forma específica para a binomial. Somamos um número grande de passos gerados por Bernoulli. Toda vez que ocorrer uma soma de variáveis estocásticas, se a variância individual de cada termo for finita e sob condições de independência dos passos (suficiente mas não necessária) a variância crescerá com N e a largura da distribuição com \sqrt{N} . Voltaremos a isto no capítulo sobre o Teorema do Limite Central.

4.5 Frequência não é probabilidade

Porque parece razoável confundir frequência e probabilidade? O que segue é importante. A probabilidade de bola vermelha ou de sucesso é p . O valor esperado do número de sucessos é $\langle m \rangle = Rp$, portanto

$$p = \frac{\langle m \rangle}{R} \quad (4.17)$$

ou seja

$$p = \left\langle \frac{m}{R} \right\rangle = \langle f \rangle \quad (4.18)$$

onde $f = m/R$ é a frequência de sucessos. Em palavras, o valor esperado da frequência é a probabilidade de sucesso. A frequência não é a probabilidade. A frequência é um número que depende do experimento realizado. Isto caracteriza a frequência como um número aleatório. A variância da frequência é

$$\begin{aligned} \sigma_f^2 &= \langle f^2 \rangle - \langle f \rangle^2 \\ &= \left\langle \left(\frac{m}{R} \right)^2 \right\rangle - \left\langle \frac{m}{R} \right\rangle^2 \\ &= \frac{1}{R^2} \sigma_m^2 \\ &= \frac{R}{R^2} p(1-p) = \frac{1}{R} p(1-p) \\ \sigma_f &= \frac{1}{\sqrt{R}} \sigma_m \end{aligned} \quad (4.19)$$

Isto significa que embora a frequência seja um número que depende do experimento particular e só o seu valor esperado seja a probabilidade de sucesso, à medida que o número de tentativas R aumenta, seu desvio padrão vai a zero com $1/\sqrt{R}$. Portanto qualquer experimento que meça a frequência encontrará valores perto da probabilidade para R grande o que pode levar alguns de vocês à possibilidade de confundir frequência com probabilidade. Isto porém não é perdoável.

O que significa perto e grande no parágrafo acima será discutido com mais cuidado no capítulo 5, onde faremos estas idéias mais precisas olhando para a desigualdade de Chebyshev e definindo

convergência em probabilidade. Seremos, então, capazes de dizer o que significa que f converge para p quando R aumenta. Também olharemos o problema relacionado de inferência de p dada a frequência no capítulo ??

4.5.1 A distribuição Multinomial

Suponha que o processo seja descrito por I_{Multi} = "na urna há N bolas de no máximo C cores, M_c da cor c , $\sum_{c=1...C} M_c = N$. As bolas extraídas são repostas na urna".

Temos, analogamente ao caso de duas cores, que a probabilidade de extrair uma bola de uma cor c é $p_c = M_c/N$. Obviamente $\sum_{c=1...C} p_c = 1$, porque afinal uma bola extraída é de alguma cor. Para uma sequência de N extrações com reposição usamos o fato que as sequências são mutuamente exclusivas e a regra da soma para obter

$$P(m_1, \dots, m_C | I_{Multi}) = C_N^{m_1, m_2, \dots, m_C} p_1^{m_1} p_2^{m_2} \dots p_C^{m_C}$$

Normalização leva a

$$\sum_{\sum m_c = N} P(m_1, \dots, m_C | I_{Multi}) = 1$$

Supomos novamente que já resolvemos o caso de de $R - 1$ extrações e consideramos a extração de mais uma bola. O número total de casos deve satisfazer

$$C_N^{m_1, m_2, \dots, m_C} = C_{N-1}^{m_1-1, m_2, \dots, m_C} + C_{N-1}^{m_1, m_2-1, \dots, m_C} + \dots + C_{N-1}^{m_1, m_2, \dots, m_C-1} \quad (4.20)$$

onde o termo do lado direito em que aparece $m_c - 1$ é o número de sequências em que faltava uma bola da cor c para chegar ao caso denotado no lado esquerdo: $\{m_1, m_2, \dots, m_C\}$ em R extrações. As C condições iniciais $C_1^{0, \dots, 0, 1, 0, \dots, 0} = 1$ são suficientes para girar a manivela da relação de recorrência 4.20. O resultado é que

$$C_N^{m_1, m_2, \dots, m_C} = \frac{N!}{m_1! m_2! \dots m_C!} \quad (4.21)$$

pois substituindo na relação de recorrência

$$\begin{aligned} C_N^{m_1, m_2, \dots, m_C} &\stackrel{?}{=} \sum_c \frac{(N-1)!}{m_1! m_2! \dots (m_c-1)! \dots m_C!} \\ &= \frac{\sum_c m_c (N-1)!}{m_1! m_2! \dots m_C!} \\ &= \frac{N(N-1)!}{m_1! m_2! \dots m_C!} \\ &= \frac{N!}{m_1! m_2! \dots m_C!} \end{aligned} \quad (4.22)$$

vemos que 4.20 é de fato satisfeita pelas expressões 4.21. Verifique que as condições iniciais são satisfeitas. Falta provar unicidade. Mas isso é simples e é deixado para os leitores interessados.

4.5.2 Urnas sem reposição: a distribuição hipergeométrica.

A diferença fundamental com relação aos casos anteriores é que vale $I_4 =$ "a extração de cada bola é feita sem reposição das anteriores, (inicialmente N bolas, M vermelhas)" e portanto em condições diferentes das anteriores. A primeira extração é igual ao caso anterior

$$P(x_1 = V|N, M, I_4) = \frac{M}{N}$$

Agora a diferença, no segundo passo o estado da urna e portanto as probabilidades dependem do resultado da primeira extração

$$P(x_2, x_1|N, M, I_4) = P(x_2|x_1, M, N, I_4)P(x_1|N, M, I_4)$$

Se as duas forem vermelhas, teremos

$$\begin{aligned} P(x_2 = V, x_1 = V|N, M, I_4) &= P(x_2 = V|x_1 = V, M, N, I_4)P(x_1 = V|N, M, I_4) \\ &= P(x_2|N-1, M-1, I_4)P(x_1 = V|N, M, I_4) \\ &= \frac{M-1}{N-1} \frac{M}{N} \end{aligned} \quad (4.23)$$

pois na segunda extração há somente $N-1$ bolas, das quais $M-1$ são vermelhas. A probabilidade que as primeira r bolas extraídas sejam vermelhas é

$$\begin{aligned} P(x_r = V, \dots, x_2 = V, x_1 = V|N, M, I_4) &= \frac{(M-r-1)\dots(M-1)M}{(N-r-1)\dots(N-1)N} \\ &= \frac{M!(N-r)!}{(M-r)!N!} \end{aligned} \quad (4.24)$$

que faz sentido mesmo que $r > M$ se for convencionado que o fatorial de números negativos é infinito. Continuamos, mas agora calculamos as probabilidades que as bolas seguintes sejam azuis. O estado da urna é de $N-r$ bolas, das quais $M-r$ são vermelhas, e a probabilidade de extrair uma bola azul é:

$$P(x_{r+1} = A|N-r, M-r, I_4) = \frac{N-r-(M-r)}{N-r} = \frac{N-M}{N-r}.$$

Repetindo

$$P(x_{r+b} = A, \dots, x_{r+1} = A|N-r, M-r, I_4) = \frac{(N-M)!(N-r-b)!}{(N-M-b)!(N-r)!}.$$

Assim chegamos a que uma sequência de r vermelhas seguidas por b azuis tem probabilidade, pela regra do produto

$$\begin{aligned} P(x_{r+b} = A, \dots, x_{r+1} = A, x_r = V, \dots, x_1 = V|N, M, I_4) &= P(x_r = V, \dots, x_1 = V|N, M, I_4) \\ &\times P(x_{r+b} = A, \dots, x_{r+1} = A|x_r = V, \dots, x_1 = V, N, M, I_4) \end{aligned}$$

que pode ser escrito como

$$\begin{aligned} &= \frac{M!(N-r)!}{(M-r)!N!} \frac{(N-M)!(N-r-b)!}{(N-M-b)!(N-r)!} \\ &= \frac{M!}{(M-r)!N!} \frac{(N-M)!(N-r-b)!}{(N-M-b)!}. \end{aligned}$$

Note que os fatoriais são de

- N e $(N - r - b)$ os números inicial e final de bolas na urna
- M e $N - M$, os números iniciais de bolas vermelhas e de azuis.
- $M - r$ e $(N - M - b)$ os números finais de bolas vermelhas e de azuis.

Ou seja não aparece nada que diga a ordem em que foram extraídas, primeiro as vermelhas depois as azuis. Isto deve ser verdade para qualquer ordem de extração, desde que os resultados finais de extração r e b sejam os mesmos. Vejamos se é assim.

Suponha que numa sequência S_1 de $r + b$ extrações a k -ésima bola vermelha ocorreu na posição l e da k' -ésima bola azul na $l + 1$, e na sequência S_2 a k -ésima bola vermelha foi extraída após $l + 1$ extrações e a k' -ésima bola azul após l . Aparte dessa troca, as sequências são iguais. Os fatores que contribuem à probabilidade são para a sequência S_1

$$\dots \frac{M - k - 1}{N - l} \frac{N - M - k' - 1}{N - l - 1} \dots$$

e para a sequência S_2

$$\dots \frac{N - M - k' - 1}{N - l} \frac{M - k - 1}{N - l - 1} \dots$$

que são iguais. Seque que a probabilidade de extrair r bolas vermelhas e b azuis, independentemente da ordem é dada pelo produto do número de sequências possíveis, $\binom{r+b}{b}$ e da probabilidade de uma sequência:

$$P(\{r, b\} | N, M, I_4) = \frac{(r+b)!}{r!b!} \frac{M!}{(M-r)!N!} \frac{(N-M)!(N-r-b)!}{(N-M-b)!}$$

e simplificando, a probabilidade ao "extrair sem reposição $r + b$ bolas de uma urna com N bolas das quais M são vermelhas e $N - M$ azuis, exatamente r sejam vermelhas" é

$$P(\{r, b\} | N, M, I_4) = \binom{r+b}{r} \frac{\binom{N-r-b}{M-r}}{\binom{N}{M}} \quad (4.25)$$

É interessante que isto pode ser escrito como

$$P(\{r, b\} | N, M, I_4) = \frac{\binom{M}{r} \binom{N-M}{b}}{\binom{N}{r+b}} \quad (4.26)$$

onde o numerador é obtido pelo produto de todas as diferentes combinações de escolhas possíveis de r bolas do total de M vermelhas vezes o número de combinações de b do total de $N - M$ azuis, dividido pelo total de possibilidades das combinações de $r + b$ do total de N bolas. Podemos ainda escrever a mesma expressão de uma forma que fica simétrica e permite generalização para mais cores. Mudando a notação chamamos de M_1 (em lugar de M) o número de bolas vermelhas, M_2 o de azuis; de r_1 o número de bolas da primeira cor, de r_2 o da segunda cor:

$$P(\{r_1, r_2\} | M_1, M_2, I_4) := P(\{r, b\} | N, M, I_4) = \frac{\binom{M_1}{r_1} \binom{M_2}{r_2}}{\binom{M_1+M_2}{r_1+r_2}}. \quad (4.27)$$

É razoável supor, e facilmente demonstrável para o caso de C cores:

$$P(\{r_1, r_2 \dots r_C\} | M_1, M_2, \dots, M_C, I_4) = \frac{\prod_{c=1 \dots C} \binom{M_c}{r_c}}{\binom{\sum_{c=1 \dots C} M_c}{\sum_{c=1 \dots C} r_c}} \quad (4.28)$$

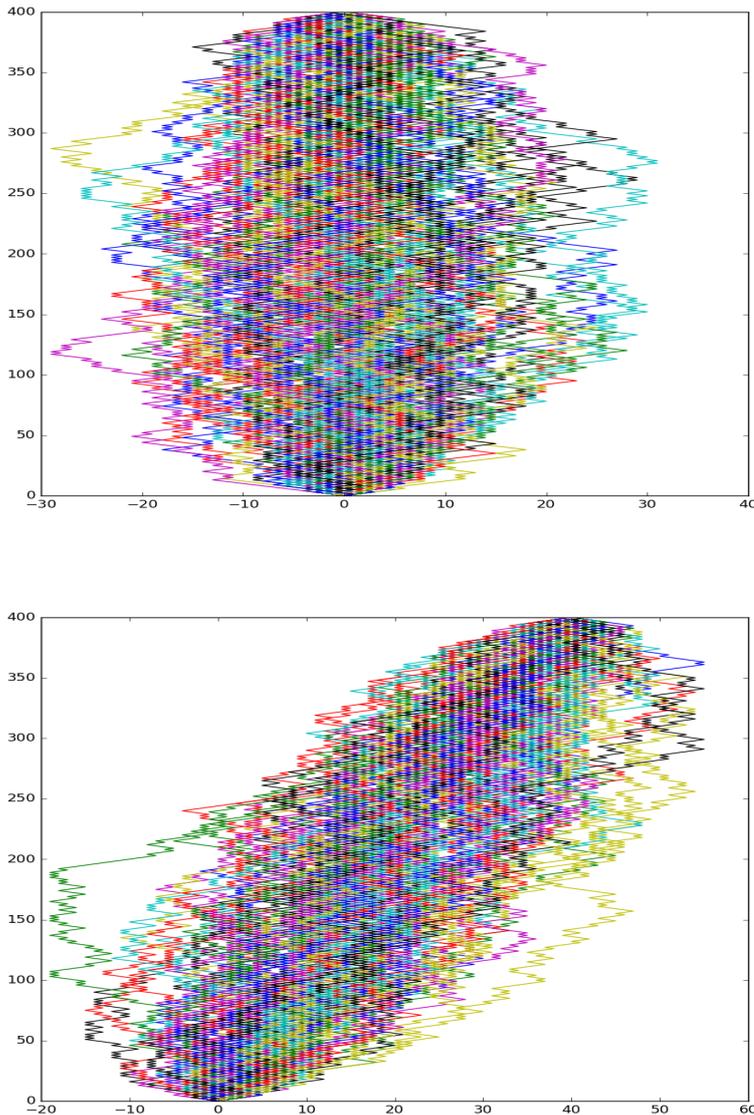


Figura 4.4: Caminhos hipergeométricos. Acima: Urna com $N = 400$ bolas das quais $M_1 = 200$ são vermelhas e $M_2 = 200$ azuis. Abaixo: Urna com $N = 400$, $M_1 = 220$ vermelhas e $M_2 = 180$ azuis. A cada bola vermelha extraída o caminhante anda para a direita, a cada bola azul, para a esquerda.

Os caminhos hipergeométricos para urnas estão mostrados nas figuras 4.4, 4.5 e 4.6. Devido a que a urna não volta ao mesmo estado após a extração as figuras são diferentes dos caminhos binomiais. Há uma difusão inicial, mas as trajetórias convergem para o mesmo lugar. Não importa a história de extração, a urna vazia será a mesma em todos os casos.

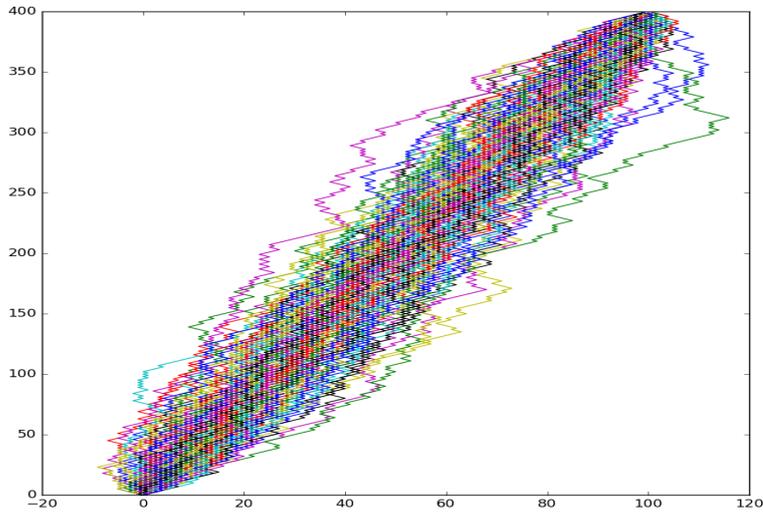


Figura 4.5: Caminhos hipergeométricos. Trajetórias para a urna com $N = 400$, $M_1 = 250$ vermelhas e $M_2 = 150$ azuis.

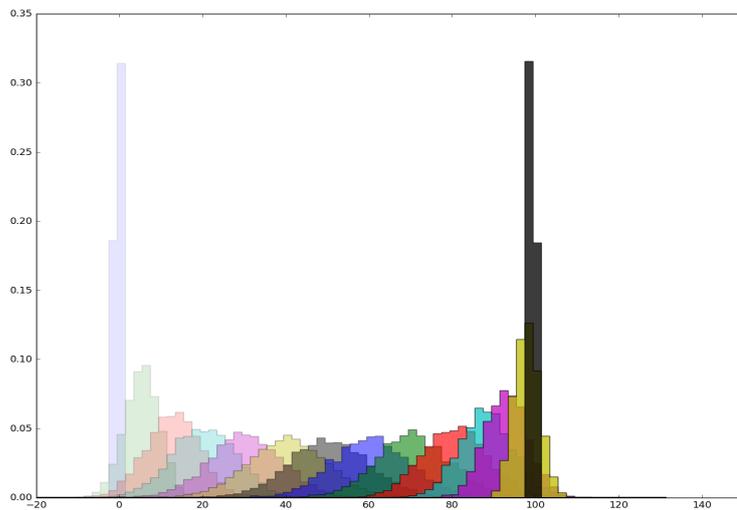


Figura 4.6: Histogramas dos caminhos hipergeométricos: simulação. Urna com $M = 400$, $M_1 = 250$ vermelhas e $M_2 = 150$ azuis. A situação é a mesma da figura anterior. Abaixo: Os histogramas foram gerados após extrair(1,20,50,80,120,160,200,240,280,320,350,370,390,399) bolas e olhar os resultados para 5000 urnas.

4.5.3 Bolas escondidas

Voltemos ao caso sem reposição. N bolas, M vermelhas, $N - M$ azuis. Extraímos uma bola mas (agora a diferença) não somos informados da sua cor. A bola é escondida fora da urna. Qual é a probabilidade $P(x_2 = V|N, M, I_4)$ que a segunda bola seja vermelha? O interesse nestes casos está no método, não no jogo em si. As regras da probabilidade são suficientes para responder isto. Tivemos duas extrações portanto o nosso interesse deve começar por analisar a distribuição conjunta $P(x_2, x_1|N, M, I_4)$. A única coisa que sabemos sobre x_1 é que foi vermelho ou azul, possibilidades excusivas e exaustivas. Portanto

$$\begin{aligned} P(x_2|N, M, I_4) &= \sum_{x_1=V,A} P(x_2, x_1|N, M, I_4) \\ &= \sum_{x_1=V,A} P(x_2|x_1, N, M, I_4)P(x_1|N, M, I_4) \end{aligned} \quad (4.29)$$

que fica escrita em termos de probabilidades que conhecemos. Isto é um exemplo ao contrário do uso de marginalização. Portanto

$$\begin{aligned} P(x_2 = V|N, M, I_4) &= P(x_2 = V|x_1 = V, N, M, I_4)P(x_1 = V|N, M, I_4) \\ &+ P(x_2 = V|x_1 = A, N, M, I_4)P(x_1 = A|N, M, I_4) \\ &= \frac{M-1}{N-1} \frac{M}{N} + \frac{M}{N-1} \frac{N-M}{N} \\ &= \frac{M}{N}. \end{aligned} \quad (4.30)$$

Ou seja, como a primeira extração não nos deu nenhuma informação, a probabilidade de extração no segundo passo continuou sendo M/N .

Podemos pensar sobre o que acontece se as duas primeiras bolas forem escondidas. O mesmo. Se não há informação não há alteração de probabilidades. Mas suponha que extraímos e escondemos uma bola. Extraímos uma segunda e é vermelha. O que isto nos diz sobre a bola escondida? Queremos saber sobre $P(x_1|x_2, N, M, I_4)$. Voltamos a pensar sobre a distribuição conjunta e usamos novamente a regra do produto

$$P(x_1|x_2, N, M, I_4) = \frac{P(x_2, x_1|N, M, I_4)}{P(x_2|N, M, I_4)}.$$

Especificamente, suponha que a segunda bola é vermelha, qual é a probabilidade que a primeira seja vermelha:

$$\begin{aligned} P(x_1 = V|x_2 = V, N, M, I_4) &= \frac{P(x_2 = V, x_1 = V|N, M, I_4)}{P(x_2 = V|N, M, I_4)} \\ &= \frac{\frac{M-1}{N-1} \frac{M}{N}}{\frac{M}{N}} = \frac{M-1}{N-1}, \end{aligned}$$

confirmando o que talvez podia ser desconfiado a o recuperar a probabilidade de extração de uma segunda bola conhecendo o resultado da primeira. Note então que o que é primeiro e o que é

segundo não interessa. O que interessa é que informação está disponível. Se não há informação nenhuma é equivalente a uma primeira extração de bola, se há informação é equivalente a uma segunda extração sabendo a primeira.

4.6 Inversão: Urna com conteúdo desconhecido

Um problema em ciência pode ser descrito como "conhecido o sistema, que previsões podemos fazer sobre o resultado de experiências?" Outro tipo de problema é o inverso, "sabendo o resultado das experiências, o que podemos dizer sobre um sistema desconhecido?"

Considere que o conteúdo da urna é desconhecido e retiramos bolas com reposição. Como reposição significa que em cada extração o estado da urna é o mesmo, mesmo que o nosso estado de informação tenha mudado. Em um dado ponto temos um conjunto de dados, $D_R = \{V, V, V, A, A.. \}$. O que podemos dizer sobre a fração de cores? Este tipo de problema constitui o tópico central do problema de análise de dados experimentais e inclui a idéia fundamental de modelo. Voltaremos com mais detalhes, uma e outra vez, ao longo destas notas. Precisamos definir a informação subjacente I_5 . Consideramos que há somente C cores, cada cor com número M_c de bolas, $N = \sum_c M_c$ o número total de bolas. Portanto a probabilidade de extração de uma bola de cor c seria $p_c = M_c/N$. Acabamos de ver que saberíamos calcular a probabilidade de qualquer sequência dados os p_c . Agora usamos a regra do produto, o truque que não parará de dar resultados. Por facilidade olhemos o caso de duas cores, teremos $p = M/N$ como parâmetro desconhecido. Obtivemos a distribuição binomial 4.12, para m bolas vermelhas em R extrações, quando a fração de bolas vermelhas é p :

$$P(m|p, R, I_2) = \binom{R}{m} p^m q^{R-m}. \quad (4.31)$$

A regra do produto nos dá para a distribuição conjunta de m e p

$$P(p, m|R, I_2) = P(p|R, I_2)P(m|p, R, I_2) = P(m|R, I_2)P(p|m, R, I_2), \quad (4.32)$$

de onde temos o resultado conhecido como a regra de Bayes

$$P(p|m, R, I_2) = \frac{P(p|R, I_2)P(m|p, R, I_2)}{P(m|R, I_2)}. \quad (4.33)$$

Aqui aparece algo novo. Vale a pena respirar e tomar o tempo necessário para assimilar algo que será fundamental no que segue. Temos a probabilidade do parâmetro da binomial, que por sua vez é uma probabilidade. Além disso temos duas probabilidades de p ,

- A distribuição *a priori* $P(p|R, I_2)$
- A distribuição posterior $P(p|m, R, I_2)$. Posterior à inclusão da m nos condicionantes.

Ainda temos $P(m|p, R, I_2)$ que codifica a informação que temos sobre quão provável é um valor de m caso p tenha um valor dado. Esta probabilidade recebe o nome de verossimilhança. O mundo se divide em pessoas que ficam nervosas ao falar da probabilidade de p e aqueles que acham natural falar da probabilidade deste parâmetro. Claro que p é uma probabilidade, mas se lembrarmos que é a razão entre bolas vermelhas e o total, não há motivo para nervosismo. Ainda ficam mais nervosos ao falar da probabilidade de *a priori* - antes de levar em consideração os dados. As questões levantadas aqui serão atacadas no capítulo ??.

4.7 A regra de sucessão de Laplace

O que vem a seguir é interessante por pelo menos dois motivos. Primeiro porque mostra a aplicação dos métodos desenvolvidos a um problema de urna interessante, onde as hipóteses ficam claras, pois senão não é possível fazer as contas. O segundo é histórico. A previsão feita é usada agora em problemas que não tem nada a ver com as hipóteses e se chega a algo que viola as expectativas do bom senso. Para alguns autores isto é indicação que as regras da probabilidade usadas por Laplace não fazem sentido. Isso tem acontecido e a discussão sobre porque isto ocorre e como evitar este tipo de procedimento é instrutivo para o aluno. Faço hipóteses, calculo um resultado, aplico em outro problema onde a informação é diferente e portanto espero resultados diferentes, e como os dois não batem critico a teoria. Parece mais política que ciência.

Consideremos uma urna de composição desconhecida, exceto por ter bolas de somente duas possibilidades de cores e procedemos a extrações com reposição. A reposição significa que cada extração é independente e em condições idênticas às anteriores. Outra forma de colocar o problema é considerando um processo de Bernoulli, dois estados $s = 1$ ou $s = -1$, ou sucesso e fracasso. Não sabemos o valor parâmetro p . Os dois casos são idênticos se na urna o número de bolas for infinito.

As asserções relevantes para o problema são as seguintes:

- $N =$ "foram feitas N tentativas consecutivas de Bernoulli"
- $n =$ "dado N , foram obtidos n sucessos"
- $M =$ "foram feitas M tentativas consecutivas de Bernoulli"
- $m =$ "dado M , foram obtidos m sucessos"
- $I =$ descrição do processo

O fato de usar o mesmo símbolo para um número e uma asserção deve ser perdoado por simplificar a notação.

O objetivo do exercício é determinar com base em um primeiro experimento descrito por N e n , qual é a probabilidade $P(m|nMN)$ de obter m após M .

Começamos por identificar o que não sabemos, m e p , a probabilidade de sucesso, que é o parâmetro da binomial. Como dissemos na secção anterior alguns autores tentam evitar falar de probabilidade de uma probabilidade, enfatizarmos que p é um parâmetro de uma distribuição, logo não deve haver resistência à sua estimativa e representação através de distribuições que codifiquem o que sabemos. Portanto estamos interessados na distribuição de probabilidades conjunta de m e p dado o que sabemos: $P(m, p|nMNI)$. Mas não estamos interessados em p , e portanto marginalizamos

$$P(m|nMNI) = \int_0^1 P(m, p|nMNI) dp.$$

A regra do produto leva a

$$P(m|nMNI) = \int_0^1 P(m|pnMNI)P(p|nMNI)dp. \quad (4.34)$$

Jogar M vezes o jogo sem saber o resultado não dá informação sobre p , portanto $P(p|nMNI) = P(p|nNI)$. Como sabemos que a probabilidade de obtenção de n sucessos em N tentativas é uma binomial $P(n|pNI)$ e podemos usar Bayes para inverter:

$$P(p|nNI) = \frac{P(p|NI)P(n|pNI)}{P(n|NI)} \quad (4.35)$$

O denominador $P(n|NI)$ pode ser obtido por normalização, portanto não nos preocupa. Novamente, é irrelevante saber N e não saber n , portanto temos $P(p|NI) = P(p|I)$ para o *a priori*. Fazemos a suposição que não temos, antes de ver os dados, nenhuma preferência por qualquer valor de p , portanto $P(p|I) = 1$, é a distribuição uniforme.

$$\begin{aligned} P(p|nNI) &\propto P(n|pNI) \propto p^n(1-p)^{N-n} \\ &= \frac{p^n(1-p)^{N-n}}{\int_0^1 p'^n(1-p')^{N-n} dp'} \\ &= \frac{(N+1)!}{n!(N-n)!} p^n(1-p)^{N-n}, \end{aligned} \quad (4.36)$$

que reconhecemos como a distribuição Beta(n, N) de p após n sucessos em N tentativas. Para a normalização usamos o resultado devido a Euler, ver equação 3.16

$$E_k^r = \int_0^1 p^r(1-p)^k dp = \frac{r!k!}{(r+k+1)!} \quad (4.37)$$

Voltamos ao cálculo de 4.34, notando que $P(m|pnMNI) = P(m|pMI)$, pois saber p torna desnecessária a

informação de n, N ,

$$\begin{aligned}
 P(m|nMNI) &= \int_0^1 P(m|pMI)P(p|nNI)dp \\
 &= \binom{M}{m} \frac{(N+1)!}{n!(N-n)!} \int p^m(1-p)^{M-m} p^n(1-p)^{N-n} dp \\
 &= \binom{M}{m} \frac{(N+1)!}{n!(N-n)!} E_{N+M-n-m}^{m+n} \\
 &= \frac{M!}{m!(M-m)!} \frac{(N+1)!}{n!(N-n)!} \frac{(m+n)!(N+M-n-m)!}{(N+M+1)!} \\
 &= \binom{n+m}{n} \binom{N+M-n-m}{N-n} \frac{1}{\binom{N+M+1}{N+1}}
 \end{aligned}$$

Esta expressão horrível pode ser simplificada em casos particulares. Por exemplo, Laplace considerou o caso em que após N eventos com n sucessos, queremos a probabilidade de $m = 1$ sucesso em $M = 1$ tentativas.

$$\begin{aligned}
 P(m = 1|n, M = 1, N, I) &= \binom{n+1}{n} \binom{N-n}{N-n} \frac{1}{\binom{N+2}{N+1}} \\
 &= \frac{(n+1)! (N+1)!}{n! (N+2)!} \\
 &= \frac{n+1}{N+2} \tag{4.38}
 \end{aligned}$$

No caso particular, mas que concentrou a atenção de estudiosos por séculos, onde temos $n = N$ sucessos em N tentativas, a probabilidade de que a próxima seja um sucesso é

$$\begin{aligned}
 P(m = 1|n = N, M = 1, N, I) &= \binom{n+1}{n} \binom{N-n}{N-n} \frac{1}{\binom{N+2}{N+1}} \\
 &= \frac{N+1}{N+2}.
 \end{aligned}$$

Este resultado recebe o nome de *regra da sucessão*. Aqui Laplace cometeu o seu maior erro, não no uso das regras da probabilidade nem de contas. Simplemente fez uma piada que foi mal entendida por muitos estudiosos que o seguiram. A estimativa bíblica da idade do universo era da ordem de 5000 anos $\approx 1.82613 \times 10^6$ dias. Em todos esses dias nasceu o sol. Qual seria a probabilidade de que o sol saísse amanhã? Pela regra da sucessão 4.38, seria $1 - 5 \times 10^{-7} = 0.9999995$. A chance de sair seria 182614 vezes maior que a de não sair. Na frase seguinte à piada, retomando um aspecto mais sério, disse que ⁶

Mas este número é incomparavelmente maior para ele que, reconhecendo na totalidade dos fenômenos o principal regulador dos dias e estações, visto que nada no momento presente pode deter a sua marcha"

Laplace.

Isto significa que deve ficar claro ao usuário, que se tiver mais informação, no caso do sol todo o conhecimento de Dinâmica e Astronomia, deve por todos os meios usá-la. O cálculo acima então

⁶ "Mais ce nombre est incomparablement plus fort pour celui qui connaissant par l'ensemble des phénomènes, le principe régulateur des jours et des saisons, voit que rien dans le moment actuel, ne peut en arrêter le cours". Essai philosophique sur les probabilités Laplace

não se deveria aplicar a não ser a situações onde se deve aplicar: àquelas em que as hipóteses são justificáveis. Os críticos à regra da sucessão por dizer que dá resultados ridículos para a saída do sol amanhã, devem responder se acham natural dizer que tudo o que sabemos sobre o sol, e o que significa que ele sairá, pode ser descrito como uma urna com dois tipos de bolas, pretos e brancos. Mas se você usar frequência como definição de probabilidade pode estar tentado a dizer que o sol sempre sairá, pois sempre saiu. Mas isto é igualmente ridículo, pois temos informação, na forma de teorias de evolução estelar que isso mudará.

Outra crítica é sobre o uso da distribuição *a priori* uniforme. Retomaremos o efeito da distribuição *a priori* no capítulo ???. As mudanças para distribuições razoáveis mudam pouco. A queixa em particular é que poderíamos fazer uma mudança não linear de variáveis e o que é uniforme agora deixaria de ser. Mas ao falar de urnas, parece natural falar do parâmetro p e se não há preferências *a priori* para acreditar num estado da urna, a uniforme parece bem justificada. Obviamente aquele que tiver informação diferente terá que fazer outras escolhas. Outras distribuições *a priori* podem e devem ser usadas, em outras condições de informação.

4.8 Poisson: um limite da binomial

Suponha que em um experimento temos N partículas que podem decair em um dado intervalo de tempo Δt e uma probabilidade p de detectar o resultado do decaimento da partícula. O tempo morto do detector é nulo. O número m de sucessos, ou detecções em Δt é dado pela binomial

$$P(m|p, N, I_2) = \binom{N}{m} p^m (1-p)^{N-m} \quad (4.39)$$

Queremos tomar o limite de N muito grande, p muito pequeno. Há várias formas de fazê-lo, um resultado extremamente útil é quando

$$N \rightarrow \infty, p \rightarrow 0, Np \rightarrow \lambda = \text{constante}$$

pois, considerando que

$$\begin{aligned} p^m \frac{N!}{(N-m)!} &= p^m N(N-1)(N-2) \cdots (N-m+1) \\ &= pN \left(1 - \frac{1}{N}\right) pN \left(1 - \frac{2}{N}\right) \cdots pN \left(1 - \frac{m-1}{N}\right) pN \\ &\rightarrow \lambda^m \end{aligned} \quad (4.40)$$

e

$$\begin{aligned} (1-p)^{N-m} &= \left(1 - \frac{\lambda}{N}\right)^{N-m} \approx \left(1 - \frac{\lambda}{N}\right)^N \\ &\rightarrow e^{-\lambda}. \end{aligned}$$

Temos então a distribuição de Poisson (que talvez deveria também ter o nome de de Moivre)

$$\begin{aligned}
 P(m|p, N, I_2) &= \frac{N!}{m!(N-m)!} p^m (1-p)^{N-m} \\
 &= \left(\frac{1}{m!}\right) \left(\frac{N!}{(N-m)!} p^m\right) \left((1-p)^{N-m}\right) \\
 \rightarrow P(m|\lambda) &= \frac{\lambda^m}{m!} e^{-\lambda}. \quad (4.41)
 \end{aligned}$$

Lembramos que o valor médio

$$\langle m \rangle = \lambda, \quad (4.42)$$

e o segundo momento

$$\langle m^2 \rangle = \lambda + \lambda^2, \quad (4.43)$$

que leva à variância

$$\sigma_{Poisson}^2 = \lambda. \quad (4.44)$$

Para calcular momentos superiores podemos usar

$$\lambda \frac{\partial}{\partial \lambda} P(m|\lambda) = -\lambda P(m|\lambda) + m P(m|\lambda) \quad (4.45)$$

pois teremos, multiplicando por m^k e somando sobre m :

$$\begin{aligned}
 \langle m^{k+1} \rangle &= \lambda \langle m^k \rangle + \sum_m \lambda \frac{\partial}{\partial \lambda} m^k P(m|\lambda) \\
 &= \lambda \langle m^k \rangle + \lambda \frac{\partial}{\partial \lambda} \langle m^k \rangle \quad (4.46)
 \end{aligned}$$

Volteremos a falar desta distribuição ao analisar dados experimentais.

4.9 Sequências Imaginadas, mãos quentes e falácia do jogador.

É fácil imaginar o experimento de lançar uma moeda. Jogo a moeda bem para o alto, bate no teto e cai no chão. Observo e anoto o resultado. Agora surge a pergunta: é fácil imaginar um segundo lançamento? Parece fácil. Se o primeiro lançamento foi, porque não seria o segundo? E cem lançamentos? Este problema foi proposto aos estudantes do primeiro curso de Probabilidades no IFUSP em 2016. OS dados brutos são apresentados na figura

Notamos imediatamente que os dados gerados por pessoas não seguem o modelo pedido. Uma pequena deriva à direita nos dados é compatível com $p = .52$ poderia ser vista nos dados, na figura 4.7, mas ainda não temos instrumentos para estimar isto. Os histogramas dos dados e da binomial são bem diferentes, figura 4.8. O histograma dos dados é muito mais estreito que o da binomial. A figura 4.9 mostra um característica interessante do processo. Escolhemos um ponto qualquer na sequencia e perguntamos se os próximos $k - 1$ tem o mesmo símbolo. Isto é, perguntamos se um dado sítio numa trajetoria é seguido por símbolos semelhantes de forma a que há uma sequência de k repetições. Fazemos isso para

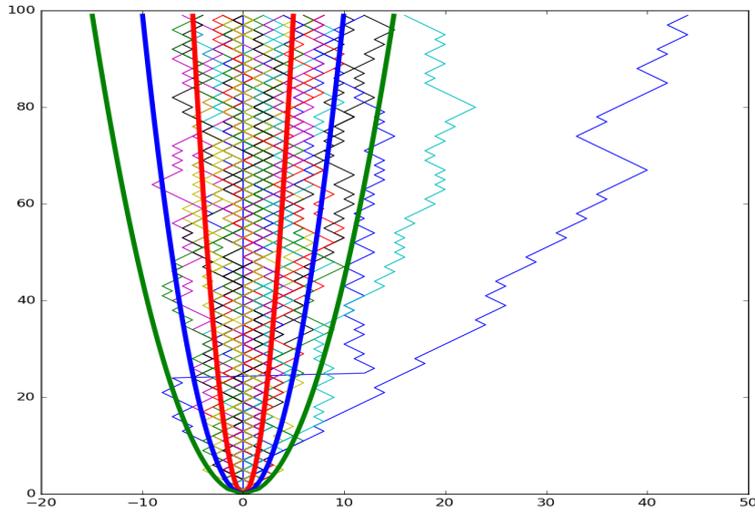


Figura 4.7: As trajetórias imaginadas por cada estudante. $K = 40$ estudantes responderam. As curvas sólidas são os desvios padrão ($\sigma, 2\sigma, 3\sigma$) que uma binomial com $p = 0.5$ teria após $N = 100$ passos.

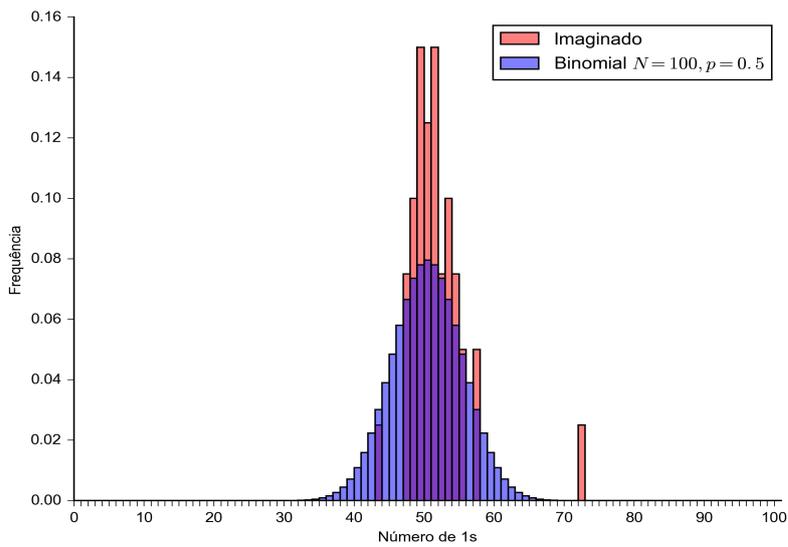
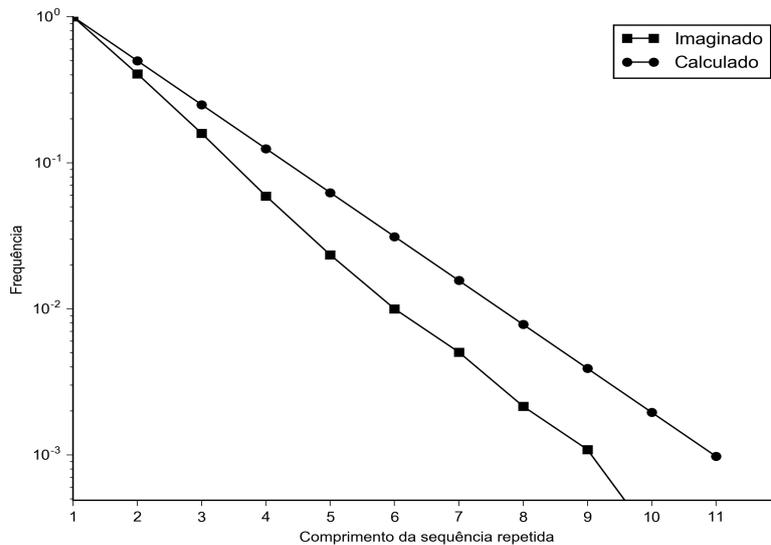


Figura 4.8: Histogramas do número de caras nos dados imaginados e o histograma do binomial simulado. $K = 40$ trajetórias de $N = 100$ jogadas.

Figura 4.9: Freqüências de seqüências de k símbolos.



todos os sítios, e fazemos isso para k de 1 até 11. O logaritmo da razão f_I entre todas as vezes que ocorre e o total de símbolos KN aparece nas ordenadas da figura 4.9. Para a binomial, é simples de calcular, a razão vai como $f_B = p^{k-1}$, pois as jogadas são independentes (círculos). Mas a linha de baixo (quadrados), para os dados, mostra que há uma repressão sistemática por parte de uma pessoa em compensar uma seqüência e inverter o símbolo imaginado. A representação dos dados em termos do logaritmo da freqüência é interessante para leis de formas exponenciais. Supomos um modelo $f = A2^{-\alpha(k-1)}$ e vemos da figura que $\alpha_B = 1$ para a binomial e $\alpha_I = 1.25$ para o processo imaginado. Explicar o valor não trivial pode ser interessante para quem estiver interessado nos aspectos psicológicos do problema. Um modelo muito simples é o de memória de um passo. Isto será retomado ao olhar para processos Markovianos. Neste caso podemos considerar como modelo apropriado um com $P(x_{t+1}|x_t)$ dado por

$$P(x_{t+1} = 1|x_t = 1) = 1 - P(x_{t+1} = -1|x_t = 1) = \frac{1}{2^\alpha} \approx 0.42$$

$$P(x_{t+1} = -1|x_t = -1) = 1 - P(x_{t+1} = 1|x_t = -1) = \frac{1}{2^\alpha} \approx 0.42$$

Se você não lembrasse do passado isso não poderia ocorrer. Ver vários símbolos iguais sugere que o próximo deve ser diferente. Imaginem o contrário, um observador olha para um processo binomial e se surpreende que houve vários símbolos iguais. O que faz? Aposta que o próximo também deve ser iguais pois "se tudo fosse normal"deveria haver uma compensação. Acredita que o processo está quente e se dispõe a apostar mais alto, porque a máquina que gera as jogadas "está quente". Perceber que

cometemos éstas falácias de análise pode ser útil para pessoas dispostas a perder dinheiro em apostas.

5

Teorema do Limite Central

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 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} .

5.1 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 distribuídos ¹. Estamos 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 5.1. 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

¹ Independentes e igualmente distribuídos: usualmente abreviado por i.i.d.

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

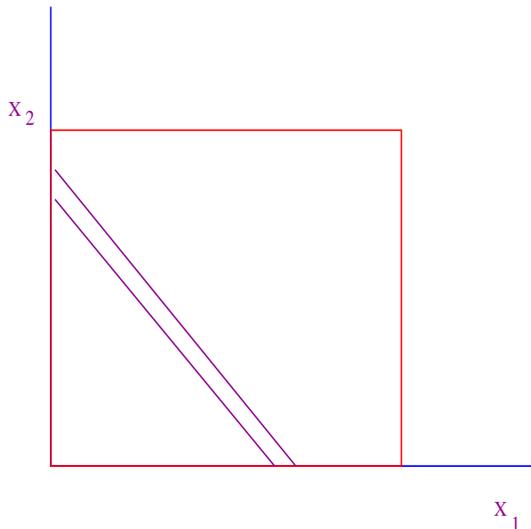


Figura 5.1: No plano X_1X_2 temos a região onde o valor de Y está entre y e $y + dy$. Todos os pares x_1 e x_2 nela contribuem para a probabilidade de Y

$$P(y|N = 2)dy = \int_{y \leq x_1 + x_2 \leq y + dy} dx_1 dx_2 P(x_1)P(x_2), \quad (5.1)$$

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 ².

$$P(y|N = 2)dy = \int \chi_{y \leq x_1 + x_2 \leq y + dy} dx_1 dx_2 P(x_1)P(x_2), \quad (5.2)$$

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 $\delta_n(A)$:

$$\begin{aligned} \delta_n(y \leq x_1 + x_2 \leq y + \Delta y_n) &= \frac{1}{\Delta y_n}, \quad \text{se } y \leq x_1 + x_2 \leq y + \Delta y_n \\ &= 0, \quad \text{se não} \end{aligned} \quad (5.3)$$

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

$$P(y|N = 2) = \int dx_1 dx_2 P(x_1)P(x_2)\delta(y - x_1 + x_2), \quad (5.4)$$

$$P(y|N = 2) = \int dx P(x)P(y - x), \quad (5.5)$$

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 :

$$P(y) = \int dx_1 dx_2 P(y, x_1, x_2). \quad (5.6)$$

² χ_A é chamada a função característica do intervalo ou conjunto A , não confunda com a função característica da distribuição de probabilidades definida abaixo.

Da regra do produto

$$P(y) = \int dx_1 dx_2 P(y|x_1, x_2)P(x_1, x_2). \quad (5.7)$$

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 5.5. A vantagem disto é que podemos facilmente obter expressões para funções gerais. Se $y = f(x_1, x_2)$, então

$$P(y) = \int dx_1 dx_2 \delta(y - f(x_1, x_2))P(x_1, x_2) \quad (5.8)$$

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) ³ e a inversa

$$\hat{P}(k) = \int_{-\infty}^{\infty} e^{-ikx} P(x)dx \quad (5.9)$$

$$P(x) = \int_{-\infty}^{\infty} e^{ikx} \hat{P}(k) \frac{dk}{2\pi} \quad (5.10)$$

³ Para que exista é suficiente ainda que P seja seccionalmente contínua em cada intervalo $[-M, N]$ e definir $\hat{P} = \lim_{N, M \rightarrow \infty} \int_{-M}^N e^{-ikx} P(x)dx$

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 5.9 e a troca da ordem de integração e somatória nos mostra que os coeficientes estão relacionados aos momentos:

$$\begin{aligned} \hat{P}(k) &= \int_{-\infty}^{\infty} \sum_{s=0}^{\infty} \frac{(-ikx)^s}{s!} P(x) \\ &= \sum_{s=0}^{\infty} \frac{(-ik)^s}{s!} \langle x^s \rangle \end{aligned} \quad (5.11)$$

Tomemos a TF dos termos da equação 5.5, e usando :

$$\delta(k) = \int \frac{dx}{2\pi} e^{ikx} \quad (5.12)$$

obtemos

$$\begin{aligned}\hat{P}(k|N=2) &= \int dy dx e^{-iky} P(x) P(y-x), \\ &= \int \frac{dx dy dk_1 dk_2}{(2\pi)^2} \hat{P}(k_1|1) \hat{P}(k_2|1) e^{-iky+ik_1x+ik_2(y-x)}\end{aligned}\quad (5.13)$$

Integrando sobre x e usando a representação da delta:

$$\begin{aligned}&= \int \frac{dy dk_1 dk_2}{2\pi} \hat{P}(k_1|1) \hat{P}(k_2|1) e^{-iky+ik_2y} \delta(k_1 - k_2), \\ &= \hat{P}(k|1) \hat{P}(k|1) = \hat{P}^2(k|N=1)\end{aligned}\quad (5.14)$$

Para a soma de $N = n$ variáveis x_i

$$P(y|N=n) = \int \prod_{i=1\dots n} dx_i P(x_1) P(x_2) \dots P(y - \sum_{i=1}^{n-1} x_i), \quad (5.15)$$

ou, introduzindo uma integral mais

$$P(y|N=n) = \int \prod_{i=1\dots n} dx P(x_1) P(x_2) \dots P(x_n) \delta(y - \sum_{i=1}^n x_i), \quad (5.16)$$

obtemos

$$\hat{P}(k|N=n) = \hat{P}^n(k|N=1) \quad (5.17)$$

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 5.17 e dado que produtos, ao tomar logaritmos, viram somas, temos

$$\log \hat{P}(k|N=n) = \sum_i^n \log \hat{P}(k|N=1) = n \log \hat{P}(k|N=1). \quad (5.18)$$

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:

$$\log \hat{P}(k|N=n) = \sum_{s=0}^{\infty} C_s(n) \frac{(-ik)^s}{s!}. \quad (5.19)$$

A equação 5.18 nos indica o motivo do nome dos cumulantes: a aditividade (ou acúmulo) ante convoluções

$$\begin{aligned}C_s(\sum_i x_i) &= \sum_i C_s(x_i), \\ C_s(n) &= n C_s(1),\end{aligned}\quad (5.20)$$

onde a equação 5.20 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

$$C_s = \frac{1}{(-i)^s} \frac{d^s \log \hat{P}}{dk^s} \Big|_{k=0} \quad (5.21)$$

Podemos calcular alguns dos primeiros,

$$\begin{aligned} \log \hat{P}(k|N=1) &= \log \int e^{-ikx} P(x) dx \\ &= \log \int \sum_{s=0}^{\infty} \frac{(-ikx)^s}{s!} P(x) dx \\ &= \log \left(1 + \sum_{s=1}^{\infty} \frac{(-ik)^s}{s!} \langle x^s \rangle \right) \\ &= \sum_{s_1=1}^{\infty} \frac{(-ik)^{s_1}}{s_1!} \langle x^{s_1} \rangle - \frac{1}{2} \sum_{s_1, s_2=1}^{\infty} \frac{(-ik)^{s_1+s_2}}{s_1! s_2!} \langle x^{s_1} \rangle \langle x^{s_2} \rangle \\ &+ \frac{1}{3} \sum_{s_1, s_2, s_3=1}^{\infty} \frac{(-ik)^{s_1+s_2+s_3}}{s_1! s_2! s_3!} \langle x^{s_1} \rangle \langle x^{s_2} \rangle \langle x^{s_3} \rangle + \dots \quad (5.22) \end{aligned}$$

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 $\langle x^s \rangle$:

$$\begin{aligned} C_0 &= 0, \\ C_1 &= \langle x \rangle, \\ C_2 &= \langle x^2 \rangle - \langle x \rangle^2, \\ C_3 &= \langle x^3 \rangle - 3 \langle x^2 \rangle \langle x \rangle + 2 \langle x \rangle^3, \\ C_4 &= \langle x^4 \rangle - 4 \langle x^3 \rangle \langle x \rangle - 3 \langle x^2 \rangle^2 + 12 \langle x^2 \rangle \langle x \rangle^2 \\ &- 6 \langle x \rangle^4, \quad (5.23) \end{aligned}$$

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:

$$nC_1^x = C_1^Y(n) = \sqrt{n} C_1^Z(n) = n C_1^W(n), \quad (5.24)$$

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

$$nC_s^x = C_s^Y(n) = n^{s/2} C_s^Z(n) = n^s C_s^W(n), \quad (5.25)$$

Portanto

$$\begin{aligned} C_s^Y(n) &= n C_s^x \\ C_s^Z(n) &= \frac{1}{n^{s/2-1}} C_s^x \\ C_s^W(n) &= \frac{1}{n^{s-1}} C_s^x, \quad (5.26) \end{aligned}$$

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

$$u_s^Y(n) = \frac{C_s^Y(n)}{(C_2^Y(n))^{s/2}} = n^{1-\frac{s}{2}} u_s^x, \quad (5.27)$$

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.

5.2 O Teorema do Limite Central I

5.2.1 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 5.19, a função característica de Y é

$$\hat{P}(k|n) = \exp\left(\sum_{s=0}^{\infty} C_s^Y(n) \frac{(-ik)^s}{s!}\right) \quad (5.28)$$

e a distribuição é

$$P(Y|N = n) = \int_{-\infty}^{\infty} \exp(iky - \sum_{s=0}^{\infty} C_s^Y(n) \frac{(-ik)^s}{s!}) \frac{dk}{\sqrt{2\pi}}. \quad (5.29)$$

Até aqui é exato. Desprezando os cumulantes de ordem superior a segunda, pela eq. 5.27

$$P(Y|N = n) = \int_{-\infty}^{\infty} \exp(iky + ikC_1^Y(n) - k^2 C_2^Y(n)/2) \frac{dk}{\sqrt{2\pi}} \quad (5.30)$$

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

$$P(Y|N = n) = \frac{1}{\sqrt{2\pi n\sigma_x^2}} e^{-\frac{(y-n\mu)^2}{2n\sigma_x^2}} \quad (5.31)$$

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

$$\begin{aligned} P(Z|N = n) &= \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(z-\sqrt{n}\mu)^2}{2\sigma_x^2}} \\ P(W|N = n) &= \frac{1}{\sqrt{2\pi\frac{\sigma_x^2}{n}}} e^{-\frac{(w-\mu)^2}{2\frac{\sigma_x^2}{n}}} \end{aligned} \quad (5.32)$$

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. 5.31 e 5.32) 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.

5.3 A média e a concentração em torno da média

A distribuição de W dada pela eq. 5.31 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 5.26, temos que $C_2^W(n) = \sigma_x^2/n$, satisfaz

$$\begin{aligned} C_2^W(n) &= \int_{-\infty}^{\infty} dw(w^2 - \langle w \rangle^2)P(W = w|N = n) \\ &= \int_{-\infty}^{\infty} dw(w - \langle w \rangle)^2 P(W = w|N = n) \\ &\geq \int_{|w - \langle w \rangle| \geq \epsilon} dw(w - \langle w \rangle)^2 P(W = w|N = n) \\ &\geq \int_{|w - \langle w \rangle| \geq \epsilon} dw \epsilon^2 P(W = w|N = n) \\ &\geq \epsilon^2 \text{Prob}(|w - \langle w \rangle| \geq \epsilon). \end{aligned} \quad (5.33)$$

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 :

$$\text{Prob}(|w - \langle w \rangle| \geq \epsilon) \leq \frac{C_2^W(n)}{\epsilon^2} \quad (5.34)$$

Mas $C_2^W(n)$ depende de maneira simples de n . Extraindo 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:

$$\text{Prob}(|w - \langle w \rangle| \geq \epsilon) \leq \frac{C_2^x}{n\epsilon^2} \quad (5.35)$$

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' onde p, q tem outro valores (entre zero e um). Consideremos as jogadas independentes, 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 $f = m/n$

⁴ por 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.

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

$$P(m|N = n I') = \frac{n!}{m!(n-m)!} p^m q^{n-m} \quad (5.36)$$

- **(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.

5.4 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 5.29 que foram *desprezados* anteriormente para chegar até a equação 5.30. 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.

5.5 O Teorema do Limite Central III

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

5.5.1 A distribuição uniforme

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

$$\hat{P}(k|1) = \frac{1}{L} \int_{-L/2}^{L/2} e^{-ikx} dx = \frac{2}{kL} \sin\left(\frac{kL}{2}\right) \quad (5.37)$$

$$P(Y = y|N = n) = \int_{-\infty}^{\infty} [\hat{P}(k|1)]^n e^{iky} \frac{dk}{2\pi} = \int_{-\infty}^{\infty} \left[\frac{2}{kL} \sin\left(\frac{kL}{2}\right)\right]^n e^{iky} \frac{dk}{2\pi} \quad (5.38)$$

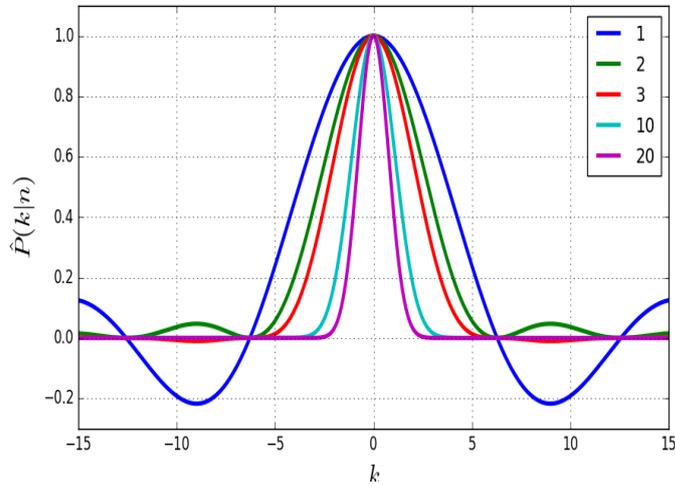


Figura 5.2: A função característica $\hat{P}(k|N = n)$ para a soma de n variáveis uniformemente distribuídas.

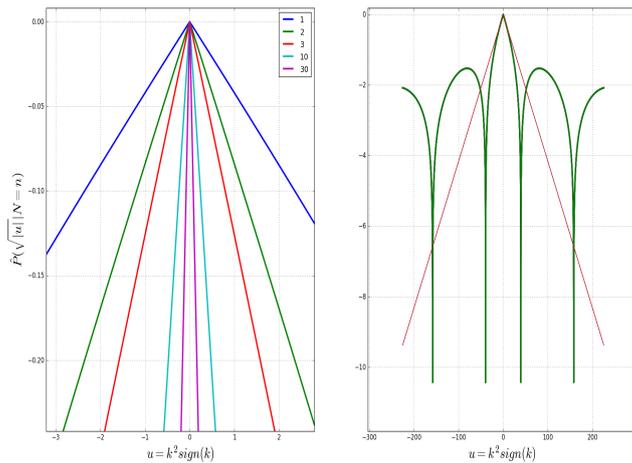


Figura 5.3: A função característica $\log(|\hat{P}(\sqrt{|u|}|N = n)|)$ como função de $u = k^2 \text{sign}(k)$, para a soma de n variáveis uniformemente distribuídas. Nesta representação gaussianas aparecem como retas $\propto -\text{abs}(u)$. Esquerda: na região central parecem gaussianas. Direita: fora da região central diferem de gaussianas. Nesta figura dividimos por \sqrt{n} e todas as curvas colapsam. As retas são para a gaussiana mais próxima $\sigma^2/2n = 1/24$. Note a diferença da escala de u nas abscissas.

A figura 5.2 mostra que a função característica fica mais parecida com uma gaussiana e na figura 5.3 vemos que efetivamente o $\log(|\hat{P}(\sqrt{|u|}|N = n)|)$ com $u = \pm k^2$ fica cada vez mais perto de $-\sigma^2|u|$ (gaussiana).

Exercício Mostre que $\sigma^2/2 = n/24$.(verificar??)

5.5.2 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

$$\hat{P}(k|1) = \int_0^\infty ae^{-xa}e^{-ikx}dx = \frac{a}{a + ik} \tag{5.39}$$

$$P(Y = y|N = n) = \int_{-\infty}^\infty \left(\frac{a}{a + ik}\right)^n e^{iky} \frac{dk}{2\pi} \tag{5.40}$$

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

$$P(Y = y|N = n) = \Theta(y)a^n \frac{y^{n-1}e^{-ay}}{(n-1)!}. \tag{5.41}$$

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

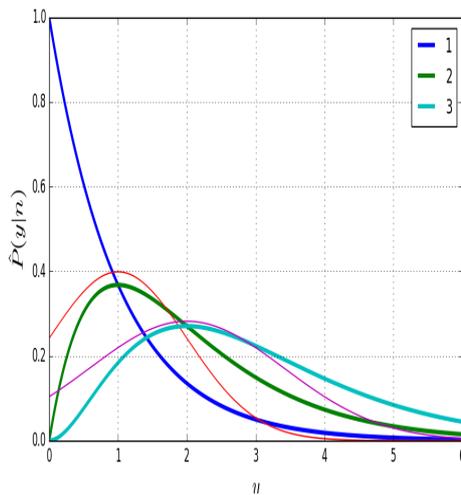


Figura 5.4: A densidade de probabilidade $P(y|N = n, a = 1)$ para a soma de n variáveis exponencialmente distribuídas, $n = 1, 2, 3$. Para as duas últimas mostramos as gaussianas com $\mu = \sigma^2 = n - 1$

A figura 5.4 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 5.5. 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.

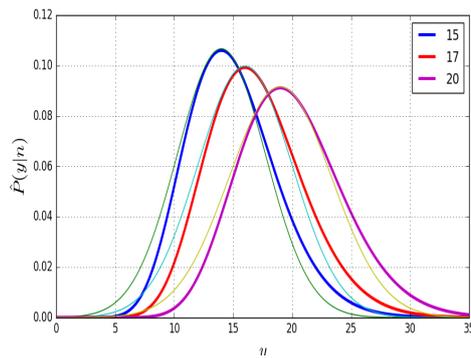


Figura 5.5: A densidade de probabilidade $P(y|N = n, a = 1)$ para a soma de n variáveis exponencialmente distribuídas, $n = 15, 17, 20$. Junto estão mostradas as gaussianas com $\mu = \sigma^2 = n - 1$.

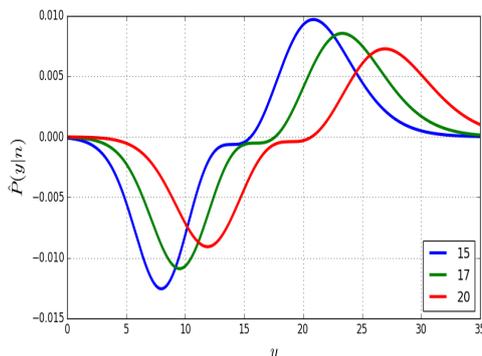


Figura 5.6: A diferença entre a densidade de probabilidade $P(y|N = n, a = 1)$ para a soma de n variáveis exponencialmente distribuídas, $n = 15, 17, 20$ e as gaussianas com $\mu = \sigma^2 = n - 1$. Os mesmos parâmetros da figura anterior. Note que a região central é bem aproximada. Há uma região de transição, ao afastar-se para as caudas, e finalmente as caudas vão rapidamente para zero, assim como a sua diferença

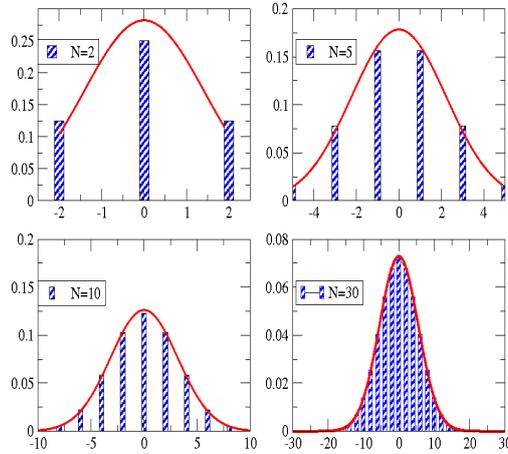
5.5.3 A distribuição binomial revisitada

A distribuição de Bernoulli é dada por $P(x) = p\delta(x - 1) + q\delta(x + 1)$. O número de aplicações que usam 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 $Y = \sum_{i=1 \dots n} x_i$. 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 5.7 a binomial foi dividida por ΔY . De outra 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 $y + 2$.

5.5.4 Caminho Aleatório

Novamente olhamos para a distribuição binomial. Olhe para a figura 5.8. Definimos o caminho aleatório através de Difusão: K (= 10000 na figura 5.8) seqüências de N passos de um

Figura 5.7: A binomial (dividida por $\Delta Y = 2$, barras) e a densidade gaussiana correspondente (linha contínua), para $N = 2, 5, 10$ e 30



processo binomial, definidos por

$$y_n = y_{n-1} + x_n \tag{5.42}$$

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,

$$\log N! = \left(N + \frac{1}{2}\right) \log N - N + \frac{1}{2} \log 2\pi + \log \left(1 + \frac{1}{12N} + \frac{1}{288N^2} + \mathcal{O}\left(\frac{1}{N^3}\right)\right) \tag{5.43}$$

que é uma expansão assintótica, isto é melhora quando N aumenta

⁵. Não precisamos todos esses termos, basta $\log N \approx N \log N - N$ onde desprezamos $\mathcal{O}(\log N)$

⁵ Jeffreys, 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.

$$\begin{aligned} \log P(m|N = nI') &= \log n! - \log m! - \log(n - m)! + m \log p + (n - m) \log q \\ &= n \log n - n - m \log m + m - (n - m) \log(n - m) \\ &\quad + n - m + m \log p + (n - m) \log q \end{aligned} \tag{5.44}$$

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

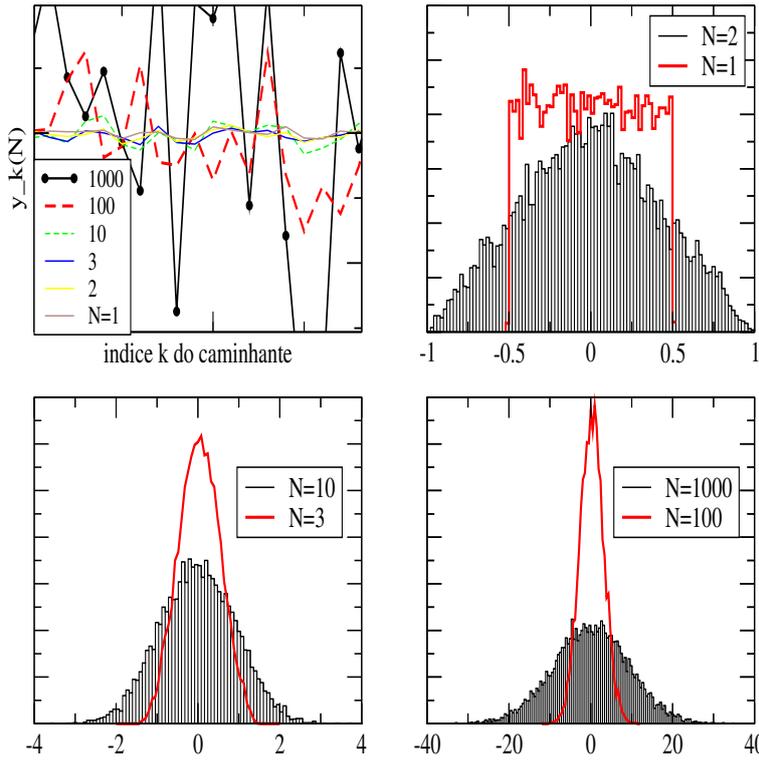


Figura 5.8:

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 5.27, 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

$$P(m|N = nI) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2\sigma^2} (m - m^*)^2 \quad (5.45)$$

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

6

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 também está relacionado à atualização de 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 ou atribuir 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, escolhemos 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.

6.1 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 certa informação que pode não ser verdadeira. É melhor não saber do que achar que se sabe algo que está errado ¹.

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 nossa 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.

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

¹ He who knows best knows how little he knows. Thomas Jefferson

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 $\{x_i\}_{k=1\dots n}$. Estes *microestados* são mutuamente exclusivos. Queremos atribuir as probabilidades p_k . 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. Dentro de um grupo g os estados são indexados por i . Podemos escrever

$$k(i, g) = \sum_{g'}^{g-1} m_{g'} + i$$

e fixos os $\{m_g\}$, dado g e i obtemos k . A probabilidade de que X tenha um valor x_k é dada pela regra do produto em termos das probabilidades de estar em um grupo e de ter um índice i dado que está num grupo:

$$p_k = P(k) = P(i, g | \{m_g\}) = P(i | g \{m_g\}) P(g | \{m_g\})$$

e pela regra da soma, a probabilidade que esteja no grupo g , por $P(g | \{m_g\}) = \sum_{i \in g} P(i | g \{m_g\}) = P_g$ assim

$$P(i | g \{m_g\}) = \frac{P(k)}{P(g | \{m_g\})}.$$

Se for dado que está no grupo g , a probabilidade que esteja no estado i é $p(i | g) = \frac{p_{k(i,g)}}{P_g}$. 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

$$H[p_k] = H_G[P_g] + \sum_{g=1}^N P_g H_g[p(i|g)] \quad (6.1)$$

Que isto é suficiente para determinar a forma funcional da

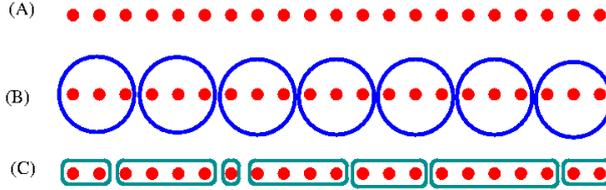


Figura 6.1: (A) Cada ponto (vermelho) representa um possível estado (mutuamente exclusivos) de uma variável. (B) Agrupamos em grupos com o mesmo tamanho. (C) Os agrupamentos são arbitrários.

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 6.1.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 $p(i|g) = N/n = 1/m$

$$H[1/n \dots 1/n] = H_G[1/N, \dots, 1/N] + \sum_{g=1}^N \frac{1}{N} H_g[1/m, \dots, 1/m] \quad (6.2)$$

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

$$F(n) = F(N) + \sum_{g=1}^N \frac{1}{N} F(m) \quad (6.3)$$

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

$$F(mN) = F(N) + F(m) \quad (6.4)$$

A solução óbvia é dada por

$$F(n) = k \log n. \quad (6.5)$$

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 6.1.C), salvo o fato $\sum_{g=1}^N m_g = n$, mas ainda temos que $p_k = 1/n$ é uniforme. Temos então que a probabilidade $P_g = m_g/n$ é arbitrária, mas a probabilidade condicional dentro de cada grupo é uniforme: $p(i|g) = 1/m_g$

$$H[1/n, \dots, 1/n] = H_G[P_g] + \sum_{g=1}^N P_g H_g[1/m_g, \dots, 1/m_g] \quad (6.6)$$

e substituímos a entropia da distribuição uniforme:

$$F(n) = H_G[P_g] + \sum_{g=1}^N P_g F(p(i|g)), \quad (6.7)$$

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

$$H_G[P_g] = F(n) - \sum_{g=1}^N P_g F(m_g). \quad (6.8)$$

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

$$H_G[P_g] = \left(\sum_{g=1}^N P_g \right) k \log n - k \sum_{g=1}^N P_g \log(m_g) \quad (6.9)$$

$$H_G[P_g] = -k \sum_{g=1}^N P_g \log(m_g/n) \quad (6.10)$$

e finalmente

$$H_G[P_g] = -k \sum_{g=1}^N P_g \log(P_g) \quad (6.11)$$

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 ² 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 ³.

² Aparentemente 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.

³ Não cabe no curso introdutório de Mecânica Estatística

6.1.1 Um método variacional

O caminho para as aplicações está aberto. O método de escolha das distribuições de probabilidade é de maximizar a ignorância. De todas as distribuições compatíveis com os vínculos, escolhemos a que introduz a menor quantidade de informação adicional aos vínculos. Reduzimos a inferência a um método variacional. Isto não deve vir como uma surpresa, pois toda a Física é essencialmente redutível a métodos variacionais. A pergunta profunda é se esses métodos estão relacionados ou se são independentes. A resposta que está emergindo é que todos esses métodos estão relacionados. E isto transcende a Física e se estende por outras áreas. 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 ⁴. Em particular tem se conseguido nos últimos anos uma formulação muito mais satisfatória e menos ad hoc dos princípios por trás deste método.

⁴ ver A.C

A generalização do método de máxima entropia para inferência em problemas com variáveis em variedades contínuas pode levar a problemas conceituais que serão atacados a seguir.

6.2 Entropia Relativa

As deduções aqui deveriam ser lidas após treinamento sobre métodos variacionais, que será abordado nos próximos capítulos. Portanto talvez seja recomendável pular esta seção numa primeira leitura. A entropia de Shannon vai ser usada para inferências em quase todo o resto do livro. Mas esta não é a última palavra em mecanismos de inferência. A utilidade de entropia de Shannon foi a de *atribuição* de probabilidades. Isso deve dar a impressão que do nada, as probabilidades são construídas através da imposição de certos vínculos sobre funções dos graus de liberdade. Dificilmente a inferência é feita a partir de nenhum conhecimento prévio. Para

qualquer situação é mais razoável que um estado de crenças seja modificado pela imposição de novos vínculos. Isso nos leva à definir Informação como tudo aquilo que leva a modificar as crenças codificadas em distribuições de probabilidades. Procuramos um mecanismo que sirva para *atualizar* as crenças a partir de uma crença *a priori*. A diferença importante pode ser vista no confronto de *atualizar* a partir de um estado de conhecimento *apriori* versus *atribuir* a partir do vácuo informacional.

Qual deve ser o mecanismo de mudar crenças? Há em geral muitas distribuições de probabilidade que satisfazem um certo conjunto de vínculos. A idéia é fazer um ranking de todas elas. Isto significa atribuir um número a cada distribuição de probabilidades e escolher aquela que maximiza o ranking de preferências. Ou seja devemos construir um funcional que atribui um número a cada distribuição. Claro que esse número deve depender no estado de crença antes da imposição dos novos vínculos. Como escolher o funcional? Novamente procuramos casos simples onde sabemos algo. Impomos como desejo que o funcional dê a resposta esperada neles. O método de obtenção do funcional é o de eliminação de todos os funcionais que falham em satisfazer esses casos simples. Resta algum funcional após este processo? Poderia ser que não, mas não é o caso. Chamaremos este funcional de Entropia Relativa $S[p||q]$, entre a distribuição p candidata a ser escolhida pelo processode inferência e a distribuição *a priori* q . Também pode ser chamado de (menos) a divergência de Kullback-Leibler (KL). Por economia, será simplesmente chamada entropia. A menos de pequenos detalhes, a dedução segue A. Caticha de forma simplificada.

6.2.1 *Desiderata Entrópica*

Os desejos expostos a seguir são considerados em adição aos do Capítulo 1. Portanto a teoria de probabilidades é a estrutura matemática adequada. Mas agora faremos algumas demandas adicionais. Pedimos ao leitor que encontre argumentos contra os seguintes desejos para uma teoria de infêrencia:

- DE_1 : **Generalidade**: O método deve ser Universal.

Queremos lidar com informação incompleta e o algoritmo de procedimento deve ser o mesmo independente de qual o tipo de problema de inferência que estamos tratando. A descrição do problema, através da escolha dos graus de liberdade apropriados e dos vínculos impostos levará em conta a natureza explícita do problema.

- DE_2 : **Parcimônia** A atualização deve ser minimalista.

De todas as distribuições $p(x)$ que satisfazem o vínculo devemos escolher a que menos mude a distribuição *a priori* $q(x)$. Claro que o que quer dizer mínimo deve ser definido. Mas há um caso simples

em que mínimo é fácil de decidir. No caso em que não há nova informação, então a menor mudança é não fazer nada. Nosso método deve ser tal que $p(x) = q(x)$ após a incorporação de nenhuma informação. Parece muito trivial mas será útil.

- DE_3 : **Localidade** no espaço de configurações.

Quando o espaço de configurações pode ser dividido em duas partes disjuntas, informação que diz respeito somente a uma das partes, ao ser incorporada, não deve alterar a atribuição de probabilidades às configurações da outra.

- DE_4 : **Invariância**: A escolha dos rótulos dos graus de liberdade é uma convenção: não deve alterar o resultado da Inferência.

Para sistemas de variáveis que tomam valores num sub espaço de \mathbb{R}^N , a escolha do sistema de coordenadas não deve interferir no resultado do problema. Assim o funcional deve ser invariante ante mudanças (contínuas diferenciáveis) do sistema de coordenadas.

- DE_5 : **Independência**: Sistemas independentes devem ser descritos da mesma forma quando estudados separadamente ou em conjunto.

Há situações em que sabemos que há sistemas independentes. Consideremos uma garrafa de café e a galáxia X9. Ao fazer predições sobre a termodinâmica do café podemos (i) tratar o sistema de café sozinho, ou (ii) incluir os graus de liberdade da galáxia X9 além dos do café, e marginalizar a distribuição, integrando sobre as configurações da galáxia. Os dois métodos devem dar o mesmo resultado sob pena de ao realizar medidas experimentais, poder distinguir entre os casos em que a galáxia foi ou não incluída. Note que poderíamos, caso dessem diferentes, fazer astronomia olhando para uma garrafa de café.

6.2.2 DE_3 : Localidade

Começamos com as consequências de DE_3 . As configurações de um sistema formam o conjunto $\chi = \{x_1, x_2, \dots, x_n\}$. Podemos separar em dois conjuntos $i \in A$ e $j \in B$, tal que $A \cap B = \emptyset$ e $A \cup B = \chi$, ou seja mutuamente exclusivos e exaustivos. Um caso simples é quando a informação é dada através dos seguintes vínculos

$$P_A = \sum_{x_i \in A} p_i, \quad (6.12)$$

$$P_B = \sum_{x_j \in B} p_j, \quad (6.13)$$

$$\text{com } P_A + P_B = 1, \quad (6.14)$$

$$G = \sum_{x_j \in B} p_j g(x_j), \quad (6.15)$$

onde $p_k = p(x_k)$. Maximizar a entropia sujeita a estes vínculos leva a

$$0 = \delta \left[S - \lambda_1 \left(\sum_A p(x_i) - P_A \right) - \lambda_2 \left(\sum_B p(x_j) - P_B \right) - \lambda_3 \left(\sum_B p(x_j)g(x_j) - G \right) \right]$$

para variações independentes de quaisquer dos p . Temos n equações:

$$\frac{\partial S}{\partial p_i} = \lambda_1 \quad (6.16)$$

$$\frac{\partial S}{\partial p_j} = \lambda_2 + \lambda_3 g(x_j). \quad (6.17)$$

$$(6.18)$$

Em geral teríamos, para $k \in A$ ou $k \in B$, chamando $\frac{\partial S}{\partial p_k} = f_k$, a dependência

$$\frac{\partial S}{\partial p_k} = f_k(p_1, \dots, p_n, q_1, \dots, q_n). \quad (6.19)$$

Comparando as equações 6.16 e 6.19 vemos que

$$\lambda_1 = f_i(p_1, \dots, p_n, q_1, \dots, q_n) \quad (6.20)$$

$$= f_i(p_i, q_i). \quad (6.21)$$

A segunda equação decorre de que mudando a informação que altera p_j do conjunto B , poderíamos mudar o lado direito sem alterar o multiplicador de Lagrange que é constante. Ainda mais, podemos considerar todas as possíveis divisões de χ em subconjuntos exaustivos e mutuamente exclusivos. Eliminaríamos a dependência em todos os índices diferentes de i . Assim

$$\frac{\partial S}{\partial p_k} = f_k(p_k, q_k) \quad (6.22)$$

Integrando, vemos que a forma geral do funcional é

$$S = \sum F_k(p_k, q_k)$$

onde F_k com $f_k = \frac{\partial F_k}{\partial p_k}$ são funções ainda desconhecidas. A generalização para variáveis que tomam valores reais:

$$S = \int F(p(x), q(x), x) dx. \quad (6.23)$$

6.2.3 DE₄: Invariância

Vamos começar com um problema em uma dimensão. Uma densidade de probabilidades de uma variável X , $m(x)$ satisfaz

$$\int m(x) dx = 1$$

Podemos mudar variáveis $x \rightarrow x'$ e a densidade mudará $m(x) \rightarrow m'(x')$. A probabilidade de uma região dx em x deve ser preservada, portanto

$$m(x)dx = m'(x')dx'$$

e isso deve valer para a transformação de qualquer densidade. Em mais de uma dimensão teremos

$$m(x) = m'(x')J(x')$$

onde $J(x') = \det \left| \frac{\partial x'}{\partial x} \right|$ é o Jacobiano da transformação $x \rightarrow x'$.

A estratégia a seguir é a seguinte: fazemos a inferência no sistema de coordenadas x , obtendo um resultado. A seguir, a fazemos no referencial x' . Transformamos os x' de volta a x e comparamos. Os dois resultados deveriam dar igual. Mas não ocorre a não ser que imponhamos condições sobre a função F que aparece na equação 6.23. Isso restringe ainda mais os funcionais sobreviventes. Antes de prosseguir, facilita introduzir uma densidade $m(x)$ que por agora está à nossa disposição escolher. A motivação é que a razão de duas densidades tem propriedades simples de transformação

$$\frac{p(x)}{m(x)} = \frac{p'(x')}{m'(x')} \tag{6.24}$$

pois o Jacobiano se cancela ⁵. Assim consideramos que a equação 6.23 pode ser escrita

⁵ Cabe ressaltar que deveríamos ter cuidado de afirmar que as regiões em que as diferentes distribuições se anulam são as mesmas e são excluídas destas considerações.

$$\begin{aligned} S &= \int F(p(x), q(x), x)dx = \int \frac{1}{m(x)} F\left(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}, x\right)m(x)dx, \\ &= \int \Phi\left(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}, m(x), x\right)m(x)dx, \end{aligned} \tag{6.25}$$

onde introduzimos a ainda desconhecida função

$$\Phi(u, v, m, x) = \frac{1}{m} F(um, vm, x).$$

Queremos mudar as coordenadas de forma geral, mas é conveniente se restringir a um caso fácil. O vínculo imposto será

$$\int p(x)a(x) = A,$$

onde a função $a(x)$ é um escalar, o que significa que $a(x) \rightarrow a'(x') = a(x)$, não muda ante transformações de coordenadas. A normalização é um exemplo de vínculo escalar, portanto não é preciso considerá-lo separadamente. O cálculo variacional $\delta(S + \text{vínculos}) = 0$, leva a

$$\begin{aligned} 0 &= \frac{\delta}{\delta p(x)} \left(S - \lambda \int p(x)a(x)dx \right), \\ &= \dot{\Phi}\left(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}, m(x), x\right) - \lambda a(x), \end{aligned} \tag{6.26}$$

onde

$$\dot{\Phi}(u, v, m, x) = \frac{d}{du} \Phi(u, v, m, x).$$

Analogamente, começando no sistema de coordenadas x' , chegamos a

$$0 = \dot{\Phi}\left(\frac{p'(x')}{m'(x')}, \frac{q'(x')}{m'(x')}, m(x'), x'\right) - \lambda' a'(x'). \quad (6.27)$$

Usando a equação 6.24, esta última equação pode ser reescrita como

$$0 = \dot{\Phi}\left(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}, m'(x'), x'\right) - \lambda' a'(x'). \quad (6.28)$$

Para vínculos escalares, segue que, a razão

$$\frac{\dot{\Phi}\left(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}, m(x), x\right)}{\dot{\Phi}\left(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}, m'(x'), x'\right)} = \frac{\lambda}{\lambda'} \quad (6.29)$$

é uma constante, pois os multiplicadores de Lagrange não dependem de x nem x' . Como isto vale para qualquer transformação contínua e diferenciável de $x' = \phi(x)$, podemos olhar para diferentes escolhas de ϕ e eliminar candidatos. Primeiro, olhamos para transformações com Jacobiano igual a 1 em todo o espaço. Segue que $m(x) = m'(x')$, portanto a única diferença entre o numerador e o denominador é no quarto argumento. Como a razão é constante, $\dot{\Phi}$ não pode depender do quarto argumento x ou x' . Agora consideramos, novamente com Jacobiano igual a 1, a função $\phi(x) = x, x \in \bar{D}$ e $\phi(x) \neq x, x \in D$. Se $x \in \bar{D}$, a razão é 1:

$$\begin{aligned} 1 &= \frac{\dot{\Phi}\left(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}, m(x)\right)}{\dot{\Phi}\left(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}, m'(x')\right)} \\ &= \frac{\lambda}{\lambda'} \end{aligned}$$

Mas para $x \in D$, não muda a razão. Portanto $\dot{\Phi}$ não pode depender agora do seu terceiro argumento e

$$1 = \frac{\dot{\Phi}\left(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}\right)}{\dot{\Phi}\left(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}\right)} \quad (6.30)$$

é trivialmente satisfeita. Segue que o funcional só pode ter esta estrutura:

$$S = \int \dot{\Phi}\left(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}\right) m(x) dx. \quad (6.31)$$

6.2.4 DE_2 : Parcimônia

Temos à nossa disposição este intruso $m(x)$. Podemos escolher qualquer densidade? Impondo o desejo de parcimônia veremos que não é possível essa liberdade. Estamos interessados em não fazer nada. O aluno não deve se entusiasmar desnecessariamente. Se não

chega informação, não deveria mudar nossa escolha de $p(x)$.
 Portanto a solução do problema variacional

$$\begin{aligned}\frac{\delta}{\delta p(x)} \left(S - \lambda \int p(x) dx \right) &= 0, \\ \dot{\Phi} \left(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)} \right) &= \lambda,\end{aligned}\quad (6.32)$$

deve levar à escolha $p(x) = q(x)$. Mas nem todo funcional do tipo 6.31 leva a esse resultado. Isto restringe mais a família de funcionais. O problema 6.32 é uma equação diferencial ordinária

$$\begin{aligned}\Phi(u, v) &= \lambda \\ \Phi(u, v) &= \lambda u + c(v)\end{aligned}$$

portanto, se seguíssemos este caminho teríamos

$$\begin{aligned}S &= \int \left(\lambda \frac{p(x)}{m(x)} + c \left(\frac{q(x)}{m(x)} \right) \right) m(x) dx \\ &= \lambda \int p(x) dx + \text{const},\end{aligned}$$

o que é uma trivialidade, pois a normalização diz que isto é uma constante e toda distribuição normalizada teria a mesma preferência. Portanto escolher a função Φ leva a um beco sem saída. Mas há outra saída para satisfazer a equação 6.32, basta escolher $m(x) = q(x)$, neste caso

$$\dot{\Phi} \left(\frac{p(x)}{q(x)}, \frac{q(x)}{q(x)} \right) = \lambda, \quad (6.33)$$

que é constante de forma trivial quando $p(x) = q(x)$. Novamente temos um avanço, restringindo ainda mais as estruturas possíveis:

$$S = \int \Phi \left(\frac{p(x)}{q(x)} \right) q(x) dx. \quad (6.34)$$

6.2.5 DE_5 : Independência

Este último desejo será usado para determinar a forma explícita de Φ na equação 6.34. Além da normalização, para o sistema composto pela garrafa de café, com graus de liberdade x_1 e a galáxia X9 com graus de liberdade x_2 , temos os seguintes vínculos

$$\begin{aligned}\int p_1(x_1) f_1(x_1) dx_1 &= F_1 \\ \int p_2(x_2) f_2(x_2) dx_2 &= F_2.\end{aligned}$$

Tratando os dois sistemas de forma independente temos que $p_1(x_1)$ e $p_2(x_2)$ são dados pelas soluções de

$$\dot{\Phi} \left(\frac{p_1(x_1)}{q_1(x_1)} \right) = \lambda_{01} + \lambda_1 f_1(x_1) \text{ e } \dot{\Phi} \left(\frac{p_2(x_2)}{q_2(x_2)} \right) = \lambda_{02} + \lambda_2 f_2(x_2) \quad (6.35)$$

Tratando os sistemas juntos, o problema variacional é

$$0 = \frac{\delta}{\delta p(x_1, x_2)} \left\{ S - \mu_0 \left(\int p(x_1, x_2) dx - 1 \right) \right. \\ \left. - \mu_1 \left(\int f_1(x_1) p(x_1, x_2) dx_1 dx_2 - F_1 \right) \right. \\ \left. - \mu_2 \left(\int f_2(x_2) p(x_1, x_2) dx_1 dx_2 - F_2 \right) \right\},$$

$$\dot{\Phi} \left(\frac{p(x_1, x_2)}{q_1(x_1)q_2(x_2)} \right) = \mu_0 + \mu_1 f_1(x_1) + \mu_2 f_2(x_2) \quad (6.36)$$

Impomos que a solução deve satisfazer $p(x_1, x_2) = p_1(x_1)p_2(x_2)$ e definimos $y = \frac{p_1(x_1)p_2(x_2)}{q_1(x_1)q_2(x_2)}$. Assim

$$\dot{\Phi}(y) = \mu_0 + \mu_1 f_1(x_1) + \mu_2 f_2(x_2). \quad (6.37)$$

Derivamos duas vezes, primeiro com respeito a x_2

$$\ddot{\Phi}(y) \frac{\partial y}{\partial x_2} = (\mu_2 f_2(x_2))', \quad (6.38)$$

e depois com respeito a x_1

$$\ddot{\Phi}(y) \frac{\partial y}{\partial x_1} \frac{\partial y}{\partial x_2} + \dot{\Phi}(y) \frac{\partial^2 y}{\partial x_1 x_2} = 0. \quad (6.39)$$

Isto parece complicado mas é bem simples. Note que, com a linha indicando a derivada com respeito a x_1 ou x_2

$$\frac{\partial y}{\partial x_1} = \left(\frac{p_1(x_1)}{q_1(x_1)} \right)' \frac{p_2(x_2)}{q_2(x_2)} \\ \frac{\partial y}{\partial x_2} = \frac{p_1(x_1)}{q_1(x_1)} \left(\frac{p_2(x_2)}{q_2(x_2)} \right)' \\ \frac{\partial^2 y}{\partial x_1 x_2} = \left(\frac{p_1(x_1)}{q_1(x_1)} \right)' \left(\frac{p_2(x_2)}{q_2(x_2)} \right)'. \quad (6.40)$$

Multiplique as duas primeiras e divida pela terceira equação, segue que $\Phi(y)$ satisfaz a equação diferencial ordinária:

$$y \ddot{\Phi}(y) + \dot{\Phi}(y) = 0. \quad (6.41)$$

A solução geral contém três constantes arbitrárias

$$\Phi(y) = Ay \log y + By + C. \quad (6.42)$$

A família de funcionais entropia é reduzida a

$$S[p||q] = A \int p(x) \log \frac{p(x)}{q(x)} dx + B \int p(x) dx + C \int q(x) dx.$$

A constante C pode ser tomada igual a zero, pois simplesmente adiciona uma constante sem mudar preferências. A constante B também pode ser tomada zero, pois a normalização de $p(x)$ será sempre um vínculo. Finalmente e demonstrando otimismo os

físicos tomam $A < 0$ negativo para que o processo variacional seja o de maximização. Podemos tomar $A > 0$ e minimizar o que é chamado de divergência de Kullback-Leibler. A escolha de $A = -1$ é simplesmente uma escolha de unidades da entropia. Como veremos isto levará a que a temperatura e a energia sejam medidos nas mesmas unidades. O resultado final é dado por

$$S[p||q] = - \int p(x) \log \frac{p(x)}{q(x)} dx \quad (6.43)$$

É fácil mostrar que os resultados $p_1(x_1)p_2(x_2)$ usando a equação 6.35 e $p(x_1, x_2)$ da equação 6.36, coincidem.

Exercício Para variáveis discretas mostre que $H_{SJ} = - \sum_i p_i \log p_i$ difere por uma constante de $S[p||q] = - \sum_i p_i \log \frac{p_i}{q_i}$, se $q_i = 1/N$ é constante e portanto o valor das distribuições que as extremizam coincidem.

Isto permite concluir que a entropia de Shannon faz referência a uma distribuição *a priori*, a distribuição uniforme. O fato de não aparentar de forma explícita não quer dizer que não esteja lá. É claro que o ponto onde a distribuição uniforme faz a sua entrada no raciocínio é quando a função $F(n)$ crescente é introduzida.

Exercício Suponha \mathcal{F} uma família paramétrica de densidades, e dois membros $p_1 = p(x|\theta_1)$ e $p_2 = p(x|\theta_2)$ de \mathcal{F} . Suponha que os parâmetros θ estejam em algum subconjunto de \mathbb{R}^K . Mostre que em geral $S[p_1||p_2] \neq S[p_2||p_1]$, mas para valores pequenos da distância Euclideana, $|\theta_1 - \theta_2|$ há simetria de S ante $\theta_1 \leftrightarrow \theta_2$. Isso permite definir uma distância entre duas densidades e uma geometria Riemanniana no espaço dos θ , que agora passa a ser mais que um simples subconjunto de \mathbb{R}^K , mas o que é conhecido como uma variedade estatística. Considere $\theta_2 = \theta_1 + d\theta$ e expanda a entropia até segunda ordem em $d\theta$. Dessa forma é possível definir o tensor métrico $g_{\mu\nu}$. Encontre uma expressão para o tensor métrico que é conhecido pelo nome de Fisher-Rao. Encontre a distância entre duas gaussianas multivariada com a mesma covariância e médias diferentes.

6.3 Apêndice: Multiplicadores de Lagrange

Uma montanha é descrita pela altura $z = f(x, y)$, com x e y as coordenadas de um ponto de altura z e f um função suficientemente bem comportada, tem um máximo que pode ser obtido igualando as derivadas parciais de f a zero. Suponha que uma estrada cruza a superfície da montanha e as coordenadas da estrada são descritas por $\phi(x, y) = c$. Qual é a altura máxima de um carro que viaja pela estrada?

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 :

$$F_\lambda(x, y) = f(x, y) + \lambda(\phi(x, y) - c) \quad (6.44)$$

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_{livre} mas para a função $F_\lambda(x, y)$. O problema é novamente simples⁶. Resolvemos o sistema

$$\frac{\partial F_\lambda}{\partial x} = 0 \quad (6.45)$$

$$\frac{\partial F_\lambda}{\partial y} = 0, \quad (6.46)$$

onde λ ainda não foi especificado. A resposta depende do valor escolhido para λ , isto é define uma curva ρ , parametrizada por λ : $(x^*(\lambda), y^*(\lambda))$ 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 quando se 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.

⁶ a não ser que não seja....

7

Entropia: como usar

7.1 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

$$\sum P_i = 1 \quad (7.1)$$

$$\sum x_i P_i = \mu \quad (7.2)$$

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

$$H[P] - \lambda_1 \{ \sum x_i P_i - \mu \} - \lambda_0 \{ \sum P_i - 1 \} \quad (7.3)$$

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

$$\delta [- \sum P_i \log P_i - \lambda_1 \{ \sum x_i P_i - \mu \} - \lambda_0 \{ \sum P_i - 1 \}] = 0 \quad (7.4)$$

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:

$$(1) \quad \delta H[P] = H[P + \delta P] - H[P] = - \sum_i \delta P_i [\log P_i + 1] \quad (7.5)$$

$$(2) \quad \delta \{ -\lambda_1 (\sum_i x_i P_i - \mu) \} = -\lambda_1 \sum_i x_i \delta P_i \quad (7.6)$$

$$(3) \quad \delta \{ -\lambda_0 (\sum_i P_i - 1) \} = -\lambda_0 \sum_i \delta P_i \quad (7.7)$$

Juntando os três termos:

$$- \sum_i \delta P_i \{ \log P_i + 1 + \lambda_1 x_i + \lambda_0 \} = 0 \quad (7.8)$$

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:

$$\log P_i + 1 + \lambda_1 x_i + \lambda_0 = 0 \quad (7.9)$$

e segue que

$$P_i = e^{-1-\lambda_0-\lambda_1 x_i} \quad (7.10)$$

Os vínculos determinam os multiplicadores de Lagrange. Primeiro olhamos para o vínculo da normalização que será imposto escolhendo λ_0 . Definimos Z , que será uma grandeza central em todo desenvolvimento futuro, $Z = e^{1+\lambda_0}$, portanto é uma função de λ_0 . Mas ao impor o vínculo

$$P_i = \frac{1}{Z} e^{-\lambda_1 x_i} \quad (7.11)$$

$$1 = \sum_i P_i = \frac{1}{Z} \sum_i e^{-\lambda_1 x_i} \quad (7.12)$$

passa a ser uma função do multiplicador que resta, λ_1

$$Z(\lambda_1) = \sum_i e^{-\lambda_1 x_i} \quad (7.13)$$

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

$$\frac{d \log Z}{d \lambda_1} = -\frac{1}{Z} \sum_i x_i e^{-\lambda_1 x_i} = -\sum_i x_i P_i \quad (7.14)$$

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

$$\langle x \rangle = -\frac{d \log Z}{d \lambda_1} \quad (7.15)$$

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 $P(x)$.

7.1.1 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:

$$\delta [-\sum P_i \log P_i - \lambda_0 \{\sum P_i - 1\}] = 0 \quad (7.16)$$

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 $P_2 = 1 - p$

$$H(p) = -p \log p - (1 - p) \log(1 - p) \quad (7.17)$$

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.

7.1.2 3 estados

Suponha que seja dada a seguinte informação: “ X toma valores 0, 1 ou 2”, ou seja, em linguagem que vai ser usada mais adiante: “as configurações do sistema são os valores $x = 0, 1, 2$ ”, ou ainda: “o sistema tem um grau de liberdade X que toma valores $x = 0, 1, 2$ ”. Isto descreve o sistema. É dada informação adicional que descreve algo sobre o estado em que se encontra o sistema: “o valor médio de x , é conhecido: $\langle x \rangle = m$ ”. A distribuição de máxima entropia compatível com o vínculo é obtida a partir da equação 7.3. O resultado, já eliminado o multiplicador λ_0

$$P_i = \frac{e^{-\lambda_1 x_i}}{Z(\lambda_1)} \tag{7.18}$$

com λ_1 determinado pela eq. 7.15 e Z pela eq. 7.13 . Portanto, chamando $u = \exp(-\lambda_1)$, a função de partição é soma sobre as configurações

$$Z = \sum_i u^i = 1 + u + u^2 \tag{7.19}$$

e

$$P_i = \frac{u^i}{1 + u + u^2} \tag{7.20}$$

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

$$u^2(2 - m) + u(1 - m) - m = 0 \tag{7.21}$$

$$u = -\frac{1 - m \pm \sqrt{6m - 3m^2 + 1}}{2(2 - m)} \tag{7.22}$$

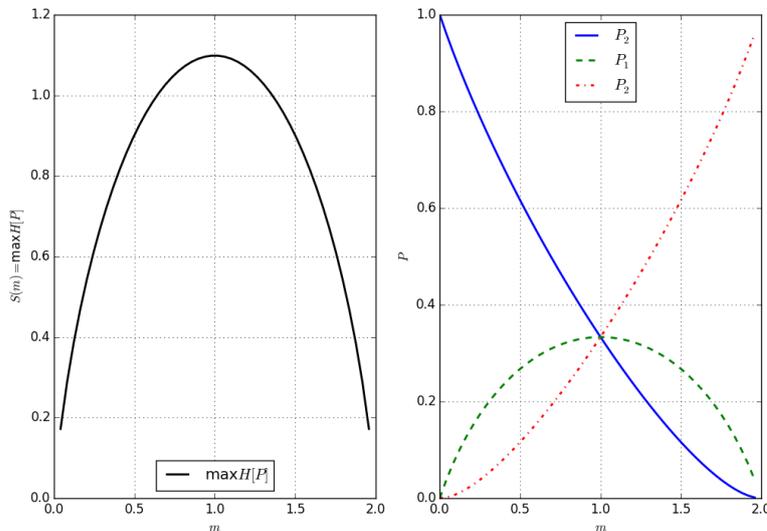


Figura 7.1: Esquerda: Máximo da entropia como função da média m . Direita: as probabilidades P_0, P_1 e P_2 que maximizam a entropia de Shannon como função da informação $\langle x \rangle = m$.

A equação 7.21 tem duas raízes. O sinal positivo leva a $u \geq 0$. A entropia de Shannon é definida no simplex $P_0 + P_1 + P_2 = 1$. O gráfico de $H[P]$ deveria ser desenhado num espaço tridimensional (bidimensional se o vínculo da normalização for incluído). Mas desenhamos $S(m) = \max H[P]$ que é função de um único parâmetro m . A figura 7.1 mostra o máximo da entropia H como função de m . Este valor é que vai ser identificado com a entropia termodinâmica e não a expressão geral. Do lado esquerdo aparecem as probabilidades. Sem nenhuma informação sobre m , o máximo de H é obtido para $P_0 = P_1 = P_2 = 1/3$, que é esperado por simetria ou razão insuficiente.

7.1.3 Variáveis reais: a gaussiana

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

$$\begin{aligned} \int P(x)dx &= 1 \\ \int xP(x)dx &= \mu \\ \int x^2P(x)dx &= \sigma^2 + \mu^2 \end{aligned} \quad (7.23)$$

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

$$\delta \left[- \int dx P(x) \log \left(\frac{P(x)}{p_0(x)} \right) + \lambda_0 \left\{ \int P(x)dx - 1 \right\} + \right. \quad (7.24)$$

$$\left. \lambda_1 \left\{ \int xP(x)dx - \mu \right\} + \lambda_2 \left\{ \int x^2P(x)dx - \sigma^2 - \mu^2 \right\} \right] = 0 \quad (7.25)$$

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 ¹ 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 ,

$$\{1 + \log P(x) + \lambda_0 + \lambda_1 x + \lambda_2 x^2\} \delta P(x) = 0, \quad (7.26)$$

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

$$P(x) = e^{1 - \lambda_0 - \lambda_1 x - \lambda_2 x^2} \quad (7.27)$$

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

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (7.28)$$

¹ Ou seja voltamos à entropia de Shannon. O que significa uma distribuição uniforme no eixo real inteiro? Veja Jeffreys para priors não normalizáveis.

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 adicional não conhecida e portanto não garantida.

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:

$$y_i = f_\theta(x_i) + \eta_i \quad (7.29)$$

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

$$P(\theta | \{x_i, y_i\}, I) = \frac{p_0(\theta) P(\{y_i\} | \theta, \{x_i\}, I)}{P(D)}, \quad (7.30)$$

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 7.29 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.

7.2 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

$$K[p|q] = \sum_i p_i \log \frac{p_i}{q_i}. \quad (7.31)$$

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

$$K[p|q] = \sum_i p_i \log(np_i), \quad (7.32)$$

$$K[p|q] = \log n - H[p], \quad (7.33)$$

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

$$\log x \leq x - 1, \quad (7.34)$$

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 7.34 usada na forma de $H[p]$ leva a

$$\begin{aligned} H[p] &= -\sum_i p_i \log(p_i), \\ &\geq -\sum_i p_i (p_i - 1) \\ &\geq 1 - \sum_i p_i^2 \geq 0, \end{aligned} \quad (7.35)$$

que significa 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

$$\begin{aligned} K[p|q] &= -\sum_i p_i \log \frac{1}{np_i}, \\ &\geq -\sum_i p_i \left(\frac{1}{np_i} - 1 \right) \\ &\geq \sum_i \frac{1}{n} - \sum_i p_i = 0, \end{aligned} \quad (7.36)$$

com a igualdade ocorrendo somente se $p_i = q_i$. Juntando as duas cotas com a relação entre K e H , equação 7.33, temos que

$$0 \leq H[p] \leq \log n, \quad (7.37)$$

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.

7.2.1 Independência

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

$$P(X, Y|I) = P(X|YI)P(Y|I). \quad (7.38)$$

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

$$\begin{aligned} K[P(X, Y|I)|P(X|I)P(Y|I)] &= \sum_{X,Y} P(X, Y|I) \log \frac{P(X, Y|I)}{P(X|I)P(Y|I)} \\ &= -H[P(X, Y|I)] + \sum_{X,Y} P(X, Y|I) \log P(X|I) + \sum_{X,Y} P(X, Y|I) \log P(Y|I) \\ &= -H[P(X, Y|I)] + H[P(X|I)] + H[P(Y|I)] \end{aligned} \quad (7.39)$$

e como $K \geq 0$ obtemos que

$$H[P(X, Y|I)] \leq H[P(X|I)] + H[P(Y|I)]. \quad (7.40)$$

A equação 7.39 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 7.40, 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.

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 :

$$P(x, y|\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)}. \quad (7.41)$$

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 $P(\sigma_x | \sigma_y)$.

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 ζ_x e ζ_y que fazem um ângulo θ . Trace dois vetores perpendiculares às 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 para 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 7.41.

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 ².

² Mas não neste curso introdutório.

7.3 Caso Geral para informação sobre valores esperados

Voltamos ao problema de como usar o formalismo de atribuição de probabilidades. Uma variável X toma valores x_i . Você pode pensar por agora que cada x_i é um número ou como será útil posteriormente, um conjunto de valores. Em geral podemos incluir vários vínculos do tipo

$$\langle g_k(x) \rangle = G_k, \quad (7.42)$$

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

$$\delta \left[- \sum_i P_i \log P_i - \sum_{k=0}^K \lambda_k \left\{ \sum_i g_k(x_i) P_i - \mu \right\} \right] = 0 \quad (7.43)$$

que leva a

$$- \sum_i \delta P_i \left\{ \log P_i + 1 + \sum_{k=0}^K \lambda_k g_k(x_i) \right\} = 0 \quad (7.44)$$

Alguns estudantes confundem, neste ponto, os diferentes tipos de somas, sobre configurações \sum_i e sobre vínculos \sum_k . Cuidado. Como anteriormente, as variações são independentes e cada termo entre chaves é nulo, o que leva a

$$P_i = e^{-1 - \sum_{k=0}^K \lambda_k g_k(x_i)} = e^{-1 - \lambda_0 - \sum_{k=1}^K \lambda_k g_k(x_i)} \quad (7.45)$$

$$P_i = \frac{e^{-\sum_{k=1}^K \lambda_k g_k(x_i)}}{Z} \quad (7.46)$$

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

$$Z(\lambda_1 \dots \lambda_K) = \sum_i e^{-\sum_{k=1}^K \lambda_k g_k(x_i)} \quad (7.47)$$

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

$$G_k = \langle g_k \rangle = - \left(\frac{\partial \log Z}{\partial \lambda_k} \right)_{\lambda_i, i \neq k} \quad (7.48)$$

onde a primeira igualdade é a informação dada, e a segunda é fácil de mostrar. **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.

7.4 Conexão com a Termodinâmica: estrutura matemática

7.4.1 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

$$S(G_1, G_2, \dots, G_K) = \log Z(\lambda_1, \lambda_2, \dots, \lambda_K) + \sum_{k=1}^K \lambda_k G_k \quad (7.49)$$

Este resultado é importante porque mostraremos que é relação análoga à relação termodinâmica $S = -\frac{F}{T} + \frac{E}{T}$, ou mais comumente escrita com $F = E - TS$, mas ainda não sabemos como. Na forma 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:

$$\frac{\partial S(G_1, G_2, \dots, G_K)}{\partial G_i} = \sum_{j=1}^K \frac{\partial \log Z}{\partial \lambda_j} \frac{\partial \lambda_j}{\partial G_i} + \sum_{j=1}^K \frac{\partial \lambda_j}{\partial G_i} G_j + \lambda_i \quad (7.50)$$

Usando a equação 7.48, vemos que as duas somas se cancelam e obtemos finalmente que

$$\frac{\partial S(G_1, G_2, \dots, G_K)}{\partial G_i} = \lambda_i \quad (7.51)$$

Um caso particular desta relação é $(\frac{\partial S}{\partial E})_{V,N} = \frac{1}{T}$ que veremos na seção 7.7.1.

Os resultados 7.49 e 7.51 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 $\frac{1}{T} = (\frac{\partial S}{\partial E})_{V,N}$.

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 e os símbolos T e F ainda não podem ser entendidos como representantes de grandezas físicas, mas mostra aonde queremos chegar.

7.4.2 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 7.51:

$$\frac{\partial \lambda_i}{\partial G_j} = \frac{\partial^2 S(G_1, G_2, \dots, G_K)}{\partial G_i \partial G_j} = \frac{\partial^2 S(G_1, G_2, \dots, G_K)}{\partial G_j \partial G_i} = \frac{\partial \lambda_j}{\partial G_i} \quad (7.52)$$

e da equação 7.48

$$\frac{\partial G_k}{\partial \lambda_i} = -\frac{\partial^2 \log Z}{\partial \lambda_k \partial \lambda_i} = \frac{\partial G_i}{\partial \lambda_k} \quad (7.53)$$

Aqui é interessante fazer uma pausa e relembrar 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 S é diferenciável duas vezes, 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 predições 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 Clássico ³ 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 nos apressar e jogar fora o método, mas desconfiar da maneira como foi aplicado. Há muitos exemplos, além da Mecânica Quântica, onde isto ocorre.

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 Na época não era necessário o adjetivo clássico.

7.5 *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. Esta segunda justificativa não faz sentido atualmente ⁴.

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 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

$$\{qp\}_{mol}\{qp\}_{atom}\{qp\}_{elet}\{qp\}_{nucl}\{qp\}_{prot}\{qp\}_{neutr}\{qp\}_{quark}\{qp\}_{gluon}\{qp\}? \quad (7.54)$$

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”⁵.

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,

⁴ Esta história talvez não seja verdadeira. Mas o rigor histórico sobre anedótas de como a ciência poderia ter chegado aonde chegou, não é o objetivo destas notas. Sobre a história temos informação incompleta e portanto só podemos atribuir probabilidades...

⁵ A.C.

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 do 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 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

$$P(\sigma_1 = +1, \dots, \sigma_N = -1, \theta_1, \dots, \theta_N) \prod_i d\theta_i = p \dots q P(\theta_1) \dots P(\theta_N) \prod_i d\theta_i \quad (7.55)$$

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

$$\begin{aligned} P(m|N\{p, q, P(\theta)\}) &= \sum_{\sigma_1, \dots, \sigma_N} \int \prod_i d\theta_i P(\sigma_1, \dots, \sigma_N, \theta_1, \dots, \theta_N) \delta_{m, (N + \sum_i \sigma_i) / 2} \\ &= \frac{N!}{m!(N-m)!} p^m q^{N-m} \end{aligned} \quad (7.56)$$

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.⁶ 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 idênticas.

⁶ As 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 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

$$P(\{\sigma_i\}_{i=1,N}|I) = \int \prod_i d\theta_i P(\{(\theta_i, \sigma_i)\}_{i=1,N}|I) \quad (7.57)$$

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 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.

7.6 Conexão com a Termodinâmica: Física

7.6.1 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 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

$$H[P] = -k \int_{E \leq \mathcal{H}(q_i, p_i) \leq E + \delta E, V} \prod_i dq_i dp_i P(q_i, p_i) \log P(q_i, p_i) \quad (7.58)$$

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:

$$P(q_i, p_i) = c \quad (7.59)$$

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 7.59 é determinada pela imposição do vínculo de normalização,

$$P(q_i, p_i) = \frac{1}{\Omega(E, N, V)} \quad (7.60)$$

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

$$\Omega(E, N, V) = \int_{E \leq \mathcal{H}(q_i, p_i) \leq E + \delta E} \prod_i dq_i dp_i. \quad (7.61)$$

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

$$S(E, N, V) = \max H[P] = \int_{E \leq \mathcal{H}(q_i, p_i) \leq E + \delta E} \prod_i dq_i dp_i \frac{1}{\Omega(E, N, V)} \log \Omega(E, N, V) \quad (7.62)$$

e o extremo da entropia é

$$S(E, N, V) = \log \Omega(E, N, V). \quad (7.63)$$

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 7.63.

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.

7.7 Conexão com a Termodinâmica: T, P, μ

7.7.1 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. Mas pode haver troca de energia entre os dois subsistemas. A

separação é feita mantendo

$$E_1 + E_2 = E \quad (7.64)$$

$$V_1 + V_2 = V, \quad (7.65)$$

$$N_1 + N_2 = N, \quad (7.66)$$

tal que todos os termos das equações 7.65 e 7.66 conhecidos enquanto só o lado direito da equação 7.64 é conhecida . Como é feita a distribuição de energia entre os dois subsistemas?

Suponhamos que o Hamiltoniano é bem descrito por

$$\mathcal{H}(\{q_i, p_i, Q_i, P_i\}) = \mathcal{H}_1(\{q_i, p_i\}) + \mathcal{H}_2(\{Q_i, P_i\}) \quad (7.67)$$

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

$$\begin{aligned} H[P(\{q_i, p_i, Q_i, P_i\})] &= - \int_{E_1 \leq \mathcal{H}_1(q_i, p_i) \leq E_1 + \delta E_1, E_2 \leq \mathcal{H}_2(Q_i, P_i) \leq E_2 + \delta E_2} P(\{q_i, p_i, Q_i, P_i\}) \log P(\{q_i, p_i, Q_i, P_i\}) \prod_i dq_i dp_i dQ_i dP_i \\ &= - \int_{E_1 \leq \mathcal{H}_1(q_i, p_i) \leq E_1 + \delta E_1} \prod_i dq_i dp_i P(\{q_i, p_i\}) \log P(\{q_i, p_i\}) \\ &\quad - \int_{E_2 \leq \mathcal{H}_2(Q_i, P_i) \leq E_2 + \delta E_2} \prod_i dQ_i dP_i P(\{Q_i, P_i\}) \log P(\{Q_i, P_i\}) \\ H[P(\{q_i, p_i, Q_i, P_i\})] &= H[P(\{q_i, p_i\})] + H[(P\{Q_i, P_i\})] \end{aligned} \quad (7.68)$$

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

$$\begin{aligned} \Omega(E, N, V) &= \int_{E_1 \leq \mathcal{H}_1(\{q_i, p_i\}) \leq E_1 + \delta E_1, E_2 \leq \mathcal{H}_2(\{Q_i, P_i\}) \leq E_2 + \delta E_2} \prod_i dq_i dp_i dQ_i dP_i \\ &= \int_{E_1 \leq \mathcal{H}_1(\{q_i, p_i\}) \leq E_1 + \delta E_1} \prod_i dq_i dp_i \int_{E_2 \leq \mathcal{H}_2(\{Q_i, P_i\}) \leq E_2 + \delta E_2} \prod_i dQ_i dP_i \\ &= \Omega(E_1, N_1, V_1) \Omega(E_2, N_2, V_2), \end{aligned} \quad (7.69)$$

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

$$S(E_1, E_2, N_1, N_2, V_1, V_2) = \log \Omega(E_1, N_1, V_1) \Omega(E_2, N_2, V_2), \quad (7.70)$$

$$S(E_1, E_2, N_1, N_2, V_1, V_2) = S(E_1, N_1, V_1) + S(E_2, N_2, V_2) \quad (7.71)$$

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 sujeita ao vínculo $E = E_1 + E_2$. Derivando a equação 7.71 e igualando a zero

$$\begin{aligned} 0 &= \frac{\partial S(E_1, E_2, N_1, N_2, V_1, V_2)}{\partial E_1} = \frac{\partial S(E_1, N_1, V_1)}{\partial E_1} + \frac{\partial S(E_2, N_2, V_2)}{\partial E_1} \\ &= \frac{\partial S(E_1, N_1, V_1)}{\partial E_1} - \frac{\partial S(E_2, N_2, V_2)}{\partial E_2} \end{aligned} \quad (7.73)$$

$$\left(\frac{\partial S(E_1, N_1, V_1)}{\partial E_1} \right)_{V_1, N_1} = \left(\frac{\partial S(E_2, N_2, V_2)}{\partial E_2} \right)_{V_2, N_2}. \quad (7.74)$$

A teoria prevê que os dois subsistemas que podem trocar energia a volume e número de partículas constantes terão a derivada da entropia com respeito à energia iguais. 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,

$$\left(\frac{\partial S(E, N, V)}{\partial E} \right)_{V, N} = \frac{1}{T} \quad (7.75)$$

e reescrevemos a equação 7.74 como $1/T_1 = 1/T_2$. Temos uma previsão a partir de vínculos e da hipótese que o estado macroscópico do sistema é determinado por E, V e N . Há outros estados do sistema, mas esses não são determinados por poucas variáveis. Estes não são chamados de estados de equilíbrio. Dizemos que os sistemas 1 e 2 estarão em equilíbrio quando suas temperaturas forem iguais. Ainda não, mas mais a frente isto nos permitirá interpretar (ver seções 7.4.1 e 7.9) do ponto de vista experimental o parâmetro de Lagrange associado a um vínculo, no caso a temperatura com um vínculo sobre o energia. Os parâmetros intensivos da termodinâmica serão associados a multiplicadores de Lagrange que impõem vínculos sobre valores de grandezas extensivas.

7.7.2 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

$$E_1 + E_2 = E \quad (7.76)$$

$$V_1 + V_2 = V, \quad (7.77)$$

$$N_1 + N_2 = N, \quad (7.78)$$

mas do lado esquerdo, só os termos da equação 7.78 são conhecidos. Novamente procuramos o máximo da equação

$$S(E_1, E_2, N_1, N_2, V_1, V_2) = S(E_1, N_1, V_1) + S(E_2, N_2, V_2) \quad (7.79)$$

mas agora devemos satisfazer duas condições:

$$\left(\frac{\partial S(E_1, E_2, N_1, N_2, V_1, V_2)}{\partial E_1} \right)_{V_1, N_1, N_2} = 0; \quad \left(\frac{\partial S(E_1, E_2, N_1, N_2, V_1, V_2)}{\partial V_1} \right)_{E_1, N_1, N_2} = 0 \quad (7.80)$$

que levam a

$$\left(\frac{\partial S(E_1, N_1, V_1)}{\partial E_1} \right)_{V_1, N_1} = \left(\frac{\partial S(E_2, N_2, V_2)}{\partial E_2} \right)_{V_2, N_2}. \quad (7.81)$$

e

$$\left(\frac{\partial S(E_1, N_1, V_1)}{\partial V_1} \right)_{V_1, N_1} = \left(\frac{\partial S(E_2, N_2, V_2)}{\partial V_2} \right)_{V_2, N_2}. \quad (7.82)$$

e usando a definição termodinâmica de pressão

$$\frac{P}{T} = \left(\frac{\partial S(E, N, V)}{\partial V} \right)_{E, N} \quad (7.83)$$

vemos que significam respectivamente

$$\frac{1}{T_1} = \frac{1}{T_2} \quad e \quad \frac{P_1}{T_1} = \frac{P_2}{T_2} \quad (7.84)$$

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.

7.7.3 Potencial químico

Consideremos a situação em que a parede que separa os dois subsistemas é rígida mas não impermeável. Só os termos da equação 7.77 são determinados pelo arranjo experimental. Podemos supor que o sistema é composto por vários tipos de moléculas

$$N_1^a + N_2^a = N^a \quad (7.85)$$

$$N_1^b + N_2^b = N^b \quad (7.86)$$

$$N_1^c + N_2^c = N^c \quad (7.87)$$

mas a parede só permite a passagem de moléculas do tipo a , sendo os termos da equações 7.86 e 7.87 conhecidos e fixos. Obteremos, por maximização da entropia do sistema conjunto que

$$\left(\frac{\partial S(E_1, N_1^a, N_1^b, N_1^c, V_1)}{\partial N_1^a} \right)_{E_1, N_1^b, N_1^c, V_1} = \left(\frac{\partial S(E_2, N_2^a, N_2^b, N_2^c, V_2)}{\partial N_2^a} \right)_{E_2, N_2^b, N_2^c, V_2}. \quad (7.88)$$

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

$$-\frac{\mu}{T} = \left(\frac{\partial S(E, N, V)}{\partial N} \right)_{E, V} \quad (7.89)$$

$$\frac{1}{T_1} = \frac{1}{T_2} \quad \text{e} \quad -\frac{\mu_1^a}{T_1} = -\frac{\mu_2^a}{T_2} \quad (7.90)$$

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. Este é um problema complicado.

7.8 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 resultado 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. Faremos apenas um breve comentário. A entropia cruzada ou relativa, deduzida no final do capítulo anterior, mostra que se usarmos algo que não é extensivo cometeremos uma violação dos desejos listado como desiderata entrópica. Qual é o desejo violado? O de independência. 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.

A equação 7.67 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

$$S(E, V, N) = \sum_{i=1, K} S(E_i, V_i, N_i) \quad (7.91)$$

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

$$S(E, V, N) = KS(E/K, V/K, N/K) \quad (7.92)$$

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

$$dS(E, V, N) = \left(\frac{\partial S}{\partial E} \right)_{VN} dE + \left(\frac{\partial S}{\partial V} \right)_{EN} dV + \left(\frac{\partial S}{\partial N} \right)_{EV} d\mu \quad (7.93)$$

e usando as definições dos parâmetros intensivos

$$dS(E, V, N) = \frac{1}{T} dE + \frac{P}{T} dV - \frac{\mu}{T} d\mu. \quad (7.94)$$

Um pouco mais de termodinâmica pode ser obtida da expressão de extensividade equação 7.92 e escrita como (com $\kappa = 1/K$)

$$\kappa S(E, V, N) = S(\kappa E, \kappa V, \kappa N). \quad (7.95)$$

Derivando com respeito a κ

$$\begin{aligned} \frac{\partial \kappa S(E, V, N)}{\partial \kappa} &= \frac{\partial S(\kappa E, \kappa V, \kappa N)}{\partial \kappa}, \\ S(E, V, N) &= \frac{\partial S(\kappa E, \kappa V, \kappa N)}{\partial(\kappa E)} \frac{\partial(\kappa E)}{\partial \kappa} + \\ &+ \frac{\partial S(\kappa E, \kappa V, \kappa N)}{\partial(\kappa V)} \frac{\partial(\kappa V)}{\partial \kappa} + \frac{\partial S(\kappa E, \kappa V, \kappa N)}{\partial(\kappa N)} \frac{\partial(\kappa N)}{\partial \kappa} \\ &= \frac{\partial S(\kappa E, \kappa V, \kappa N)}{\partial(\kappa E)} E + \frac{\partial S(\kappa E, \kappa V, \kappa N)}{\partial(\kappa V)} V + \frac{\partial S(\kappa E, \kappa V, \kappa N)}{\partial(\kappa N)} N \end{aligned}$$

e tomando $\kappa = 1$

$$S = \frac{1}{T}E + \frac{P}{T}V - \frac{\mu}{T}N.$$

$$E - TS + PV - \mu N = 0, \quad (7.96)$$

que é uma consequência de um teorema devido a Euler para funções homogêneas, definidas pela propriedade 7.92. Tomando o diferencial da expressão 7.96

$$dE - TdS - SdT + PdV + VdP - \mu dN - Nd\mu = 0 \quad (7.97)$$

e usando a equação 7.94 chegamos à relação de Gibbs-Duhem:

$$-SdT + VdP - Nd\mu = 0, \quad (7.98)$$

que faz a predição que as variações das quantidades intensivas não são independentes, mas dadas as variações de duas, a variação da terceira é fixa, em primeira ordem.

7.9 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 7.3. 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 7.51,

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

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

$$\langle \mathcal{H} \rangle = \int \prod_i dq_i dp_i P(q_i, p_i) \mathcal{H}(q_i, p_i) \quad (7.99)$$

é fixo, podemos supor que não o conhecemos, mas tem um valor definido. Se soubessemos esse valor poderíamos aplicar os resultados da seção 7.3 e obter a distribuição - de Boltzmann, de Boltzmann-Gibbs, canônica - em analogia à equação 7.46, que agora toma a forma

$$P(\{q_i, p_i\}) = \frac{e^{-\beta \mathcal{H}}}{Z} \quad (7.100)$$

onde o multiplicador de Lagrange que chamamos β em lugar de λ , pode ser obtido da equação 7.51

$$\beta = \frac{\partial S(E, V, N)}{\partial E} \quad (7.101)$$

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:

$$\beta = \frac{1}{T} \quad (7.102)$$

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 soubéssemos 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 dada pela equação 7.13 adquire uma interpretação importante e toma a forma

$$Z = \int \prod_i dq_i dp_i e^{-\beta \mathcal{H}} \quad (7.103)$$

e a equação 7.48 fica assim

$$E = \langle \mathcal{H} \rangle = -\frac{\partial \log Z}{\partial \beta} \quad (7.104)$$

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

$$S(E, V, N) = - \int \prod_i dq_i dp_i P(q_i, p_i) \log P(q_i, p_i) \quad (7.105)$$

$$= - \int \prod_i dq_i dp_i \frac{e^{-\beta \mathcal{H}}}{Z} \log \left(\frac{e^{-\beta \mathcal{H}}}{Z} \right) \quad (7.106)$$

$$= - \int \prod_i dq_i dp_i \frac{e^{-\beta \mathcal{H}}}{Z} (-\beta \mathcal{H} - \log Z) \quad (7.107)$$

$$= \beta \langle \mathcal{H} \rangle - \beta \log Z \quad (7.108)$$

$$= E - \beta \log Z. \quad (7.109)$$

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

$$F(T, V, N) = -\frac{1}{\beta} \log Z(\beta, V, N) \quad (7.110)$$

$$F = E - TS \quad (7.111)$$

7.9.1 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 sé-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.

7.9.2 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 7.42 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 7.89. Voltaremos ao ensemble grande canônico em problemas importantes como o gás ideal clássico e quântico que veremos nos próximos capítulos.

8

Aplicações Simples

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

8.1 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:

$$\mathcal{H} = -\mu_0 h \sum_{i=1}^N \sigma_i \quad (8.1)$$

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

8.1.1 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

$$N = N_+ + N_- \quad (8.2)$$

$$E = -\mu_0 h (N_+ - N_-) \quad (8.3)$$

de onde segue que

$$N_+ = \frac{1}{2}(N - \mathcal{E}) = \frac{N}{2}(1 - \epsilon) \quad (8.4)$$

$$N_- = \frac{1}{2}(N + \mathcal{E}) = \frac{N}{2}(1 + \epsilon) \quad (8.5)$$

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

$$\Omega(E, N) = \frac{N!}{N_+! N_-!} = \frac{N!}{\left(\frac{1}{2}(N - \mathcal{E})\right)! \left(\frac{1}{2}(N + \mathcal{E})\right)!} \quad (8.6)$$

Usando a fórmula de Stirling para a expansão do fatorial obtemos

$$\begin{aligned} S(E, N) &= N \log N - N_+ \log N_+ - N_- \log N_- \\ &= -N \left(\frac{1 - \epsilon}{2} \log(1 - \epsilon) + \frac{1 + \epsilon}{2} \log(1 + \epsilon) \right) \end{aligned} \quad (8.7)$$

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

$$\begin{aligned} \frac{1}{T} &= \frac{\partial S(E, N)}{\partial E} = \frac{\partial \epsilon}{\partial E} \frac{\partial S}{\partial \epsilon} \\ &= \frac{1}{2\mu_0 h} \log \left(\frac{1 - \epsilon}{1 + \epsilon} \right) \end{aligned} \quad (8.8)$$

que pode ser invertida, dando

$$E = \mu_0 h N \epsilon = -\mu_0 h N \tanh \mu_0 h T \quad (8.9)$$

e para a magnetização por grau de liberdade

$$m := \left\langle \frac{\sum_{i=1}^N \sigma_i}{N} \right\rangle = \mu_0 \tanh \frac{\mu_0 h}{T} \quad (8.10)$$

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

$$\chi_T = \left(\frac{\partial m}{\partial h} \right)_T = \frac{\mu_0^2}{T} \frac{1}{\cosh^2 \frac{\mu_0 h}{T}} \quad (8.11)$$

mostrada na figura 8.3. 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.

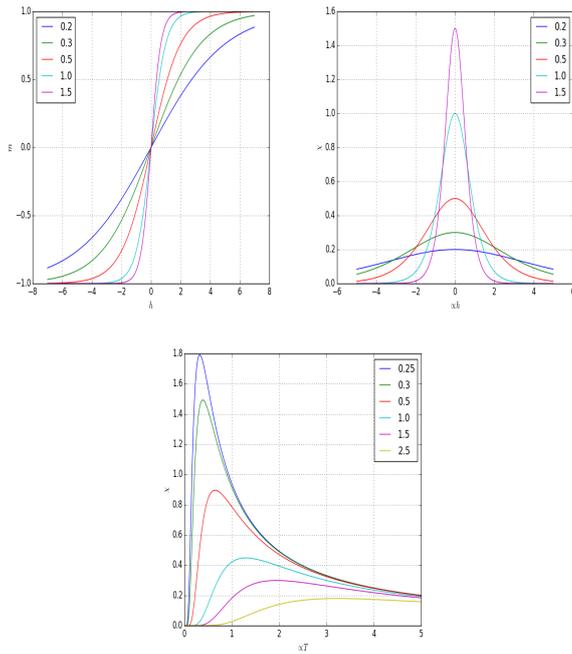


Figura 8.1: Esquerda: A magnetização por grau de liberdade m como função do campo h externo para diferentes valores da temperatura inversa β ($\mu_0 = 1$). Centro: Suscetibilidade magnética como função do campo h , para diferentes valores de β . Direita: $\chi(T)$ para diferentes valores de h .

8.1.2 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

$$P(\{\sigma_i\}) = \frac{e^{\beta\mu_0 h \sum_{i=1}^N \sigma_i}}{Z} \quad (8.12)$$

com a função de partição

$$Z(\beta, N, h) = \sum_{\{\sigma_i\}} e^{\beta\mu_0 h \sum_{i=1}^N \sigma_i} \quad (8.13)$$

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

$$\begin{aligned} Z(\beta, N, h) &= \prod_{i=1}^N \sum_{\sigma_i = \pm 1} e^{\beta\mu_0 h \sigma_i} \\ &= (2 \cosh \beta\mu_0 h)^N \end{aligned} \quad (8.14)$$

Note que da equação 8.13 obtemos

$$\frac{\partial \log Z}{\partial h} = \beta\mu_0 \left\langle \sum_{i=1}^N s_i \right\rangle \quad (8.15)$$

e da equação 8.14

$$\frac{\partial \log Z}{\partial h} = \beta \mu_0 N \tanh(\beta \mu_0 h) \quad (8.16)$$

e portanto chegamos a uma equação igual a equação 8.10.

Exercício Resfriamento magnético: Na figura 8.3 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 8.3. 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*.

8.2 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 :

$$E \leq \mathcal{H}(\{q_i, p_i\}) \leq E + \delta E, \quad (8.17)$$

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

$$\Omega_{dist}(E, V, N) = \int_{\mathcal{H}=E, V} \prod dq_i dp_i \quad (8.18)$$

mas

$$\Omega(E, V, N) = \int_{\mathcal{H}=E, V} \frac{\prod dq_i dp_i}{N! h^{3N}} \quad (8.19)$$

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 8.19 é muito simples. Para cada partícula temos uma integral sobre os valores possíveis das coordenadas, portanto cada partícula contribui com um fator V .

$$\Omega(E, V, N) = \frac{V^N}{N!} \int_{\mathcal{H}=E} \prod dp_i, \quad (8.20)$$

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 $1/\text{radial}=\text{angular}$. 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

$$\Omega(E, V, N) = \frac{V^N}{N!} \frac{2\pi^{\frac{3N}{2}}}{\Gamma(3N/2)} E^{\frac{3N-1}{2}} \delta E \quad (8.21)$$

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 7.63 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

$$S(E, V, N) = N \left(\log \frac{V}{N} + \frac{3}{2} \log \frac{E}{N} - \frac{1}{2} + \frac{3}{2} \log \frac{3\pi}{2} \right) + \dots \quad (8.22)$$

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 $v = V/N$.

$$s(u, v) = \frac{S}{N} = \log v + \frac{3}{2} \log u - \frac{1}{2} + \frac{3}{2} \log \frac{3\pi}{2}. \quad (8.23)$$

A seguir investigamos as equações de estado obtidas ao tomar as derivadas da entropia com respeito à energia, volume e número de partículas

$$\left(\frac{\partial S}{\partial E} \right)_{V, N} = \frac{1}{T} = \frac{3N}{2E} \quad (8.24)$$

$$\left(\frac{\partial S}{\partial V} \right)_{E, N} = \frac{P}{T} = \frac{N}{V} \quad (8.25)$$

$$\left(\frac{\partial S}{\partial N} \right)_{E, V} = \frac{-\mu}{T} = \log \frac{V}{N} + \frac{3}{2} \log \frac{E}{N} - 3 + \frac{3}{2} \log \frac{3\pi}{2} \quad (8.26)$$

¹ Uma definição pragmática de paradoxo: algo que está errado mas é fruto de um raciocínio a primeira vista correto.

(VERIFICAR CONTAS) onde reconhecemos as célebres equações de estado:

$$PV = NT \quad (8.27)$$

$$E = \frac{3}{2}NT \quad (8.28)$$

8.3 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

$$\mathcal{H}(\{q_i, p_i\}) = \sum_{i=1}^N \frac{p_i^2}{2m} \text{ fosse}$$

$$E = \langle \mathcal{H}(\{q_i, p_i\}) \rangle. \quad (8.29)$$

A densidade de probabilidades é

$$P(\{q_i, p_i\}) = \frac{\exp(-\beta \mathcal{H}(\{q_i, p_i\}))}{Z} \quad (8.30)$$

onde

$$Z = \int \frac{\prod dq_i dp_i}{N!} \exp(-\beta \mathcal{H}), \quad (8.31)$$

que impõe a normalização

$$1 = \int \frac{\prod dq_i dp_i}{N!} P(\{q_i, p_i\}). \quad (8.32)$$

A equação relevante é

$$E = -\frac{\partial \log Z}{\partial \beta} \quad (8.33)$$

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 :

$$\zeta_i = \int dq_i dp_i \exp(-\beta \mathcal{H}_i), \quad (8.34)$$

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

$$Z = \frac{1}{N!} \zeta^N. \quad (8.35)$$

A integração sobre as coordenadas é análoga à secção anterior, dando V^N .

Para integrar os momentos usamos:

$$\int_{-\infty}^{\infty} e^{-\frac{\beta}{2m} p^2} dp = \sqrt{2\pi \frac{m}{\beta}} \quad (8.36)$$

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

$$Z = \frac{V^N}{N!} \left(2\pi \frac{m}{\beta}\right)^{\frac{3N}{2}} \quad (8.37)$$

$$\zeta = V \left(2\pi \frac{m}{\beta}\right)^{\frac{3}{2}} \quad (8.38)$$

e a equação 8.33 dá:

$$E = \frac{3N}{2\beta} = 3N \frac{T}{2}. \quad (8.39)$$

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 7.109 através de $-\beta F = \log Z$, portanto $S = \log Z + \beta E$. A pressão no ensemble canônico, obtida através a equação 8.25 $\left(\frac{\partial S}{\partial V}\right)_{V,N} = \frac{P}{T}$, 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 $\log N!$

$$-\beta F = N \log V + \frac{3N}{2} \log \frac{2\pi m}{\beta} - N \log N + N \quad (8.40)$$

e definindo as densidades $f = F/N$, $v = V/N$ e $\epsilon = E/N$

$$-\beta f = \log v + \frac{3}{2} \log \frac{2\pi m}{\beta} + 1 \quad (8.41)$$

de onde fica claro que se o fatorial não estivesse presente as densidades não seriam intensivas.

8.4 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 $i = 1, \dots, \sum_l N_l$, para indexar as moléculas. O hamiltoniano pode ser aproximado por

$$\mathcal{H} = \sum_i \mathcal{H}_l(i), \quad (8.42)$$

e podemos supor que

$$\mathcal{H}_l(i) = \frac{p_i^2}{2m_l} + \epsilon_l(s_i) \quad (8.43)$$

² Não 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.

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

$$\zeta_l = \int dq dp \sum_s e^{-\beta \mathcal{H}_l}, \quad (8.44)$$

e a função de partição do sistema

$$Z(\beta, V, N_1 \dots N_m) = \prod_{l=1 \dots m} \frac{\zeta_l^{N_l}}{N_l!}, \quad (8.45)$$

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

$$-\beta F = \log Z = \sum_l (N_l \log \zeta_l - \log N_l + 1). \quad (8.46)$$

A pressão pode ser facilmente calculada pois cada fator ζ_l contribui com um fator V , portanto

$$-\beta F = \sum_l (N_l \log V) + \dots, \quad (8.47)$$

onde não escrevemos termos independentes do volume, segue que

$$\beta p = \left(\frac{\partial \log Z}{\partial V} \right)_{T, N_1 \dots N_l \dots N_m} \quad (8.48)$$

$$\beta p = \sum N_l \frac{1}{V} = \beta \sum_{l=1}^m p_l, \quad (8.49)$$

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} .

8.4.1 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

$$\frac{dN_1}{c_1} = \frac{dN_2}{c_2} = \dots = \frac{dN_l}{c_l} = \dots = \frac{dN_m}{c_m} \quad (8.50)$$

A energia livre $F(T, V, \{N_l\}_{l=1 \dots m})$ é mínima no equilíbrio, portanto flutuações dN_l devem satisfazer

$$0 = \Delta F = \sum_l \left(\frac{\partial F}{\partial N_l} \right)_{T, V, N_1, \dots, N_{l'} \neq l, \dots, N_m} dN_l. \quad (8.51)$$

Lembrando que estas derivadas de F são os potenciais químicos

$$\mu_l = \left(\frac{\partial F}{\partial N_l} \right)_{T, V, N_{l'} \neq l} \quad (8.52)$$

podemos ver que

$$\sum_l c_l \mu_l = 0. \quad (8.53)$$

E dada a relação entre a energia livre e a função de partição ($-\beta F = \log Z$), obtemos

$$\mu_l = -T \log \frac{\zeta_l}{N_l} = f_l - T \log N_l \quad (8.54)$$

onde introduzimos $f_l = -\beta^{-1} \log \zeta_l$, a soma $F_0 = \sum_l f_l$ e temos

$$0 = T \sum_l c_l \log \frac{\zeta_l}{N_l} = - \sum_l f_l - T \sum_l \log N_l^{c_l} \quad (8.55)$$

$$\prod_l N_l^{c_l} = e^{-\beta F_0} = K(T, V, \{C\}), \quad (8.56)$$

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

$$\frac{N_A^2 N_B}{N_C^2} = K(T, V, A, B, C) \quad (8.58)$$

8.5 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

$$\mathcal{H}(\{q_i, p_i\}) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} + V(\{q_i\}) \quad (8.59)$$

pode ser aproximado por

$$\mathcal{H}(\{q_i, p_i\}) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} + \sum_i V(q_i) \quad (8.60)$$

onde o potencial $V(q_i)$ só depende das coordenadas do átomo i e ainda mais, que é harmônico

$$\mathcal{H}(\{q_i, p_i\}) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} + \sum_i \frac{m\omega^2}{2} \mathbf{q}_i^2 \quad (8.61)$$

A função de partição é novamente gaussiana e além da equação 8.36 usamos

$$\int_{-\infty}^{\infty} e^{-\frac{\beta m \omega^2}{2} q^2} dq = \sqrt{2\pi \frac{1}{\beta m \omega^2}}. \quad (8.62)$$

A função de partição

$$Z = \left(\sqrt{2\pi \frac{m}{\beta}} \sqrt{2\pi \frac{1}{\beta m \omega^2}} \right)^{3N} \quad (8.63)$$

A energia, obtida derivando o logaritmo de Z dá:

$$E = \frac{6TN}{2} = 3NT (= 3nRT_K), \quad (8.64)$$

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.

8.6 Sólido de Einstein

Einstein considera os $3N$ osciladores descritos acima quantizados e independentes, o que significa que o microestado do sistema é descrito por $\{n_1, n_2, \dots, n_{3N}\}$ e a energia

³ N osciladores tridimensionais

$$\mathcal{H}(\{n_1, n_2, \dots, n_{3N}\}) = \hbar\omega \sum_{i=1}^{3N} \left(\frac{1}{2} + n_i \right) \quad (8.65)$$

8.6.1 Microcanônico

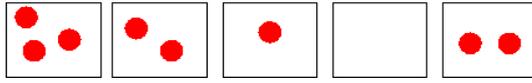
Temos a situação experimental em que o sistema está isolado e a energia é fixa, assim como o número de osciladores e o volume. As

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

$$\mathcal{H}(\{n_1, n_2, \dots, n_{3N}\}) = E = \frac{3N}{2} \hbar\omega + K\hbar\omega \quad (8.66)$$

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 objetos diferentes mas indistinguíveis entre si dentro da classe: $3N$ osciladores e K quanta de energia $\hbar\omega$:

4



⁴ Na realidade devemos considerar $3N - 1$ osciladores mas não faz diferença

Figura 8.2: Exemplo: uma das 1287 configurações distintas com $K = 8$ quanta e $N = 5$ osciladores

$$\Omega(E, N) = \frac{(3N + K)!}{(3N)!K!} \quad (8.67)$$

Definimos $\varepsilon = E/(N\hbar\omega)$, a energia por partícula medida em unidades de $\hbar\omega$. Para a entropia temos

$$S(E, N) = (3N + K) \log(3N + K) - 3N \log 3N - K \log K \quad (8.68)$$

$$= N \left(\left(\varepsilon + \frac{3}{2} \right) \log \left(\varepsilon + \frac{3}{2} \right) - \left(\varepsilon - \frac{3}{2} \right) \log \left(\varepsilon - \frac{3}{2} \right) \right) \quad (8.69)$$

onde usamos $\varepsilon = 3/2 + K/N$. Assim podemos calcular a temperatura

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E} \right)_N = \frac{1}{\hbar\omega} \log \frac{\varepsilon + \frac{3}{2}}{\varepsilon - \frac{3}{2}} \quad (8.70)$$

e invertendo temos

$$\varepsilon = \frac{3}{2} + \frac{3}{e^{\frac{\hbar\omega}{T}} - 1} \quad (8.71)$$

$$E = \frac{3N\hbar\omega}{2} + \frac{3N\hbar\omega}{e^{\frac{\hbar\omega}{T}} - 1} \quad (8.72)$$

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

$$E = \frac{3N\hbar\omega}{2} + 3N\hbar\omega \langle n \rangle \quad (8.73)$$

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

$$c_v = \frac{1}{N} \frac{\partial E}{\partial T} = 3 \frac{(\hbar\omega)^2 e^{\frac{\hbar\omega}{T}}}{T^2 (e^{\frac{\hbar\omega}{T}} - 1)^2}, \quad (8.74)$$

e tomando os limite de pequenas temperaturas vemos que o calor específico não é constante mas cai com a temperatura

$$c_v \approx 3 \frac{(\hbar\omega)^2 e^{-\frac{\hbar\omega}{T}}}{T^2}, \quad (8.75)$$

enquanto que para valores altos da temperatura

$$c_v \approx 3 \quad (8.76)$$

devemos colocar as unidades costumeiras para obter $c_v \approx 3R$ por mol.

8.6.2 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

$$P(\{n_i\}) = \frac{e^{-\beta \mathcal{H}(\{n_i\})}}{Z} \quad (8.77)$$

onde

$$Z(\beta, N) = \sum_{n_1 \dots n_{3N}} e^{-\beta \mathcal{H}} \quad (8.78)$$

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

$$Z(\beta, N) = \left[\sum_{n=0}^{\infty} e^{-\beta(n+\frac{1}{2})\hbar\omega} \right]^{3N}. \quad (8.79)$$

$$Z(\beta, N) = e^{-\frac{3N\beta}{2}\hbar\omega} \left[1 + a + a^2 + a^3 + \dots \right]^{3N}, \quad (8.80)$$

onde $a = e^{-\beta\hbar\omega}$ e somando a série geométrica

$$Z(\beta, N) = e^{-\frac{3N\beta}{2}\hbar\omega} \left[\frac{1}{1 - e^{-\beta\hbar\omega}} \right]^{3N}. \quad (8.81)$$

Usando $E = -\partial \log Z / \partial \beta$, obtemos

$$E = \frac{3N\hbar\omega}{2} + \frac{3N\hbar\omega}{e^{\beta\hbar\omega} - 1} \quad (8.82)$$

que é exatamente o resultado obtido no ensemble microcanônico

8.6.3 Gases ideais clássicos com estrutura interna quântica

a ser escrito ...

8.7 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

$$E = \langle E_\alpha \rangle, \quad N = \langle N_\alpha \rangle \quad V = \text{fixo}. \quad (8.83)$$

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:

$$E = \sum_{\alpha} P_{\alpha} E_{\alpha} \quad N = \sum_{\alpha} P_{\alpha} N_{\alpha}, \quad 1 = \sum_{\alpha} P_{\alpha}, \quad (8.84)$$

que inclui a normalização.

Maximizamos a entropia de Shannon sujeita a estes vínculos

$$H[P_{\alpha}] = - \sum_{\alpha} P_{\alpha} \log P_{\alpha} + \lambda_0 (1 - \sum_{\alpha} P_{\alpha}) + \beta (E - \sum_{\alpha} P_{\alpha} E_{\alpha}) + \lambda_1 (N - \sum_{\alpha} P_{\alpha} N_{\alpha}), \quad (8.85)$$

e obtemos

$$P_{\alpha} = e^{1-\lambda_0} e^{-\beta E_{\alpha} - \lambda_1 N_{\alpha}}. \quad (8.86)$$

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

$$P_{\alpha} = \frac{1}{\Xi(\beta, \lambda_1, V)} e^{-\beta E_{\alpha} - \lambda_1 N_{\alpha}} \quad (8.87)$$

onde

$$\Xi(\beta, \lambda_1, V) = \sum_{\alpha} e^{-\beta E_{\alpha} - \lambda_1 N_{\alpha}}. \quad (8.88)$$

Usamos, como é costume, a letra ζ 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 8.87, na entropia de Shannon 8.85

$$S(E, N, V) = - \sum_{\alpha} P_{\alpha} \log P_{\alpha}, \quad (8.89)$$

para obter

$$S(E, N, V) = \log \Xi + \beta E + \lambda_1 N. \quad (8.90)$$

O leitor deveria ver isto como semelhante à equação 7.109. 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 7.49. Primeiro notemos que

$$\frac{\partial \log \Xi}{\partial \beta} \Big|_{\lambda_1, V} = -E \quad (8.91)$$

$$\frac{\partial \log \Xi}{\partial \lambda_1} \Big|_{\beta, V} = -N \quad (8.92)$$

$$\frac{\partial \log \Xi}{\partial V} \Big|_{\beta, \lambda_1} = -\frac{P}{T} \quad (8.93)$$

Agora, usando a regra da cadeia

$$\frac{\partial S}{\partial E} \Big|_N = \frac{\partial \log \Xi}{\partial \beta} \frac{\partial \beta}{\partial E} + \frac{\partial \log \Xi}{\partial \lambda_1} \frac{\partial \lambda_1}{\partial E} + \frac{\partial \beta}{\partial E} E + \frac{\partial \lambda_1}{\partial E} N + \beta \quad (8.94)$$

$$\frac{\partial S}{\partial E} \Big|_N = \beta, \quad (8.95)$$

obtemos novamente a interpretação que β é o inverso da temperatura. Derivando com respeito a N :

$$\frac{\partial S}{\partial N} \Big|_E = \frac{\partial \log \Xi}{\partial \beta} \frac{\partial \beta}{\partial N} + \frac{\partial \log \Xi}{\partial \lambda_1} \frac{\partial \lambda_1}{\partial N} + \frac{\partial \beta}{\partial N} E + \frac{\partial \lambda_1}{\partial N} N + \lambda_1 \quad (8.96)$$

$$\frac{\partial S}{\partial N} \Big|_E = \lambda_1, \quad (8.97)$$

de onde obtemos, usando a definição do potencial químico (ver equação 7.89)

$$\lambda_1 = -\frac{\mu}{T} = -\beta\mu. \quad (8.98)$$

Segue que

$$\Xi(\beta, \mu) = \sum_{\alpha} e^{-\beta(E_{\alpha} - \mu N_{\alpha})}. \quad (8.99)$$

Para referência futura note que

$$E = \langle E_{\alpha} \rangle = -\frac{\partial \log \Xi}{\partial \beta} \Big|_{\mu, V} \quad (8.100)$$

$$N = \langle N_{\alpha} \rangle = \beta^{-1} \frac{\partial \log \Xi}{\partial \mu} \Big|_{\beta, V} \quad (8.101)$$

$$P = -\beta^{-1} \frac{\partial \log \Xi}{\partial V} \Big|_{\beta, \mu} \quad (8.102)$$

8.8 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:

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r}) + V(\mathbf{r})\psi(\mathbf{r}) = \epsilon\psi(\mathbf{r}) \quad (8.103)$$

obtendo os autoestados de energia ϵ_q , associados ao momento

$$\mathbf{p} = \hbar\mathbf{q}$$

$$\epsilon_q = \frac{\hbar^2\mathbf{q}^2}{2m}. \quad (8.104)$$

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

$$\mathbf{q} = (q_x, q_y, q_z), \text{ com } q_x = \frac{2\pi}{L}l_x, q_y = \frac{2\pi}{L}l_y, q_z = \frac{2\pi}{L}l_z \quad (8.105)$$

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:

$$\left(-\frac{\hbar^2}{2m}\sum_i^{N_\alpha}\nabla_i^2 + V(\mathbf{r})\right)\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{N_\alpha}) = E_\alpha\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{N_\alpha}), \quad (8.106)$$

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 8.103 é solução da equação 8.106.

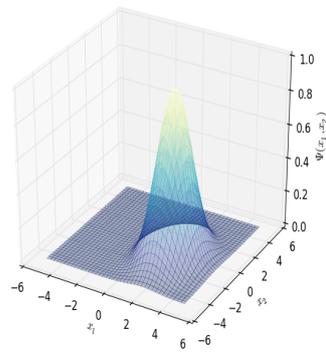
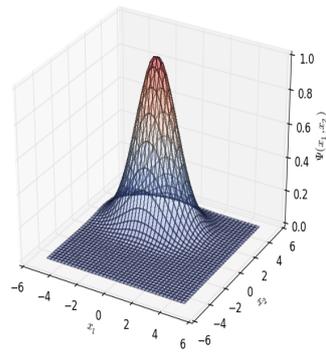
Mas isto não resolve o problema. Suponha duas partículas e duas soluções de 8.103, $\psi_a(\mathbf{r}_1)$ e $\psi_b(\mathbf{r}_2)$. A função $\Psi_{a1,b2} = \psi_a(\mathbf{r}_1)\psi_b(\mathbf{r}_2)$ é solução de 8.106 mas não leva em conta a indistinguibilidade das partículas. Note que colocar a partícula 1 no estado a e a 2 no b é diferente de colocar a 1 em b e a 2 em a , i.e.

$$\Psi_{a1,b2} \neq \Psi_{a2,b1} \quad (8.107)$$

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

$$\Psi_\pm(\mathbf{r}_1, \mathbf{r}_2) \propto \psi_a(\mathbf{r}_1)\psi_b(\mathbf{r}_2) \pm \psi_a(\mathbf{r}_2)\psi_b(\mathbf{r}_1) \quad (8.108)$$

Figura 8.3: (linhas grossas) $\Psi_{a_1, b_2} = \psi_a(x_1)\psi_b(x_2) \neq \psi_a(x_2)\psi_b(x_1) = \Psi_{a_2, b_1}$ (linhas finas)



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.

$$|\Psi_{\pm}(\mathbf{r}_1, \mathbf{r}_2)|^2 = |\Psi_{\pm}(\mathbf{r}_2, \mathbf{r}_1)|^2 \quad (8.109)$$

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 8.109 deva ser satisfeita. Isto em princípio poderia levar ao aparecimento de uma fase

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_k, \mathbf{r}_l, \dots) = e^{i\phi} \Psi(\mathbf{r}_1, \dots, \mathbf{r}_l, \mathbf{r}_k, \dots) \quad (8.110)$$

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_{α} será caracterizado por o conjunto de números de ocupação. Como cada estado-1 é caracterizado por um conjunto de números quânticos q , então o estado- N_{α} é representado por $(n_{q_1}, n_{q_2}, \dots, n_{q_l}, \dots)$, com

$$N_{\alpha} = \sum_q n_q. \quad (8.111)$$

No caso do gás ideal o q são simplesmente os momentos permitidos na caixa pelas condições de contorno e portanto

$$E_{\alpha} = \sum_q n_q \epsilon_q, \quad (8.112)$$

pois a energia total do estado- N_{α} é a soma das energias dos estados-1 ocupados e simplesmente juntamos os termos iguais.

Exercício Mostre a equação 8.112 a partir da equação 8.106.

O que significa que uma função de onda seja simétrica (+) ou antisimétrica (-) ao intercambiar o índice de duas partículas? Consideremos

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_k, \mathbf{r}_l, \dots) = \pm \Psi(\mathbf{r}_1, \dots, \mathbf{r}_l, \mathbf{r}_k, \dots) \quad (8.113)$$

para o caso de duas partículas. Segue que

$$\Psi_{BE}(\mathbf{r}_1\mathbf{r}_2) = \frac{1}{\sqrt{2}} (\psi_a(\mathbf{r}_1)\psi_b(\mathbf{r}_2) + \psi_a(\mathbf{r}_2)\psi_b(\mathbf{r}_1)) \quad (8.114)$$

$$\Psi_{FD}(\mathbf{r}_1\mathbf{r}_2) = \frac{1}{\sqrt{2}} (\psi_a(\mathbf{r}_1)\psi_b(\mathbf{r}_2) - \psi_a(\mathbf{r}_2)\psi_b(\mathbf{r}_1)) \quad (8.115)$$

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_α , que são descritos pelo conjunto de números de ocupação devem ser

$$\text{Bósons } \alpha \leftrightarrow (n_1, n_2, \dots) \text{ sem nenhuma restrição: } n_q = 0 \dots \infty, \quad (8.116)$$

$$\text{Férmions } \alpha \leftrightarrow (n_1, n_2, \dots) \text{ restrição: } n_q = 0, 1. \quad (8.117)$$

A generalização para mais de dois férmions foi feita por Slater que percebeu que a expressão 8.115 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 $\Xi(\beta, \mu, V) = \sum_\alpha e^{-\beta(E_\alpha - \mu N_\alpha)}$,

$$\begin{aligned} \Phi(\beta, \mu, V) &= -T \log \Xi \\ &= -T \log \sum_{n_{q_1}} e^{-\beta \sum_{q_1} n_{q_1} (\epsilon_{q_1} - \mu)} \\ &= -T \log \prod_{q_1} \sum_{n_{q_1}} e^{-\beta n_{q_1} (\epsilon_{q_1} - \mu)} \\ &= -T \sum_{q_1} \log \sum_{n_{q_1}} e^{-\beta n_{q_1} (\epsilon_{q_1} - \mu)} \quad (8.118) \end{aligned}$$

$$= \sum_{q_1} \Phi_{q_1}. \quad (8.119)$$

Ainda não sabemos como fazer a soma sobre os q e trataremos disso mais tarde. Por agora nos concentraremos nas somas sobre n_{q_1} , levando em conta 8.116 e 8.117. Para o caso de Bósons, a estatística Bose-Einstein será governada pelo potencial

$$\begin{aligned} \Phi_{q_1}^{BE} &= -T \log \sum_{n_{q_1}=0}^{\infty} e^{-\beta n_{q_1} (\epsilon_{q_1} - \mu)} \\ &= -T \log(1 + x + x^2 + x^3 + \dots) \\ &= -T \log\left(\frac{1}{1-x}\right) \\ &= T \log(1 - e^{-\beta(\epsilon_{q_1} - \mu)}) \quad (8.120) \end{aligned}$$

A convergência da série geométrica só ocorrerá caso $x = e^{-\beta(\epsilon_{q_1} - \mu)}$ seja menor que 1. Como o menor valor de ϵ_{q_1} , no limite de volume

infinito, é zero, devemos ter um potencial químico negativo

$$x < 1 \rightarrow \mu_{BE} < 0 \quad (8.121)$$

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:

$$\begin{aligned} \Phi_{q_l}^{FD} &= -T \log \sum_{n_{q_l}=0}^1 e^{-\beta n_{q_l}(\epsilon_{q_l} - \mu)} \\ &= -T \log(1 + x) \\ &= -T \log(1 + e^{-\beta(\epsilon_{q_l} - \mu)}), \end{aligned} \quad (8.122)$$

Juntando os resultados

$$\Phi_{q_l} = \mp T \log(1 \pm e^{-\beta(\epsilon_{q_l} - \mu)}) \quad (8.123)$$

com $-+$ para férmions e $+ -$ para bósons. Se a exponencial na expressão acima for pequena $e^{-\beta(\epsilon_{q_l} - \mu)} \ll 1$ a expansão do logaritmo ($\log(1 + x) \approx x$) leva ao mesmo resultado tanto para férmions quanto bósons:

$$\Phi_{q_l}^C = -T e^{-\beta(\epsilon_{q_l} - \mu)} \quad (8.124)$$

que veremos, recupera os resultados para o limite clássico.

Uma quantidade central na discussão que segue é o *número médio de ocupação* \bar{n}_q que pode ser calculado a partir de (ver equação 8.101)

$$\bar{n}_q = \langle n_q \rangle = \frac{\partial \Phi_q}{\partial \mu} \Big|_{\beta, V}. \quad (8.125)$$

Obtemos para bósons

$$\bar{n}_q^{BE} = \frac{1}{e^{\beta(\epsilon_{q_l} - \mu)} - 1} \quad (8.126)$$

e para férmions

$$\bar{n}_q^{FD} = \frac{1}{e^{\beta(\epsilon_{q_l} - \mu)} + 1} \quad (8.127)$$

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

$$N_{BE} = \sum_q \bar{n}_q = \sum_q \frac{1}{e^{\beta(\epsilon_{q_l} - \mu)} - 1} \quad (8.128)$$

$$N_{FD} = \sum_q \bar{n}_q = \sum_q \frac{1}{e^{\beta(\epsilon_{q_l} - \mu)} + 1} \quad (8.129)$$

Na figura 8.4 mostramos a função de Fermi. Não é o número de ocupação porque ainda não sabemos como depende o potencial

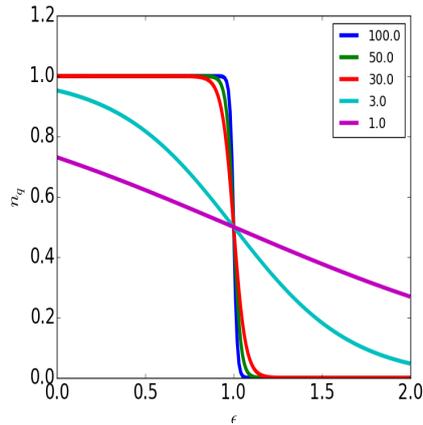


Figura 8.4: A função de Fermi: $n = (\exp(\beta(\epsilon - 1)) + 1)^{-1}$, para diferentes valores de β .

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.

8.9 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

$$\mathcal{N}_1 = \Delta l_x = \frac{L}{2\pi} \Delta q_x = \frac{L}{2\pi\hbar} \Delta p_x \quad (8.130)$$

Ao fazer a conta integrando sobre momentos num intervalo pequeno Δp_x

$$\int_{L, \Delta p_x} \frac{dx dp_x}{h} = \frac{L \Delta p_x}{h}. \quad (8.131)$$

Na equação 8.19 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

$$\mathcal{N}_3 = \frac{L^3 \Delta p_x \Delta p_y \Delta p_z}{(2\pi\hbar)^3}, \quad (8.132)$$

e em coordenadas esféricas, tomando os Δp como diferenciais, teremos a medida de integração

$$\frac{4\pi V p^2 dp}{(2\pi\hbar)^3}, \quad (8.133)$$

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 8.104

$$\epsilon_q = \frac{\hbar^2 q^2}{2m}, \quad (8.134)$$

leva, usando $d\epsilon = \frac{d\epsilon}{dp} dp = \frac{p dp}{m}$, a

$$\mathcal{N}_3 = \frac{4\pi V}{(2\pi\hbar)^3} \sqrt{2m^{\frac{3}{2}} \epsilon^{\frac{1}{2}}} d\epsilon, \quad (8.135)$$

e introduzindo $\rho(\epsilon)$, a densidade de estados de energia de uma partícula não relativística em 3 dimensões

$$\rho(\epsilon) d\epsilon = \frac{4\pi V}{(2\pi\hbar)^3} \sqrt{2m^{\frac{3}{2}} \epsilon^{\frac{1}{2}}} d\epsilon = CV \epsilon^{\frac{1}{2}} d\epsilon, \quad (8.136)$$

onde

$$C = \frac{4\pi}{(2\pi\hbar)^3} \sqrt{2m^{\frac{3}{2}}} \quad (8.137)$$

Exploraremos a seguir a aproximação de somas por integrais:

$$\sum_q (\cdot) \rightarrow \int (\cdot) \rho(\epsilon) d\epsilon. \quad (8.138)$$

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 $p = \hbar q = 2\pi\hbar/\lambda$

$$\frac{p^2}{2m} = k_B T \rightarrow \epsilon = \frac{1}{2m} \left(\frac{2\pi\hbar}{\lambda_T} \right)^2 \quad (8.139)$$

$$\lambda_T = \frac{2\pi\hbar}{\sqrt{2mk_B T}}. \quad (8.140)$$

O intervalo típico entre níveis de energia é

$$\delta\epsilon = \frac{1}{2m} \left(\frac{2\pi\hbar}{L} \right)^2 \quad (8.141)$$

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.

8.10 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 8.119 e 8.122. 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 $\rho(\epsilon)$

$$\begin{aligned} N &= \sum_q \bar{n}_q = \sum_q \frac{1}{e^{\beta(\epsilon_q - \mu)} + 1} \\ &\rightarrow \int_0^\infty \rho(\epsilon) \frac{1}{e^{\beta(\epsilon - \mu)} + 1} d\epsilon, \end{aligned} \quad (8.142)$$

onde lembramos que junto com vínculo

$$E = \int_0^\infty \rho(\epsilon) \frac{\epsilon}{e^{\beta(\epsilon - \mu)} + 1} d\epsilon, \quad (8.143)$$

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

$$N = \lim_{T \rightarrow 0} \int_0^\infty \rho(\epsilon) \frac{1}{e^{\beta(\epsilon - \mu)} + 1} d\epsilon, \quad (8.144)$$

usando que o número de ocupação se torna um função degrau, 1 para energias abaixo de μ_0 e zero acima

$$N = \int_0^{\mu_0} \rho(\epsilon) d\epsilon, \quad (8.145)$$

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 8.136. 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 a energia. Em geral se o spin for s o número será $g = 2s + 1$. portanto a densidade deve ser multiplicada por g :

$$\begin{aligned} N &= g \int_0^{\mu_0} \rho(\epsilon) d\epsilon, \\ &= gV \int_0^{\mu_0} C\epsilon^{\frac{1}{2}} d\epsilon = gVC \frac{2}{3} \epsilon_F^{\frac{3}{2}}, \end{aligned} \quad (8.146)$$

$$\begin{aligned} &= g \frac{4\pi V}{(2\pi\hbar)^3} \sqrt{2m^{\frac{3}{2}}} \frac{2}{3} \epsilon_F^{\frac{3}{2}} \\ &= \frac{gV(4\pi p_F^3)}{3(2\pi\hbar)^3}. \end{aligned} \quad (8.147)$$

Segue que

$$p_F = 2\pi\hbar \left(\frac{3}{4\pi g}\right)^{\frac{1}{3}} \left(\frac{N}{V}\right)^{\frac{1}{3}} \propto \left(\frac{N}{V}\right)^{\frac{1}{3}} \quad (8.148)$$

e

$$\epsilon_F = \left(\frac{3}{2gC}\right)^{\frac{2}{3}} \left(\frac{N}{V}\right)^{\frac{2}{3}} \quad (8.149)$$

$$\mu_0 = \epsilon_F = \frac{(2\pi\hbar)^2}{2m} \left(\frac{3}{4\pi g}\right)^{\frac{2}{3}} \left(\frac{N}{V}\right)^{\frac{2}{3}} \propto \left(\frac{N}{V}\right)^{\frac{2}{3}} \quad (8.150)$$

A energia média por partícula

$$\bar{\epsilon}(T, N) = \frac{g}{N} \int_0^{\infty} \rho(\epsilon) \frac{\epsilon}{e^{\beta(\epsilon-\mu)} + 1} d\epsilon \xrightarrow{T \rightarrow 0} \bar{\epsilon} = \frac{\int_0^{\epsilon_F} \epsilon \rho(\epsilon) d\epsilon}{\int_0^{\epsilon_F} \rho(\epsilon) d\epsilon} \quad (8.151)$$

$$\bar{\epsilon} = \frac{\int_0^{\epsilon_F} \epsilon^{\frac{3}{2}} d\epsilon}{\int_0^{\epsilon_F} \epsilon^{\frac{1}{2}} d\epsilon} = \frac{3}{5} \epsilon_F. \quad (8.152)$$

A energia total

$$E = \frac{3}{5} N \epsilon_F = 3 \frac{(2\pi\hbar)^2}{10m} \left(\frac{3}{4\pi g}\right)^{\frac{2}{3}} \left(\frac{N}{V}\right)^{\frac{2}{3}} N \propto (V)^{-\frac{2}{3}} N^{\frac{5}{3}} \quad (8.153)$$

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:

$$P = -\frac{\partial E}{\partial V} \Big|_N = \frac{2E}{3V}. \quad (8.154)$$

No caso clássico tínhamos $E = 3NT/2 = \frac{3}{2}PV$ zero energia a temperatura zero, agora também temos

$$E = \frac{3}{2}PV, \quad (8.155)$$

mas neste caso não é zero.

Exercício(verificar numeros) No caso de metais $m = 9.10^{-28}g$, $g = 2$, 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.

8.10.1 Gás ideal de Férmions a temperatura finita

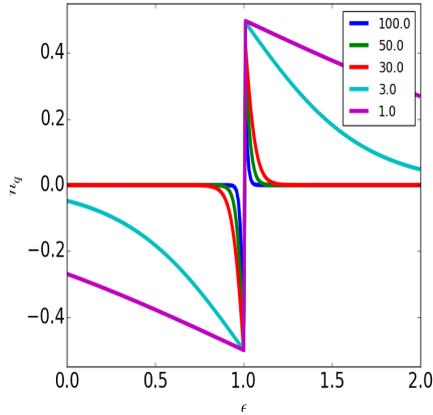


Figura 8.5: A função de Fermi a temperatura finita menos a função de Fermi a $T = 0$ (degrau): $f(\epsilon, \beta) = (\exp(\beta(\epsilon - 1)) + 1)^{-1} - \Theta(1 - \epsilon)$, para diferentes valores de β . A energia é medida em unidades do potencial químico.

Pelo exercício acima vemos que temperaturas da ordem de 300K ainda são muito baixas. Olhe para a figura 8.4 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 8.5. 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 8.5 é diferente de zero.

As equações relevantes são 8.142 e 8.143 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

$$A(T, \mu) = \int_0^{\infty} \frac{a(\epsilon)}{e^{\beta(\epsilon - \mu)} + 1} d\epsilon, \quad (8.156)$$

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

$$A(T, \mu) = A(0, \mu) + (A(T, \mu) - A(0, \mu)) \quad (8.157)$$

onde, usando a função f (ver figura 8.5)

$$A(0, \mu) = \int_0^{\mu} a(\epsilon) d\epsilon \quad (8.158)$$

$$\begin{aligned} \delta A = (A(T, \mu) - A(0, \mu)) &= \int_0^{\infty} a(\epsilon) \left(\frac{1}{e^{\beta(\epsilon - \mu)} + 1} - \Theta(\mu - \epsilon) \right) d\epsilon \\ &= \int_{-\infty}^{\infty} a(\epsilon) f(\epsilon, T) d\epsilon. \end{aligned} \quad (8.159)$$

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

$$\int_{-\infty}^{\infty} a(\epsilon)f(\epsilon, T)d\epsilon = \sum_k \frac{T^{k+1}}{k!} \frac{d^k a}{d\epsilon^k} \Big|_{\mu} \int_{-\infty}^{\infty} \left(\frac{\epsilon - \mu}{T}\right)^k f(\epsilon, T) \frac{d\epsilon}{T} \quad (8.160)$$

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

$$\int_{-\infty}^{\infty} u^{2k} f(u, T) du = 0 \quad (8.161)$$

se anulam e o primeiro termo na expansão que contribui é:

$$\begin{aligned} \int_{-\infty}^{\infty} u f(u, T) du &= \int_{-\infty}^{\infty} u \left(\frac{1}{e^u + 1} - \Theta(-u) \right) du \\ &= \int_{-\infty}^0 u \left(\frac{1}{e^u + 1} - 1 \right) du + \int_0^{\infty} \frac{u}{e^u + 1} du \\ &= - \int_{-\infty}^0 u \left(\frac{1}{1 + e^{-u}} \right) du + \int_0^{\infty} \frac{u}{e^u + 1} du \\ &= 2 \int_0^{\infty} \frac{u}{e^u + 1} du = \frac{\pi^2}{6}. \end{aligned} \quad (8.162)$$

O próximo termo na expansão será o cúbico

$$2 \int_0^{\infty} \frac{u^3}{e^u + 1} du = \frac{7\pi^4}{60}. \quad (8.163)$$

Note que o expoente de T em 8.160 é $k + 1$, portanto a série em potências de T para $A(\beta, \mu)$ na equação 8.156 tem expoentes pares:

$$A(T, \mu) = \int_0^{\mu} a(\epsilon) d\epsilon + \frac{da}{d\epsilon} \Big|_{\mu} \frac{\pi^2}{6} T^2 + \frac{d^3 a}{d\epsilon^3} \Big|_{\mu} \frac{7\pi^4}{360} T^4 + \dots \quad (8.164)$$

Agora vamos a brincar com a expressão acima:

8.10.2 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 $\rho = CVe^b$, como é nos casos de interesse. Integramos por partes:

$$\Phi = -\frac{CV}{b+1} e^{b+1} \frac{1}{e^{\beta(\epsilon - \mu)} + 1} \Big|_0^{\infty} - \int_0^{\infty} \frac{CV}{b+1} e^{b+1} \frac{1}{e^{\beta(\epsilon - \mu)} + 1} d\epsilon$$

o termo de superfície é nulo, e o segundo termo (que muda duas vezes de sinal) é proporcional à energia

$$\begin{aligned} \Phi &= - \int_0^{\infty} \frac{\epsilon}{b+1} \rho(\epsilon) \frac{1}{e^{\beta(\epsilon - \mu)} + 1} d\epsilon \\ &= - \frac{1}{b+1} \int_0^{\infty} \epsilon \rho(\epsilon) \bar{n}(\epsilon) d\epsilon \\ &= - \frac{E}{b+1} \end{aligned} \quad (8.165)$$

$$= - \frac{2}{3} E \quad (b = 1/2) \quad (8.166)$$

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 8.164 até a primeira ordem não trivial (primeira derivada de a) fica, para $b = 1/2$:

$$\begin{aligned}\Phi(\beta, \mu, V) &= -\frac{2CV}{3}\left(\frac{2}{5}\mu^{\frac{5}{2}} + \frac{\pi^2}{4}\mu^{\frac{1}{2}}T^2\right) \\ &= -\frac{4CV}{15}\mu^{\frac{5}{2}}\left(1 + \frac{5\pi^2}{8}\left(\frac{T}{\mu}\right)^2\right)\end{aligned}\quad (8.167)$$

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 8.136

$$\Phi(\beta, \mu, V) = -\frac{4}{15}\frac{4\pi Vg}{(2\pi\hbar)^3}\sqrt{2m^{\frac{3}{2}}}\mu^{\frac{5}{2}}\left(1 + \frac{5\pi^2}{8}\left(\frac{T}{\mu}\right)^2\right)\quad (8.168)$$

8.10.3 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

$$\begin{aligned}N = -\frac{\partial\Phi}{\partial\mu}\Big|_T &= \frac{2CV}{3}\left(\mu^{\frac{3}{2}} + \frac{\pi^2}{8}\mu^{-\frac{1}{2}}T^2\right) \\ &= \frac{2CV}{3}\mu^{\frac{3}{2}}\left(1 + \frac{\pi^2}{8}\left(\frac{T}{\mu}\right)^2\right).\end{aligned}\quad (8.169)$$

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

$$N(T = 0) = N_0 = \frac{2CV}{3}\mu^{\frac{3}{2}},\quad (8.170)$$

ou para um dado N o potencial químico será μ_0 , obtido invertendo 8.170

$$\mu_0 = \left(\frac{3N}{2CV}\right)^{\frac{2}{3}}.\quad (8.171)$$

Considerando N fixo temos duas expressões (8.169 e 8.171) que permitem, ao eliminar N obter uma relação entre os potenciais químicos a temperatura zero e finita:

$$N = \frac{2CV}{3}\mu_0^{\frac{3}{2}} = \frac{2CV}{3}\mu^{\frac{3}{2}}\left(1 + \frac{\pi^2}{8}\left(\frac{T}{\mu}\right)^2\right).\quad (8.172)$$

Mantendo a expansão até segunda ordem em T , esta expressão pode ser invertida para obter

$$\mu = \mu_0 \left(1 + \frac{\pi^2}{8} \left(\frac{T}{\mu} \right)^2 \right)^{-\frac{2}{3}} \quad (8.173)$$

$$= \mu_0 \left(1 - \frac{\pi^2}{12} \left(\frac{T}{\mu} \right)^2 \right) \quad (8.174)$$

$$= \mu_0 \left(1 - \frac{\pi^2}{12} \left(\frac{T}{T_F} \right)^2 \right) \quad (8.175)$$

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 $\mu > 0$.

8.10.4 A entropia S e o calor específico C_V

Usamos

$$S = - \left. \frac{\partial \Phi}{\partial T} \right|_{\mu, V}$$

e a expressão 8.167

$$S = \frac{CV\pi^2}{3} \mu^{\frac{1}{2}} T. \quad (8.176)$$

Mas por 8.170, $\mu^{\frac{1}{2}} = (3/2C)^{1/3} (N/V)^{1/3}$ e $\epsilon_F = (\frac{3}{2C})^{\frac{2}{3}} (\frac{N}{V})^{\frac{2}{3}}$

$$S = \frac{\pi^2 N}{2} \frac{2C}{3} \frac{V}{N} (3/2C)^{1/3} (N/V)^{1/3} T = \frac{\pi^2 N}{2} \frac{T}{T_F} \quad (8.177)$$

e o calor específico

$$C_V = T \left(\frac{\partial S}{\partial T} \right)_V = \frac{\pi^2 N}{2} \frac{T}{T_F}. \quad (8.178)$$

Podemos escrevê-lo de uma forma que lembra a equipartição

$$C_V = \frac{3}{2} N_{ef} \quad (8.179)$$

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.

⁵ Essencialmente nestas notas, como medimos temperatura em unidades de energia, a constante de Boltzmann $k_B = 1$.

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 ⁶.

6

8.11 Bósons

8.11.1 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 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 8.132:

$$\rho(\epsilon)d\epsilon = \mathcal{N}_3 = 2 \frac{L^3 \Delta p_x \Delta p_y \Delta p_z}{(2\pi\hbar)^3}, \quad (8.180)$$

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 ⁷.

$$\epsilon = \hbar\omega, \quad \mathbf{p} = \hbar\mathbf{q}, \quad |\mathbf{p}| = \frac{\hbar\omega}{c} \quad (8.181)$$

⁷ O campo do fóton é um vetor, tem spin 1, mas não são 3 estados porque a massa é nula

$$\rho(\epsilon)d\epsilon = 2 \frac{V 4\pi p^2}{(2\pi\hbar)^3} dp, \quad (8.182)$$

$$\rho(\epsilon)d\epsilon = \frac{8\pi V \omega^2}{(2\pi c)^3} d\omega, \quad (8.183)$$

O número de ocupação médio de um modo com frequência ω é obtido da expressão 8.126, só que com o detalhe que o potencial químico é nulo, pois como o foton 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:

$$\bar{n}_\omega^P = \frac{1}{e^{\beta\hbar\omega} - 1}. \quad (8.184)$$

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 :

$$\begin{aligned} E &= \int_0^\infty \frac{8\pi V\omega^2}{(2\pi c)^3} \frac{\hbar\omega}{e^{\beta\hbar\omega} - 1} d\omega \\ &= \frac{8\pi V\hbar}{(2\pi c)^3} \int_0^\infty \frac{\omega^3}{e^{\beta\hbar\omega} - 1} d\omega \end{aligned} \quad (8.185)$$

Podemos escrever $E = \int u(\omega, T) d\omega$, definindo a densidade de energia espectral:

$$u(\omega, T) = \frac{8\pi V\hbar}{(2\pi c)^3} \frac{\omega^3}{e^{\beta\hbar\omega} - 1} \quad (8.186)$$

que é o famoso resultado de Planck. Mudando variáveis $x = \beta\hbar\omega$ temos

$$\begin{aligned} E &= T^4 \frac{V}{\pi^2 (\hbar c)^3} \int_0^\infty \frac{x^3}{e^x - 1} dx \\ \frac{E}{V} &= \sigma_{SB} T^4 \end{aligned} \quad (8.187)$$

onde a última linha expressa o que é chamado de lei de Stephan-Boltzmann e a constante σ_{SB} pode ser calculada usando

$$\int_0^\infty \frac{x^3}{e^x - 1} dx = \frac{\pi^4}{15} \quad (8.188)$$

o que leva, após recolocar a constante de Boltzmann, a

$$\sigma_{SB} = \frac{\pi^2 k_B^4}{15 (\hbar c)^3} \quad (8.189)$$

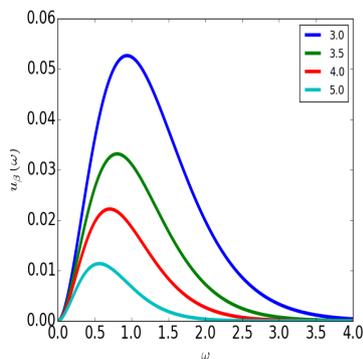


Figura 8.6: Resultado de Planck para a densidade espectral de radiação de corpo negro para diferentes valores de β . Unidades de densidade e β arbitrárias.

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.

8.11.2 Condensação de Bose-Einstein

Voltamos ao caso em que a densidade de estados é dada por 8.136. Estamos interessados em calcular

$$N = \sum_q \bar{n}_q \quad (8.190)$$

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 $\rho(\epsilon)$

$$\begin{aligned} \sum_q \bar{n}_q &= \sum_q \frac{1}{e^{\beta(\epsilon_{q_1} - \mu)} - 1} \\ \rightarrow N^*(T, \mu) &= \int_0^\infty \rho(\epsilon) \frac{1}{e^{\beta(\epsilon - \mu)} - 1} d\epsilon. \end{aligned} \quad (8.191)$$

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 8.191 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

$$\begin{aligned} N^*(T, 0) &= \int_0^\infty \rho(\epsilon) \frac{1}{e^{\beta\epsilon} - 1} d\epsilon. \\ &= CV \int_0^\infty \epsilon^{\frac{1}{2}} \frac{1}{e^{\beta\epsilon} - 1} d\epsilon. \\ &= CVT^{\frac{3}{2}} \int_0^\infty x^{\frac{1}{2}} \frac{1}{e^x - 1} dx. \end{aligned} \quad (8.192)$$

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 8.191 efetivamente calculava o número de partículas total, mas ao atingir a densidade $d_0(T)$

$$d_0(T) = \frac{N^*(T, 0)}{V} = C'T^{\frac{3}{2}} \quad (8.193)$$

com

$$C' = C \int_0^\infty x^{\frac{1}{2}} \frac{1}{e^x - 1} dx = C\Gamma\left(\frac{3}{2}\right)\zeta\left(\frac{3}{2}\right) \quad (8.194)$$

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

$$N = N_0 + N^* \quad (8.195)$$

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 8.192. 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 8.191. A temperatura em que $\mu = 0$ pode é dada por uma relação sutilmente diferente de 8.193

$$d = \frac{N}{V} = C'T_0^{\frac{3}{2}} \quad (8.196)$$

A diferença entre as equações 8.193 e 8.196 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 $T < T_0$

$$\begin{aligned} N^* &= C'VT^{\frac{3}{2}} \\ &= C'VT_0^{\frac{3}{2}}\left(\frac{T}{T_0}\right)^{\frac{3}{2}} \\ &= N\left(\frac{T}{T_0}\right)^{\frac{3}{2}} \end{aligned} \quad (8.197)$$

o número de partículas nos estados excitado diminui com a temperatura (elevada a 3/2). Portanto

$$N_0 = N - N^* = N\left(1 - \left(\frac{T}{T_0}\right)^{\frac{3}{2}}\right) \quad (8.198)$$

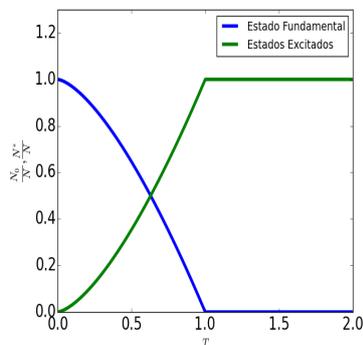


Figura 8.7: Frações dos números de partículas no estado fundamental e nos estados excitados com função de $\frac{T}{T_0}$, para N e V fixos.

Exercício Mostre que

$$\int_0^{\infty} x^s \frac{1}{e^x - 1} dx = \Gamma(s)\zeta(s) \quad (8.199)$$

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 $\zeta(3/2) =$.

8.11.3 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

$$\begin{aligned} E(T, \mu) &= \int_0^\infty \rho(\epsilon) \frac{\epsilon}{e^{\beta(\epsilon-\mu)} - 1} d\epsilon, \\ &= CVT^{\frac{5}{2}} \int_0^\infty x^{\frac{3}{2}} \frac{1}{e^x - 1} dx. \end{aligned} \quad (8.200)$$

$$\begin{aligned} E(T, \mu) &= CVT^{\frac{5}{2}} \Gamma\left(\frac{3}{2}\right) \zeta\left(\frac{3}{2}\right) \\ &= 0.770TN^* \end{aligned} \quad (8.201)$$

$$C_V = \frac{\partial E}{\partial T} \Big|_{V, N} \quad (8.202)$$

$$= 1.9N \left(\frac{T}{T_0}\right)^{\frac{3}{2}} \quad (8.203)$$

$$P = \frac{2}{3} \frac{E}{V} \propto N^* \frac{T}{V} \quad (8.204)$$

9

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. Estes 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.

9.1 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:

$$\mathcal{H} = - \sum_{\langle ij \rangle} J_{ij} s_i s_j - \sum_i h_i s_i. \quad (9.1)$$

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

$$P(J_{ij}) = \sqrt{\frac{N}{2\pi J^2}} e^{-\frac{N(J_{ij}-J_0/N)^2}{2J^2}}, \quad (9.2)$$

portanto

$$\begin{aligned} \langle\langle J_{ij} \rangle\rangle_J &= \frac{J_0}{N} \\ \sigma_J^2 = \langle\langle J_{ij}^2 \rangle\rangle_J - \langle\langle J_{ij} \rangle\rangle_J^2 &= \frac{J^2}{N} \end{aligned} \quad (9.3)$$

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.

9.2 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 .

9.2.1 Matriz de Transferência

Vamos supor um modelo de um ferromagneto nas seguintes condições experimentais: Uma amostra é colocada em um campo magnético externo h .

$$\mathcal{H} = -J \sum_{\langle ij \rangle} s_i s_{i+1} - h \sum_i s_i. \quad (9.4)$$

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

$$m(\beta, h, N) = \left\langle \frac{1}{N} \sum_i s_i \right\rangle = \sum_{\{s_1, s_2, \dots, s_N\}} \left(\frac{1}{N} \sum_i s_i \right) P(s_1, s_2, \dots, s_N) \quad (9.5)$$

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:

$$m(\beta) = \lim_{h \rightarrow 0^+} m(\beta, h). \quad (9.6)$$

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

$$\begin{aligned} \mathcal{H} &= -J \sum_{i=1}^N s_i s_{i+1} - \frac{h}{2} \sum_{i=1}^N (s_i + s_{i+1}) \\ &= - \sum_{i=1}^N \left(J s_i s_{i+1} - \frac{h}{2} (s_i + s_{i+1}) \right) = \sum_{i=1}^N \mathcal{H}_{s_i, s_{i+1}} \end{aligned} \quad (9.7)$$

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

$$P(\{s_i\}) = \frac{e^{-\beta H}}{Z}. \quad (9.8)$$

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

$$f(\beta, h) = - \lim_{N \rightarrow \infty} \frac{1}{N\beta} \log Z(\beta, h, N). \quad (9.9)$$

e a magnetização

$$m(\beta, h) = - \frac{\partial f(\beta, h)}{\partial h} = \lim_{N \rightarrow \infty} \frac{1}{N\beta} \frac{\partial \log Z(\beta, h, N)}{\partial h} \quad (9.10)$$

Com as condições de contorno periódicas

$$Z(\beta, h) = \sum_{\{s_1, s_2, \dots, s_N\}} \prod_i e^{-\beta \mathcal{H}_{s_i, s_{i+1}}} \quad (9.11)$$

A matriz de transferência \mathbf{T} é introduzida definindo seus elementos através de

$$(\mathbf{T})_{s_i, s_{i+1}} = \exp \beta \mathcal{H}_{s_i, s_{i+1}} \quad (9.12)$$

$$\mathbf{T} = \begin{pmatrix} T_{++} & T_{+-} \\ T_{-+} & T_{--} \end{pmatrix}$$

$$\mathbf{T} = \begin{pmatrix} e^{\beta(J+h)} & e^{-\beta J} \\ e^{-\beta J} & e^{\beta(J-h)} \end{pmatrix}$$

Se $\mathbf{B} = \mathbf{A}\mathbf{A}$ 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.

$$\begin{aligned} Z(\beta, h) &= \sum_{\{s_1, s_2, \dots, s_N\}} T_{s_1, s_2} T_{s_2, s_3} \dots T_{s_i, s_{i+1}} \dots T_{s_N, s_1} \\ &= \text{tr}(\mathbf{T}^N) \end{aligned} \quad (9.13)$$

Usamos dois resultados simples de álgebra linear:

(i) o traço de produto de matrizes é invariante ante permutações cíclicas: $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$. 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 $I = \mathbf{U}\mathbf{U}^{-1}$ entre os fatores \mathbf{T}

$$\begin{aligned} Z(\beta, h, N) &= \text{tr}(\mathbf{T}^N) \\ &= \text{tr}(\mathbf{U}\mathbf{U}^{-1} \mathbf{T}^N) = \text{tr}(\mathbf{U}^{-1} \mathbf{T}^N \mathbf{U}) \\ &= \text{tr}(\mathbf{U}^{-1} \mathbf{T} \mathbf{U} \mathbf{U}^{-1} \dots \mathbf{T} \mathbf{U} \mathbf{U}^{-1} \mathbf{T} \mathbf{U}) \\ &= \text{tr}(\mathbf{T}_D^N) \\ &= \lambda_1^N + \lambda_2^N \end{aligned} \quad (9.14)$$

onde $\lambda_{1,2}$ são os autovalores de \mathbf{T} obtidos das raízes de

$$\det(\mathbf{T} - \lambda I) = 0,$$

então

$$\lambda^2 - 2\lambda e^{\beta J} \cosh(\beta h) + 2 \sinh(2\beta J) = 0$$

$$\begin{aligned}\lambda_{1,2} &= e^{\beta J} \cosh(\beta h) \pm \sqrt{e^{2\beta J} \cosh^2(\beta h) - 2 \sinh(2\beta J)} \\ \lambda_{1,2} &= \left(e^{\beta J} \cosh(\beta h) \pm \sqrt{e^{2\beta J} \cosh^2(\beta h) - e^{2\beta J} + e^{-2\beta J}} \right) \\ \lambda_{1,2} &= e^{\beta J} \left(\cosh(\beta h) \pm \sqrt{\sinh^2(\beta h) + e^{-4\beta J}} \right) \quad (9.15)\end{aligned}$$

Exercício Mostre que $\lambda_1 > \lambda_2 > 0$, para qualquer escolha de β e h .

Para N grande

$$\begin{aligned}Z(\beta, h, N) &= \lambda_1^N + \lambda_2^N \\ &= \lambda_1^N \left(1 + \left(\frac{\lambda_2}{\lambda_1} \right)^N \right) \\ &\rightarrow \lambda_1^N \quad (9.16)\end{aligned}$$

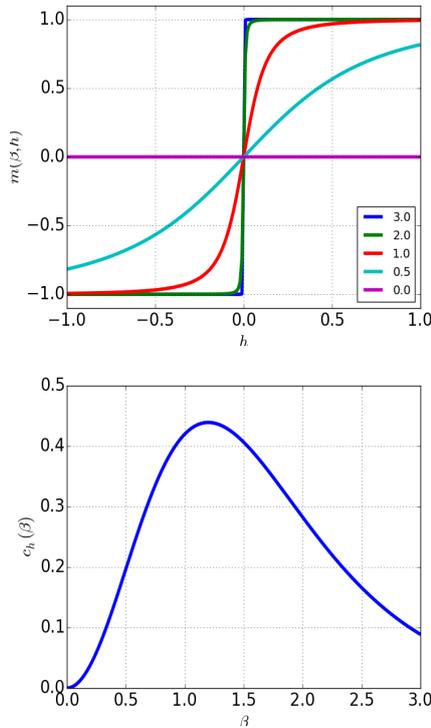


Figura 9.1: Ising 1d. Esquerda: Magnetização $m(\beta, h)$ como função de h para diferentes valores de β . Note que todas as curvas passam pela origem: não há magnetização espontânea. Direita: Calor específico Unidades: $J = 1$).

Segue que a magnetização é dada por

$$\begin{aligned}m(\beta, h) &= \frac{1}{\beta} \frac{\partial \log \lambda_1}{\partial h} \\ &= \frac{\sinh(\beta h) + \frac{\sinh(\beta h) \cosh(\beta h)}{\sqrt{\sinh^2(\beta h) + e^{-4\beta J}}}}{\cosh(\beta h) + \sqrt{\sinh^2(\beta h) + e^{-4\beta J}}} \\ &= \frac{\sinh(\beta h)}{\sqrt{\sinh^2(\beta h) + e^{-4\beta J}}} \xrightarrow{h \rightarrow 0} 0 \quad (9.17)\end{aligned}$$

portanto não existe ordem, que seria sinalizado por $m \neq 0$, a não ser que haja um campo externo $h \neq 0$.

O calor específico pode ser calculado de

$$E(T, N, h) = -\frac{\partial \log Z}{\partial \beta} = -N \frac{\partial \log \lambda_1}{\partial \beta}$$

$$E(T, N, h) = -\frac{N}{\lambda_1} \frac{\partial \lambda_1}{\partial \beta} = -\frac{N}{\lambda_1} \left(J \lambda_1 + e^{\beta J} (h \sinh \beta h + \frac{h \sinh \beta h \cosh \beta h - 2e^{-4\beta J}}{\sqrt{\sinh^2(\beta h) + e^{-4\beta J}}}) \right)$$

$$E(T, N, h = 0) = E(T, N) = -\frac{N}{\lambda_{1,h=0}} \frac{\partial \lambda_{1,h=0}}{\partial \beta} = -NJ - \frac{Ne^{-\beta J}}{2 \cosh \beta J} \quad (9.18)$$

e, usando $C_h = \frac{\partial E}{\partial T}$ chegamos a

$$c_h(\beta) = \frac{1}{N} C_h(h = 0) = (\beta J)^2 \operatorname{sech}^2(\beta J) \quad (9.19)$$

Note que o pico no calor específico é muito bem comportado e não está associado com uma transição de fase.

9.3 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 Opper 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

$$P(\{s_i\}) = \frac{e^{-\beta \mathcal{H}}}{Z}. \quad (9.20)$$

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 agora $Z(\beta, h) = \sum_{\{s_i\}} \exp(-\beta \mathcal{H})$. 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

$$f(\beta, h) = -\lim_{N \rightarrow \infty} \frac{1}{N\beta} \log Z(\beta, h). \quad (9.21)$$

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 9.21. 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

$$KL(Q||P) = \sum_{\{s_i\}} Q \log \frac{Q}{P}. \tag{9.22}$$

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. Um clique é simplesmente um conjunto finito de variáveis interagentes. Por exemplo, sítios individuais ou pares de spins contíguos. Um passo além, mas ainda tratável, é dado tomando Q 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: $Q = \prod_i Q_i(s_i)$.

9.3.1 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

$$Q(s_i|m_i) = (1 + s_i m_i)/2$$

$$Q = \prod_i Q_i(s_i|m_i) = \prod_i \frac{1 + s_i m_i}{2}. \tag{9.23}$$

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

$$\langle s_i \rangle = \sum s_i \frac{1 + s_i m_i}{2} = m_i. \quad (9.24)$$

Repetimos: que escolha fazemos para o conjunto dos $\{m_i\}$?

Minimizamos a distância de Kullback-Leibler. Equivalentemente maximizamos a entropia cruzada, com P a distribuição de Boltzmann como distribuição *a priori* e Q a ser determinada como a posterior:

$$\begin{aligned} -S[Q||P] &= KL(Q||P) = \sum_{\{s_i\}} Q(s) \log \frac{QZ}{e^{-\beta \mathcal{H}}} \\ &= \log Z + \beta E(Q) - S(Q). \end{aligned} \quad (9.25)$$

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 :

$$E(Q) = \sum_{\{s_i\}} Q(s) \mathcal{H}(s) = \langle \mathcal{H}(s) \rangle_Q \quad (9.26)$$

e a entropia de Q :

$$S(Q) = - \sum_{\{s_i\}} Q(s) \log Q(s) = - \langle \log Q \rangle_Q. \quad (9.27)$$

Substituindo Q da equação 9.23 obtém-se

$$E(Q) = - \sum_{\langle ij \rangle} J_{ij} m_i m_j - \sum_i h m_i \quad (9.28)$$

e

$$S(Q) = - \sum_i \left[\frac{1 + m_i}{2} \log \frac{1 + m_i}{2} + \frac{1 - m_i}{2} \log \frac{1 - m_i}{2} \right] \quad (9.29)$$

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 parece familiar aos físicos. Devido a convexidade de $\log x$ temos que $\log x \leq x - 1$, assim colocando $x = P_i/Q_i$ temos

$$\begin{aligned} KL(Q||P) &= - \sum Q_s \log \frac{P_s}{Q_s} \geq \sum Q_s (1 - P_s/Q_s) \\ KL(Q||P) &\geq \sum (Q_s - P_s) = 1 - 1 = 0, \end{aligned} \quad (9.30)$$

preencha os detalhes da demonstração acima. Assim a energia livre exata é limitada por $F(Q)$

$$- \log Z \leq \beta E(Q) - S(Q). \quad (9.31)$$

e vemos de outra maneira que faz sentido encontrar o valor mínimo de F . Esta é a famosa desigualdade de Bogoliuvov e aqui temos de um ponto de vista geométrico. A minimização leva a que $\langle s_i \rangle = m_i$ satisfaz

$$m_i = \tanh \left(\beta \left(\sum_j J_{ij} m_j + h_i \right) \right) \quad (9.32)$$

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

$$m = \tanh(\beta(\nu Jm + h)) \tag{9.33}$$

e a energia livre por partícula

$$f = E - \beta^{-1}S = -\frac{\nu J}{2}m^2 - hm + \beta^{-1}\left(\frac{1-m}{2} \log \frac{1-m}{2} + \frac{1+m}{2} \log \frac{1+m}{2}\right) \tag{9.34}$$

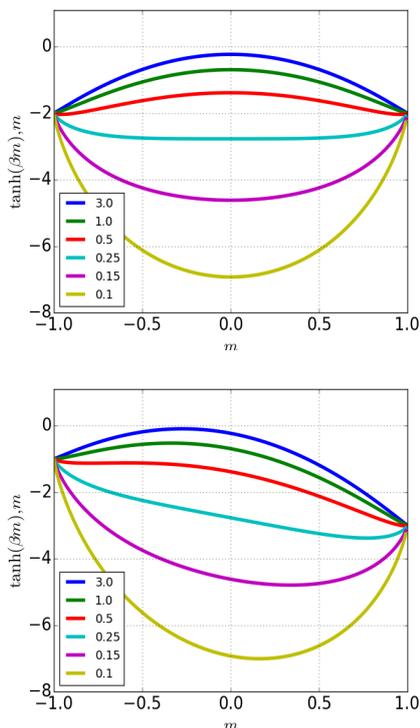


Figura 9.2: Energia livre $f = E - \beta^{-1}S$ na aproximação de Campo Médio para $J_{ij} = J = 1$ e $\nu = 4$. Esquerda: $h = 0$, Direita: $h = 1/2$.

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 9.32

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 $\tanh^{-1} y = \frac{1}{2} \log \frac{1+y}{1-y}$.)

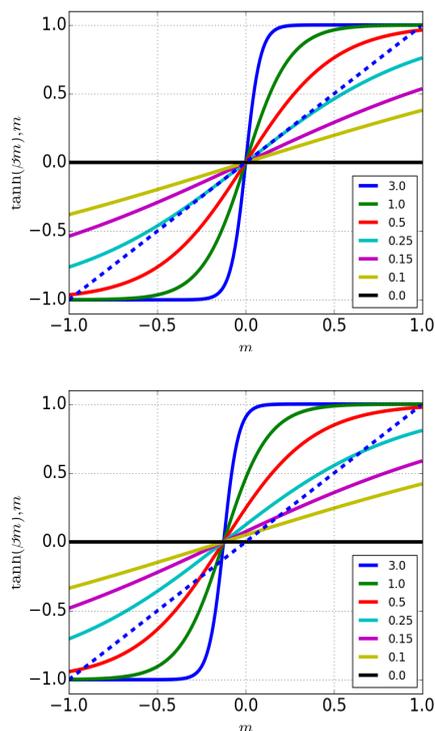


Figura 9.3: Campo Médio para $J_{ij} = J = 1$ e $\nu = 4$. A magnetização $m(\beta, h)$ é determinada pela interseção da tangente hiperbólica (curvas contínuas para diferentes valores de β) e da identidade (curva tracejada), ver equação 9.33. Esquerda: $h = 0$, Direita: $h = 1/2$.

- 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:

$$\langle s_i \rangle = \left\langle \tanh \left(\sum_j J_{ij} s_j + h_i \right) \right\rangle, \quad (9.35)$$

segundo as instruções seguintes para campo nulo. Considere o sistema obtido sem o spin no sítio i : $\mathcal{H}_{/i} = \mathcal{H} + \sum_j J_{ij} s_i s_j$. Mostre

que

$$\begin{aligned}\sum_{s_i} s_i e^{-\beta \mathcal{H}} &= 2e^{-\beta \mathcal{H}/i} \sinh(\sum_j J_{ij}) \\ \sum_{s_i} e^{-\beta \mathcal{H}} &= 2e^{-\beta \mathcal{H}/i} \cosh(\sum_j J_{ij})\end{aligned}\quad (9.36)$$

Há uma semelhança óbvia entre a equação de Callen e a aproximação de CM (eq. 9.32). 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.

¹ A região de interesse para um dado problema por ser longe da criticalidade.

9.3.2 Campo médio: Blume-Capel spin 1

A possibilidade de outros tipos de variáveis levou à introdução de variáveis clássicas de vários tipos diferentes interagindo de formas diversas. Isto pode levar a tipos de transições diferentes. Neste caso os íons magnéticos modelados tem spin 1 (triplete) (e.g. M. Blume, Phys. Rev (1966), H. W. Capel Physica 32 966 (1966)). O Hamiltoniano é

$$\mathcal{H} = - \sum_{i,j} J_{ij} S_i S_j + D \sum_i S_i^2 - h \sum_i S_i \quad (9.37)$$

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

$$\begin{aligned}\sum_{S_i=0,\pm 1} Q_i &= 1, \\ \sum_{S_i=0,\pm 1} Q_i S_i &= m_i, \\ \sum_{S_i=0,\pm 1} Q_i S_i^2 &= r_i,\end{aligned}\quad (9.38)$$

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 9.22. A energia livre é $f = E - TS$ com

$$E(Q) = \sum_{\{s_i\}} Q(s) \mathcal{H}(S) = \langle \mathcal{H}(S) \rangle_Q \quad (9.39)$$

e a entropia de Q :

$$S(Q) = - \sum_{\{s_i\}} Q(S) \log Q(S) = - \langle \log Q \rangle_Q. \quad (9.40)$$

Substituindo Q da equação 9.38 obtém-se

$$E(Q) = - \sum_{\langle ij \rangle} J_{ij} m_i m_j + D \sum_i r_i - h \sum_i m_i \quad (9.41)$$

e

$$S(Q) = - \sum_i \left[\frac{r_i + m_i}{2} \log \frac{r_i + m_i}{2} + (1 - r_i) \log(1 - r_i) + \frac{r_i - m_i}{2} \log \frac{r_i - m_i}{2} \right] \quad (9.42)$$

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

$$\frac{\partial f}{\partial m} = 0 \quad (9.43)$$

$$\frac{\partial f}{\partial r} = 0 \quad (9.44)$$

que levam, a primeira a

$$\begin{aligned} \beta(m\nu J + h) &= \frac{1}{2} \log \frac{r + m}{r - m} \\ m &= r \tanh \beta(m\nu J + h) \end{aligned} \quad (9.45)$$

onde usamos novamente que $\tanh^{-1} y = \frac{1}{2} \log \frac{1+y}{1-y}$, e a segunda a

$$e^{-2\beta D} = \frac{r^2 - m^2}{4(1 - r)^2}. \quad (9.46)$$

Esta equação pode ser simplificada substituindo m pela tangente hiperbólica (eq. 9.45), levando a

$$r = \frac{2e^{-\beta D} \cosh(\beta(JM^* + h))}{2e^{-\beta D} \cosh(\beta(JM^* + h)) + 1} \quad (9.47)$$

As duas equações 9.45 e 9.46 podem ser resolvidas numericamente por iteração para cada valor de β, h, D para obter a solução das propriedades termodinâmicas como mostra a figura 9.4.

9.4 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

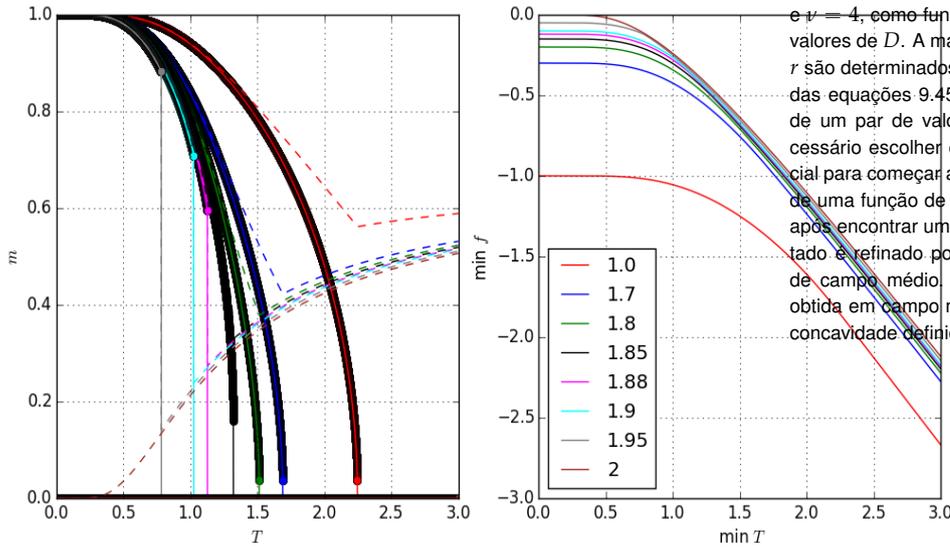


Figura 9.4: Esquerda: m (círculos) e r (tracejada) na aproximação de campo médio para o modelo de Blume Capel com $J_{ij} = J = 1$ e $r = 4$, como função de β , para diferentes valores de D . A magnetização $m(\beta, h, D)$ e r são determinados pela solução numérica das equações 9.45 e 9.47 iterando 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.

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.

9.4.1 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 sistemas de spin o método também sugerirá parâmetros de ordem úteis.

Incluimos o acoplamento J em β e o Hamiltoniano é dado por

$$H_N(s, h) = -\frac{1}{N} \sum_{\langle i, j \rangle} s_i s_j - h \sum_i s_i \tag{9.48}$$

Temos $N(N - 1)/2$ pares e

$$Z_N(\beta, h) = \sum_{\{s_i\}} e^{-\beta H} \tag{9.49}$$

define a variável

$$m = \frac{1}{N} \sum_i s_i \tag{9.50}$$

em termos da qual podemos escrever o Hamiltoniano

$$H_N(s, h) = -\frac{N}{2}m^2 - hNm \quad (9.51)$$

que obviamente extensivo e a função de partição

$$Z_N(\beta, h) = \sum_{\{s_i\}} e^{\frac{N\beta}{2}m^2 + \beta hNm} \quad (9.52)$$

A relação que será fundamental aqui para obter a solução é o fato simples que

$$1 = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx \quad (9.53)$$

Podemos mudar variáveis: $x = (M - m)\sqrt{\beta N}$ e assim obter

$$e^{\frac{1}{2}\beta Nm^2} = \frac{1}{\sqrt{2\pi/(\beta N)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\beta NM^2 + \beta NmM} dM \quad (9.54)$$

Este truque, descoberto por vários autores², 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.

² Acho que o primeiro foi Feynman, mas Hubbard e Stratonovich também costumam ser citados, embora Gauss já devia conhecê-lo.

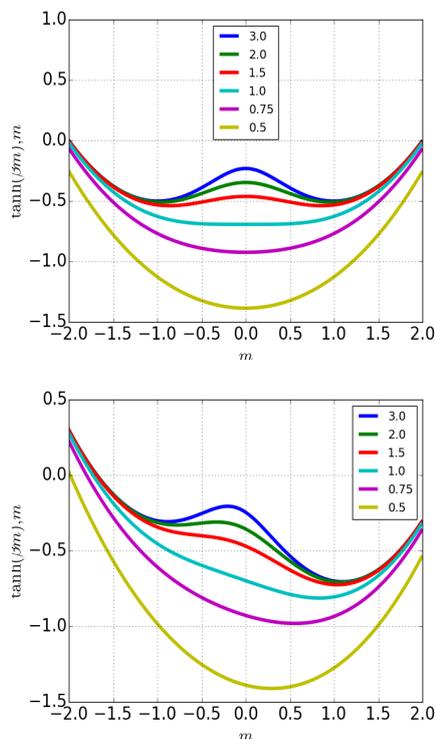


Figura 9.5: f para modelo de alcance infinito, ver equação 9.59. Esquerda: $h = 0$, Direita: $h = .1$.

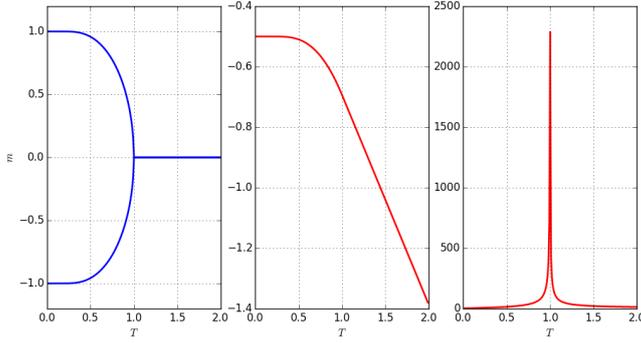


Figura 9.6: Modelo Ising de spin $\frac{1}{2}$ de alcance infinito. Esquerda: magnetização contra T . Centro: $\min f$ contra T . Direita: número de iterações de $m_{t+1} = \tanh \beta m_t$ até que $|m_{t+1} - m_t| < 10^{-6}$. A convergência da dinâmica das iterações fica muito mais lenta no ponto crítico.

O preço desta enorme simplificação é pequeno: uma única integral sobre M . A função de partição fica

$$\begin{aligned}
 Z_N(\beta, h) &= \frac{1}{\sqrt{2\pi/(\beta N)}} \sum_{\{s_i\}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\beta N M^2 + \beta N m(M+h)} dM \\
 &= \frac{1}{\sqrt{2\pi/(\beta N)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\beta N M^2} \left(\sum_{s_i} e^{\beta(M+h)s_i} \right)^N dM \\
 &= \frac{1}{\sqrt{2\pi/(\beta N)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\beta N M^2} (2 \cosh(\beta(M+h)))^N dM \\
 &= \frac{1}{\sqrt{2\pi/(\beta N)}} \int_{-\infty}^{\infty} e^{-\beta N f(M, h)} dM \quad (9.55)
 \end{aligned}$$

onde

$$f(M, h) = \frac{1}{2} M^2 - \frac{1}{\beta} \log(2 \cosh(\beta(M+h))) \quad (9.56)$$

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 9.55. 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

$$m^* = \tanh(\beta(m^* + h)), \quad (9.57)$$

expandindo em série de Taylor em torno do máximo:

$$f(M, h) = f(m^*, h) + \frac{1}{2} \left(\frac{\partial^2 f(M, h)}{\partial M^2} \right)_{|m^*} (M - m^*)^2 + \dots \quad (9.58)$$

onde

$$f(m^*, h) = \frac{1}{2} m^{*2} - \frac{1}{\beta} \log(2 \cosh(\beta(m^* + h))). \quad (9.59)$$

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

³ central..novamente, soma de N variáveis s_i , surpresa?

nome de ponto de sela. Aqui m^* é simplesmente um máximo do integrando.

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

$$Z(\beta, h, N) = e^{-N\beta f(m^*)} \frac{1}{\sqrt{2\pi/(\beta N)}} \int_{-\infty}^{\infty} e^{-\frac{\beta N}{2} \frac{\partial^2 f(M, h)}{\partial M^2} \Big|_{m^*} (M - m^*)^2} dM \quad (9.60)$$

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

$$\frac{\partial \log Z}{\partial h} = \beta \langle \sum_{i=1}^N s_i \rangle = N\beta M^* \quad (9.61)$$

onde M^* é a magnetização por partícula, ou o valor esperado $\langle s_i \rangle$. Agora tomamos a derivada do log da equação 9.60 vendo que a integral é um fator trivial que não contribui, para obter que

$$M^* = \tanh(\beta m^* + h), \quad (9.62)$$

comparando com a equação 9.57 finalmente obtemos que

$$M^* = m^*, \quad (9.63)$$

ou seja o valor de M que torna $f(M, h)$ um extremo é a magnetização do sistema.

9.4.2 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

$$f(px_1 + (1-p)x_2) \geq pf(x_1) + (1-p)f(x_2). \quad (9.64)$$

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

$$\begin{aligned} Nm &= N_1 m_1 + N_2 m_2 \\ m &= pm_1 + (1-p)m_2 \end{aligned} \quad (9.65)$$

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

$$m^2 \leq \frac{N_1}{N} m_1^2 + \frac{N_2}{N} m_2^2 \quad (9.66)$$

segue que

$$Z_N(\beta, h) \leq Z_{N_1}(\beta, h) Z_{N_2}(\beta, h) \quad (9.67)$$

e a subaditividade:

$$\frac{1}{N} \log Z_N(\beta, h) \leq \frac{1}{N} \log Z_{N_1}(\beta, h) + \frac{1}{N} \log Z_{N_2}(\beta, h). \quad (9.68)$$

É claro que podemos separar os grupos em subgrupos, até chegar a N grupos de um só spin

$$\frac{1}{N} \log Z_N(\beta, h) \leq \frac{N}{N} \log Z_1(\beta, h) = \log\left(\frac{1}{2}\beta + 2 \cosh(\beta h)\right) \quad (9.69)$$

Temos uma cota uniforme em N e podemos escrever

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log Z_N(\beta, h) \leq \log\left(\frac{1}{2}\beta + 2 \cosh(\beta h)\right) \quad (9.70)$$

uma cota superior para a energia livre no limite termodinâmico.

Para mostrar que o PS é exato começamos pela definição da equação 9.50

$$m = \frac{1}{N} \sum_i s_i$$

e, para qualquer M

$$(M - m)^2 \geq 0 \rightarrow m^2 \geq 2mM - M^2 \quad (9.71)$$

portanto

$$Z_N(\beta, h) \geq \sum_{s_1, s_2, \dots, s_N} e^{(2mM - M^2)\beta N / 2 + \beta h N m}. \quad (9.72)$$

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 nos levará a buscar uma outra desigualdade no sentido oposto, mas isso após continuar um pouco mais:

$$\begin{aligned} Z_N(\beta, h) &\geq e^{-\frac{M^2 \beta N}{2}} \sum_{s_1, s_2, \dots, s_N} e^{\beta(h+M) \sum_i s_i} \\ &\geq e^{-\frac{M^2 \beta N}{2}} \left(\sum_{s_i} e^{\beta(h+M) s_i} \right)^N \\ &\geq e^{-\frac{M^2 \beta N}{2}} (2 \cosh(\beta(h+M)))^N \end{aligned} \quad (9.73)$$

e

$$\frac{1}{N} \log Z_N(\beta, h) \geq -\frac{M^2 \beta}{2} + \log(2 \cosh(\beta(h+M))) \quad (9.74)$$

O lado direito é independente de N e temos uma cota inferior uniforme. Lembramos que o valor de M é arbitrário e podemos

escolhé-lo da melhor forma possível, tomamos o supremo. Agora fica facil tomar o limite termodinâmico:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log Z_N(\beta, h) \geq \sup_M \left[-\frac{M^2 \beta}{2} + \log(2 \cosh(\beta(h + M))) \right] \quad (9.75)$$

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:

$$Z_N(\beta, h) = \sum_M \sum_{\{s_i\}} \delta_{Mm} e^{\frac{N\beta}{2} m^2 + \beta h N m} \quad (9.76)$$

Agora usamos a profunda relação $m^2 = 2mM - M^2$ para $M = m$ obtendo

$$Z_N(\beta, h) = \sum_M \sum_{\{s_i\}} \delta_{Mm} e^{\frac{N\beta}{2} (2mM - M^2) + \beta h N m} \quad (9.77)$$

Novamente linearizamos o expoente com respeito a m . Usamos outro fato trivial: $\delta_{mM} \leq 1$, logo

$$\begin{aligned} Z_N(\beta, h) &\leq \sum_M e^{-\frac{N\beta}{2} M^2} \sum_{s_i} e^{\beta(M+h)Nm} \\ Z_N(\beta, h) &\leq \sum_M e^{-\frac{N\beta}{2} M^2 + N \log(2 \cosh(\beta(M+h)))} \end{aligned} \quad (9.78)$$

Há $N + 1$ termos na soma acima, podemos substituir M pelo argumento do supremo M^* e obter

$$Z_N(\beta, h) \leq (N + 1) e^{N(-\frac{\beta}{2} M^2 + \log(2 \cosh(\beta(M+h))))} \quad (9.79)$$

$$\frac{1}{N} \log Z_N(\beta, h) \leq \frac{1}{N} \log(N + 1) + \sup_M \left(-\frac{\beta}{2} M^2 + \log(2 \cosh(\beta(M + h))) \right) \quad (9.80)$$

Tomamos o limite $N \rightarrow \infty$ e temos uma cota superior. Juntando a equação 9.74 temos o

Teorema

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log Z_N(\beta, h) = \sup_M \left[-\frac{M^2 \beta}{2} + \log(2 \cosh(\beta(h + M))) \right] \quad (9.81)$$

mostrando que no limite termodinâmico a aproximação de ponto de sela é exata.

9.4.3 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 9.37, e $J_{ij} = J/N$, podemos repetir o que nos levou à equação 9.55

$$\begin{aligned}
 Z_N(\beta, h) &= \frac{1}{\sqrt{2\pi/(\beta N)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\beta N J M^2} \left(\sum_{S_i} e^{\beta(JM+h)s_i - \beta D S_i^2} \right)^N dM \\
 &= \frac{1}{\sqrt{2\pi/(\beta N)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\beta N J M^2} \left(2e^{-\beta D} \cosh(\beta(M+h)) + 1 \right)^N dM \\
 &= \frac{1}{\sqrt{2\pi/(\beta N)}} \int_{-\infty}^{\infty} e^{-\beta N g(M, h, D)} dM \\
 &\propto e^{-\beta N g(M^*, h, D)} \tag{9.82}
 \end{aligned}$$

onde

$$g(M, h, D) = \frac{1}{2} \beta N J M^2 - \frac{1}{\beta} \log \left(2e^{-\beta D} \cosh(\beta(JM+h)) + 1 \right) \tag{9.83}$$

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

$$\frac{\partial \log Z}{\partial D} = \beta \left\langle \sum_{i=1}^N S_i^2 \right\rangle = -N \beta r_\infty \tag{9.84}$$

que neste caso leva a

$$r_\infty = \frac{\partial g}{\partial D} = \frac{2e^{-\beta D} \cosh(\beta(JM^*+h))}{2e^{-\beta D} \cosh(\beta(JM^*+h)) + 1} \tag{9.85}$$

e para a magnetização, usando a expressão para r_∞ a

$$M^* = r_\infty \tanh(\beta(JM^*+h)), \tag{9.86}$$

Mostrando que as equações 9.86 e 9.46 tem exatamente a mesma forma, e portanto as mesmas soluções que para a aproximação de campo médio (equações 9.45 e 9.85).

Mostre que também vale o

Teorema

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log Z_N(\beta, h) = \sup_M \left[-\frac{M^2 \beta}{2} + \log(2e^{-\beta D} \cosh(\beta(M+h)) + 1) \right] \tag{9.87}$$

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.

9.5 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

$$Z(\kappa_M, g_M) = \sum_{\{s_1, s_3, \dots, s_{N-1}\}} \sum_{s_2, \dots, s_{2i}, s_{2M}} e^{-\mathcal{H}_i^M} = \sum_{\{s_1, s_3, \dots, s_{N-1}\}} \prod_{i=1}^{2^{M-1}} \sum_{s_{2i}} e^{-\mathcal{H}_i^M} \quad (9.88)$$

onde, para qualquer valor de m definimos

$$\mathcal{H}_i^m = -\kappa_m (s_{2i-1} + s_{2i+1})s_{2i} - \frac{g_m}{2} (s_{2i-1} + 2s_{2i} + s_{2i+1}) \quad (9.89)$$

lembrando que os valores iniciais, na resolução mais fina, são $\kappa_M = \beta J$ e $g_M = \beta h$. Mostraremos que

$$Z(\kappa_M, g_M) \propto \sum_{\{s_1, s_3, \dots, s_{N-1}\}} e^{-\mathcal{H}_i^{M-1}} = Z(\kappa_{M-1}, g_{M-1}) \quad (9.90)$$

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 9.89, e precisamos calcular as novas constantes κ_{M-1} e h_{M-1}

$$\sum_{s_{2i}} e^{-\mathcal{H}_i^M} = e^{\delta f_M} e^{-\mathcal{H}_i^{M-1}} = e^{\delta f_M} e^{\kappa_{M-1}s_{2i-1}s_{2i+1}+g_{M-1}(s_{2i-1}+s_{2i+1})} \quad (9.91)$$

onde δf_M é uma constante independente de s_{2i-1} e s_{2i+1} .

Precisamos impor

$$\begin{aligned} e^{\delta f_M} e^{\kappa_{M-1}+2g_{M-1}} &= e^{2\kappa_M+2g_M} + e^{-2\kappa_M} \quad \text{para } s_{2i-1} = s_{2i+1} = 1, \\ e^{\delta f_M} e^{-\kappa_{M-1}} &= e^{g_M} + e^{-g_M} \quad \text{para } s_{2i-1} = 1, s_{2i+1} = -1, \\ e^{\delta f_M} e^{-\kappa_{M-1}} &= e^{g_M} + e^{-g_M} \quad \text{para } s_{2i-1} = -1, s_{2i+1} = 1, \\ e^{\delta f_M} e^{\kappa_{M-1}-2g_{M-1}} &= e^{2\kappa_M-2g_M} + e^{-2\kappa_M} \quad \text{para } s_{2i-1} = s_{2i+1} = -1, \end{aligned}$$

a razão entre a primeira e a quarta diz como muda o campo externo efetivo $g_M \rightarrow g_{M-1}$

$$e^{4g_{M-1}} = \frac{e^{2\kappa_M+2g_M} + e^{-2\kappa_M}}{e^{2\kappa_M-2g_M} + e^{-2\kappa_M}} \quad (9.92)$$

o produto

$$\begin{aligned} e^{2\kappa_{M-1}} &= e^{-2\delta f_M} (e^{2\kappa_M+2g_M} + e^{-2\kappa_M}) (e^{2\kappa_M-2g_M} + e^{-2\kappa_M}) \\ &= 2e^{-2\delta f_M} (\cosh 4\kappa_M + \cosh 2g_M). \end{aligned} \quad (9.93)$$

Da segunda e da terceira obtemos

$$e^{\delta f_M} = 2 \cosh g_M e^{\kappa_{M-1}} \quad (9.94)$$

Eliminando δf escrevemos o mapa que leva da descrição em uma escala m para uma escala $m-1$, chamado de Grupo de Renormalização:

$$\kappa_{m-1} = \frac{1}{4} \log \left(\frac{\cosh 4\kappa_m + \cosh 2g_m}{4 \cosh^2 g_m} \right) \quad (9.95)$$

$$g_{m-1} = \frac{1}{4} \log \left(\frac{e^{2\kappa_m+2g_m} + e^{-2\kappa_m}}{e^{2\kappa_m-2g_m} + e^{-2\kappa_m}} \right) \quad (9.96)$$

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 9.7. 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

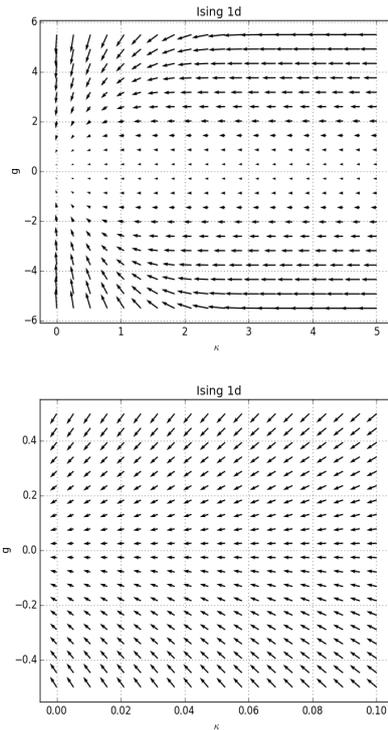


Figura 9.7: Ising 1d. Fluxo no espaço de Hamiltonianos. o lado direito é uma ampliação da figura à esquerda para valores de campo pequeno. A origem $(\kappa = 0, g = 0)$ é um ponto fixo do fluxo do grupo de renormalização.

usando a matriz de transferência, onde se mostrou que o sistema não tem magnetização espontânea. Agora rerotulamos 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:

$$Z(\kappa_m, g_m) = e^{2^{m-1} \delta f_m} Z(\kappa_{M-1}, g_{M-1}) \quad (9.97)$$

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

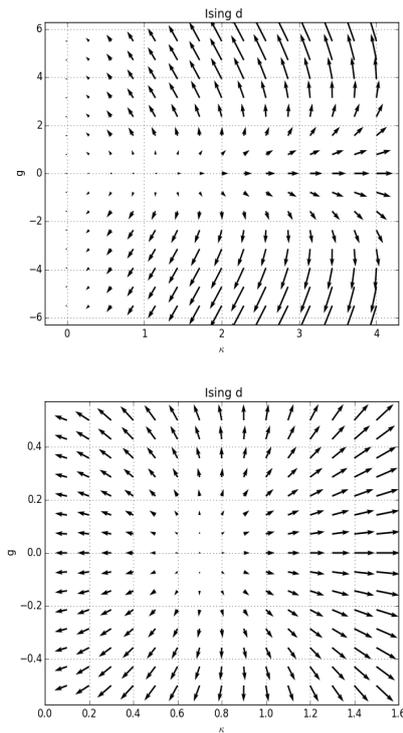


Figura 9.8: Ising 2d. Fluxo no espaço de Hamiltonianos. o lado direito é uma ampliação da figura à esquerda para valores de campo pequeno. A origem $(\kappa = 0, g = 0)$ e $(\kappa = \infty, g = 0)$ são pontos fixos estáveis do fluxo do grupo de renormalização. Mas há um novo ponto fixo $(\kappa = \kappa_c, g = 0)$ que é instável.

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 o equivalente das equações 9.98 e 9.96 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 9.8 e topologicamente está de acordo com cálculos mais trabalhosos. As equações são

$$\kappa_{m-1} = \frac{1}{4} \log \left(\frac{\cosh 8\kappa_m + \cosh 4g_m}{4 \cosh^2 2g_m} \right) \quad (9.98)$$

$$g_{m-1} = \frac{1}{4} \log \left(\frac{e^{4\kappa_m + 4g_m} + e^{-4\kappa_m}}{e^{4\kappa_m - 4g_m} + e^{-4\kappa_m}} \right) \quad (9.99)$$

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 9.8). 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 ζ . Uma vez que a descrição do sistema é feita numa resolução maior que ζ , 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 ζ . Quando a resolução so distingue regiões maiores que ζ , 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.

9.6 *Argumentos de Energia- Entropia*

Monte Carlo

A descrição das propriedades dos sistemas até aqui estudados foi reduzida ao cálculo de integrais em alta dimensão, valores esperados de funções

$$\langle f \rangle = \int f P d\mu.$$

Obviamente a medida $d\mu$ pode ser discreta, como é o caso quando tratamos variáveis discretas. É essencial encontrar meios numéricos de realizar estes cálculos, dado que o conjunto de modelos exatamente integráveis é bem menor que o de modelos interessantes. A classe de métodos de Monte Carlo é sem dúvida a mais importante de todas as ferramentas numéricas à nossa disposição. A arte de fazer Monte Carlos não será abordada aqui, simplesmente uma iniciação às idéias sem entrar em detalhes que se tornam necessários para seu uso profissional.

10.1 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

$$I_T = \frac{1}{N} \left(\frac{1}{2} f(x_1) + \sum_{i=2}^{N-1} f(x_i) + \frac{1}{2} f(x_N) \right), \quad (10.1)$$

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. O que 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.

10.2 Monte Carlo

10.2.1 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 $\langle x \rangle$ e $\langle x^2 \rangle$ existem e são finitos.¹ A variância σ_x^2 é definida por

$$\sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2$$

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

$$P(y) = \frac{1}{\sqrt{2\pi\sigma_y}} e^{-\frac{(y-\langle y \rangle)^2}{2\sigma_y^2}}$$

¹ Definimos os momentos $\langle x^n \rangle = \int x^n P(x) dx$

A aproximação é boa na região central da gaussiana e melhora quando N cresce. O valor médio de y e sua variancia são

$$\langle y \rangle = \langle x \rangle \text{ e } \sigma_y = \frac{\sigma_x}{\sqrt{N}}$$

Notem que se o objetivo for encontrar o valor esperado de x , que é $\langle x \rangle$, e não for possível realizar a integral, podemos estimar $\langle x \rangle$ a partir de y (isso pode ser generalizado para o cálculo de $\langle f \rangle = \int fP(x)dx$.) 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 amostras são independentes. 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 $\sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2$ 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 variancia de y .

Exercício: Distribuição de Cauchy Considere o problema acima, exceto que $\sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2$ 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.

10.2.2 Monte Carlo

A idéia básica é aproximar uma integral I por I_{MC}

$$I = \int_a^b f(x) dx \simeq I_{MC} = \frac{1}{N} \sum_{i=1}^N f(x_i) \tag{10.2}$$

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. (10.2) 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

$$y(x) = \int_0^x w(z)dz \tag{10.3}$$

$$y(0) = 0, y(1) = 1$$

então $dy = w(x)dx$ e a integral toma a forma

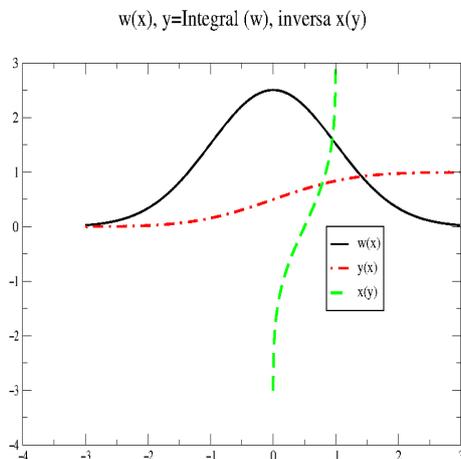
$$I = \int f(x(y))dy$$

e a aproximação Monte Carlo é

$$I = \int_a^b f(x)w(x)dx \simeq I_{MC} = \frac{1}{N} \sum_{i=1}^N f(x(y_i)) \tag{10.4}$$

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. [10.3]). 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 10.3, mas não é em geral possível fazê-lo de forma analítica.

² Uma 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



10.2.3 *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

$$P_Y(y)dy = P_X(x)dx$$

$$P_Y(y)dy = P_X(x) \left| \frac{\partial x}{\partial y} \right| dy \tag{10.5}$$

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

$$P_Y(y)dy = dy, \quad 0 \leq y_i < 1, i = 1...N$$

e zero fora.

10.2.4 *Distribuição Exponencial*

Suponha que queremos gerar amostras de uma distribuição exponencial. i.e $P_X(x) = \exp(-x)$. Integrando a eq. (10.5) obtemos

$$y(x) = \int_0^{y(x)} P_X(x) \frac{dx}{dy} dy$$

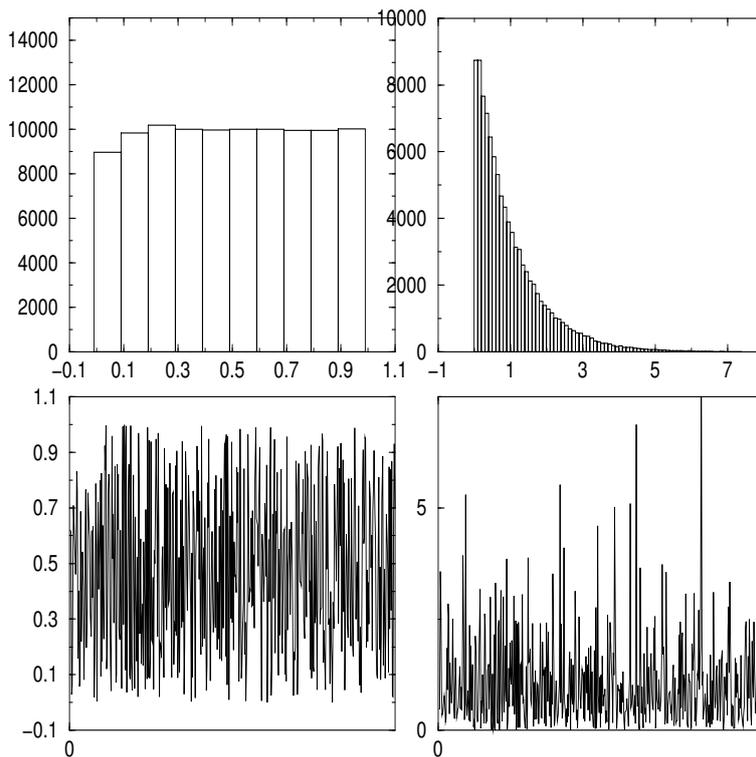
$$y(x) = \int_0^x P_X(x) dx = \int_0^x e^{-z} dz$$

$$y(x) = 1 - \exp(-x)$$

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

$$x = -\log(\text{RAND}(\text{SEED}))$$

Compare na figura a distribuição uniforme (esquerda) e a a exponencial (direita) (abaixo : série temporal, acima : histogramas)



10.2.5 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 (10.5). 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.(10.5) temos:

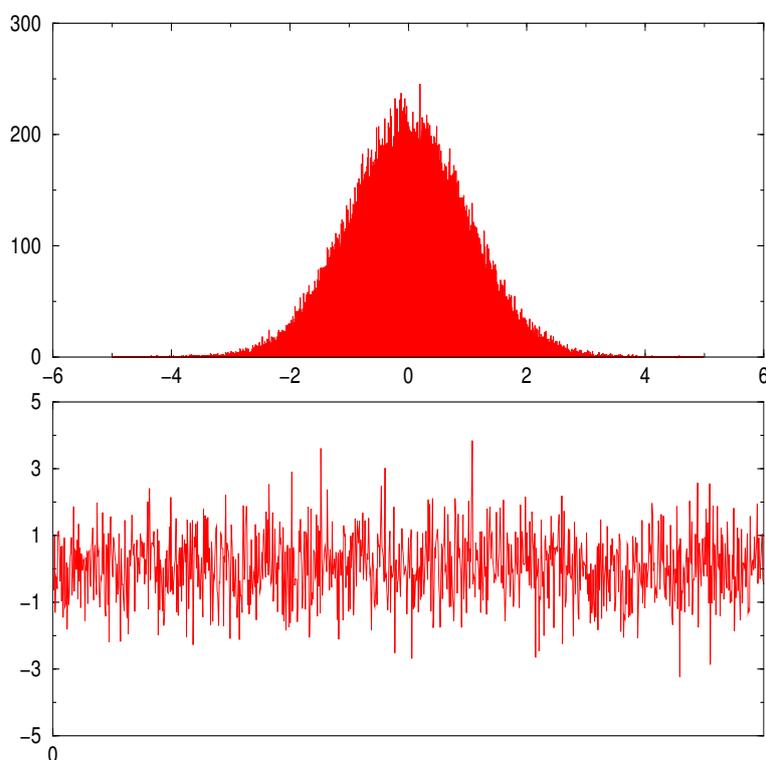
$$\int \int P_Y(y(x_1, x_2)) \left| \frac{\partial y}{\partial x} \right| dy_1 dy_2 = \int \int P_X(x) dx_1 dx_2$$

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.



10.2.6 Métodos Estáticos: rejeição

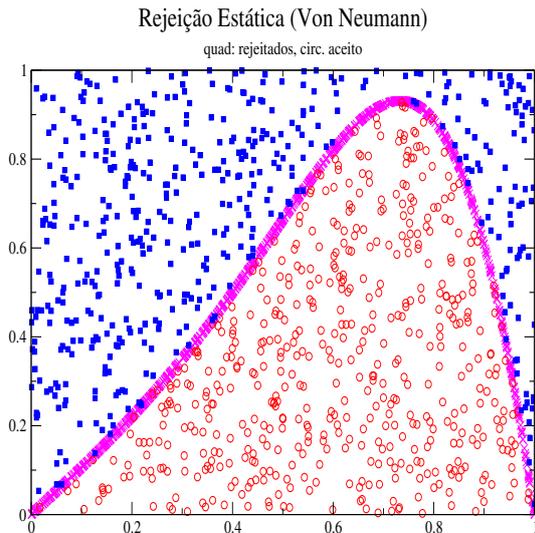
Raramente é possível realizar as integrais que permitem descobrir a transformação exata de variáveis e devemos então encontrar uma forma gerar diretamente os x com a distribuição $w(x)$. Os métodos que apresentaremos podem ser divididos em duas classes, estáticos e dinâmicos. Na primeira os números são gerados independentemente um dos outros³, enquanto que na segunda classe, construiremos um processo dinâmico que usará 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.

³ Tão independentemente quanto o gerador de números pseudo-aleatórios o permitir.



A sequência de números aceitos x são as abscissas dos círculos na figura acima.

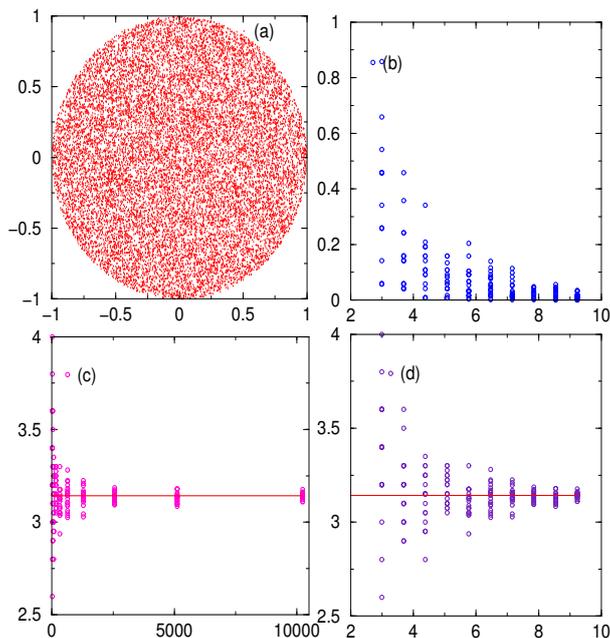
10.2.7 Círculo

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 :

- (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

Fig. 10.2.8: (a) pontos aceitos, (b) erro abs, (c) pi vs N, (d) pi vs logN



10.2.8 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 princípio- 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.

⁴ outra notação comum é $\Gamma(x_n \rightarrow x_{n+1})$

⁵ A cadeia de Markov é caracterizada pelas probabilidades de transição e pela distribuição inicial de probabilidades de x

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

$$w(x) = \int w(z)\Gamma(x|z)dz, \quad (10.6)$$

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

$$\Delta P(x|t) = P(x|t + 1) - P(x|t) = \int P(z|t)\Gamma(x|z)dz - P(x|t) \int \Gamma(z|x)dz,$$

$$\Delta P(x|t) = P(x|t + 1) - P(x|t) = \int [P(z|t)\Gamma(x|z) - P(x|t)\Gamma(z|x)] dz \quad (10.7)$$

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. [10.6, 10.7]

$$\Delta w(x) = \int [w(z)\Gamma(x|z) - w(x)\Gamma(z|x)] dz = 0.$$

Há várias escolhas possíveis de Γ para satisfazer esta relação. A escolha mais simples sugere impor uma condição

$$w(z)\Gamma(x|z) = w(x)\Gamma(z|x) \quad (10.8)$$

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. [10.8] $w(x)$ 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 [10.8]. Embora todas levem a algoritmos corretos, no sentido que

$$I = \int_a^b f(x)w(x)dx \simeq I_{MC} = \frac{1}{N} \sum_{i=1}^N f(x_i) \quad (10.9)$$

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.

10.2.9 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

$$\frac{F(a)}{F(1/a)} = a, \text{ para todo } a \quad (10.10)$$

Para escolher $T(\cdot|\cdot)$ é útil definir uma distância entre configurações. Se estivermos falando de graus de liberdade no espaço euclidiano, a soma das distâncias entre uma partícula nas configurações x e z é uma boa escolha. Para um sistema de Ising podemos medir a distância pelo número de spins diferentes entre duas configurações. Definimos uma região $\mathcal{B}(x|z)$, uma bola, e a função indicadora $\chi_{\mathcal{B}}$, que toma valores 1 dentro da bola e zero fora. A escolha mais comum, para a probabilidade de tentativa de mudança é tomar

$$T(z|x) = \text{constante dentro } \mathcal{B}(x|z) = \frac{1}{|\mathcal{B}(x|z)|}$$

e zero fora da de \mathcal{B} . 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

$$F(a) = \min(1, a).$$

Para verificar que satisfaz 10.10, note que há dois casos:

- $a \geq 1$ segue que $\frac{\min(1,a)}{\min(1,a^{-1})} = \frac{1}{a^{-1}} = a$
- $a < 1$ segue que $\frac{\min(1,a)}{\min(1,a^{-1})} = a$

o que leva ao seguinte

Algoritmo de Metropolis:

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 $w(x_n)$. Defina $r = \frac{w(x_T)}{w(x_n)}$ (para não ficar recalculando $w(x)$, que pode ser muito caro.)
 - Se $r \geq 1$ (i.e. $w(x_T) \geq w(x_n)$) então **aceita** : $x_{n+1} = x_T$
 - Se $r < 1$ (i.e. $w(x_T) < w(x_n)$) então escolhe um número aleatório uniforme $0 \leq \xi < 1$ e
 - aceita** : $x_{n+1} = x_T$ se $r \geq \xi$ (i.e. $w(x_T) \geq w(x_n)\xi$)
 - rejeita** : $x_{n+1} = x_n$ se $r < \xi$ (i.e. $w(x_T) \leq w(x_n)\xi$)
4. Guarda informação sobre x_n .
5. Se critério de parada não for satisfeito, volta a 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 de 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 médio $\langle E(x) \rangle$ 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 $\exp(-\beta(E(x_T) - E(x_n))) < \xi$.
4. Guarda informação sobre x_n .
5. Se um critério de parada não for satisfeito, volta a 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 automaticamente rejeitada. Se o aumento de energia for muito grande é grande a probabilidade que seja rejeitada, mas se não for, é grande a de ser aceita. A escala de grande ou pequeno é determinada pela razão dos fatores de Boltzmann de cada configuração.

10.2.10 Atrator da dinâmica

Daremos argumentos em defesa da posição que a distribuição $w(x)$ é um ponto fixo atrativo da equação 10.7 para o algoritmo de Metropolis.

Defina os conjuntos

$$B_x(z) = \{z|T(z|x) > 0\}C_+(x) = \{x|P(x) > w(x)\}, \quad (10.11)$$

$$C_0(x) = \{x|P(x) = w(x)\}, \quad (10.12)$$

$$C_-(x) = \{x|P(x) < w(x)\}. \quad (10.13)$$

Defina $\delta P(x) = P(x|t) - w(x)$, a diferença entre a distribuição em um dado instante t e a de equilíbrio. É claro que dado que as distribuições estão normalizadas, teremos

$$\begin{aligned} 0 &= \int \delta P(x) dx \\ &= \int_{C_+(x)} \delta P(x) dx + \int_{C_-(x)} \delta P(x) dx \end{aligned}$$

onde o primeiro termo contém as contribuições positivas e o segundo as negativas. A equação da dinâmica

$$\begin{aligned} \Delta P(x|t) &= \int_{B_x(z)} [P(z|t)\Gamma(x|z) - P(x|t)\Gamma(z|x)] dz \\ &= \int_{B_x(z)} [(w(z) + \delta P(z))\Gamma(x|z) - (w(x) + \delta P(x))\Gamma(z|x)] dz \\ &= \int_{B_x(z)} [\delta P(z)\Gamma(x|z) - \delta P(x)\Gamma(z|x)] dz. \end{aligned} \quad (10.14)$$

Agora integramos sobre o conjunto de configurações $C_+(x)$

$$\Delta P_+ = \int_{C_+(x)} \Delta P(x|t) dx = \int_{B_x(z), C_+(x)} [\delta P(z)\Gamma(x|z) - \delta P(x)\Gamma(z|x)] dz dx,$$

e separamos as configurações z em $C_{\pm}(z)$

$$\begin{aligned} \Delta P_+ &= \int_{C_+(x), C_+(z)} [\delta P(z)\Gamma(x|z) - \delta P(x)\Gamma(z|x)] dz dx \\ &+ \int_{C_+(x), C_-(z)} [\delta P(z)\Gamma(x|z) - \delta P(x)\Gamma(z|x)] dz dx. \end{aligned}$$

A integral sobre $C_+(x), C_+(z)$ é nula devido a que é simétrica e antisimétrica ante trocas $z \leftrightarrow x$. Portanto

$$\begin{aligned} \Delta P_+ &= \int_{C_+(x), C_-(z)} [\delta P(z)\Gamma(x|z) - \delta P(x)\Gamma(z|x)] dz dx. \\ &\leq 0, \end{aligned} \quad (10.15)$$

pois o termo com sinal positivo tem $\delta P(z)$ que é sempre negativo em $C_-(z)$, e o termo com sinal negativo tem $\delta P(x)$ que é sempre positivo em $C_+(x)$.

Analogamente, integrando a equação 10.14 sobre o conjunto de configurações $C_-(x)$, vemos que

$$\Delta P_- = \int_{C_-(x)} \Delta P(x|t) dx \quad (10.16)$$

$$= \int_{C_-(x), C_+(z)} [\delta P(z)\Gamma(x|z) - \delta P(x)\Gamma(z|x)] dz dx.$$

$$\geq 0. \quad (10.17)$$

Estes resultados sugerem que a diferença entre $P(x|t)$ e $w(x)$ diminui ao iterar a dinâmica. A dinâmica é tal que nas regiões em que $P(x|t)$ é maior do que deveria ser, diminui. Onde é menor aumenta.

10.2.11 Modelo de Ising

Novamente visitaremos o laboratório de Ising. Há vários livros que tratam deste problema de forma mais completa. Aqui apresentaremos somente alguns detalhes que permitirão que o leitor comece seu simulador de Monte Carlo em um problema que tem propriedades críticas interessantes.

Que um algoritmo funcione bem para o modelo de Ising em duas dimensões não é garantia que servirá para outros modelos. Do ponto de vista de um programa legível, convém escrevê-lo em módulos. Não buscamos velocidade mas sim facilidade de leitura. Os módulos necessários são descritos a seguir:

Pseudo código

⁶ A configuração do sistema é guardada numa matriz sigma de dimensões $L \times L$. Isto é $\text{sigma}(I_x, I_y) = \pm 1$.

A função `Controle(param)` controla que função é executada e com que parâmetros: `param=(L, Nterm, NMC, deltaNMC, interT)`

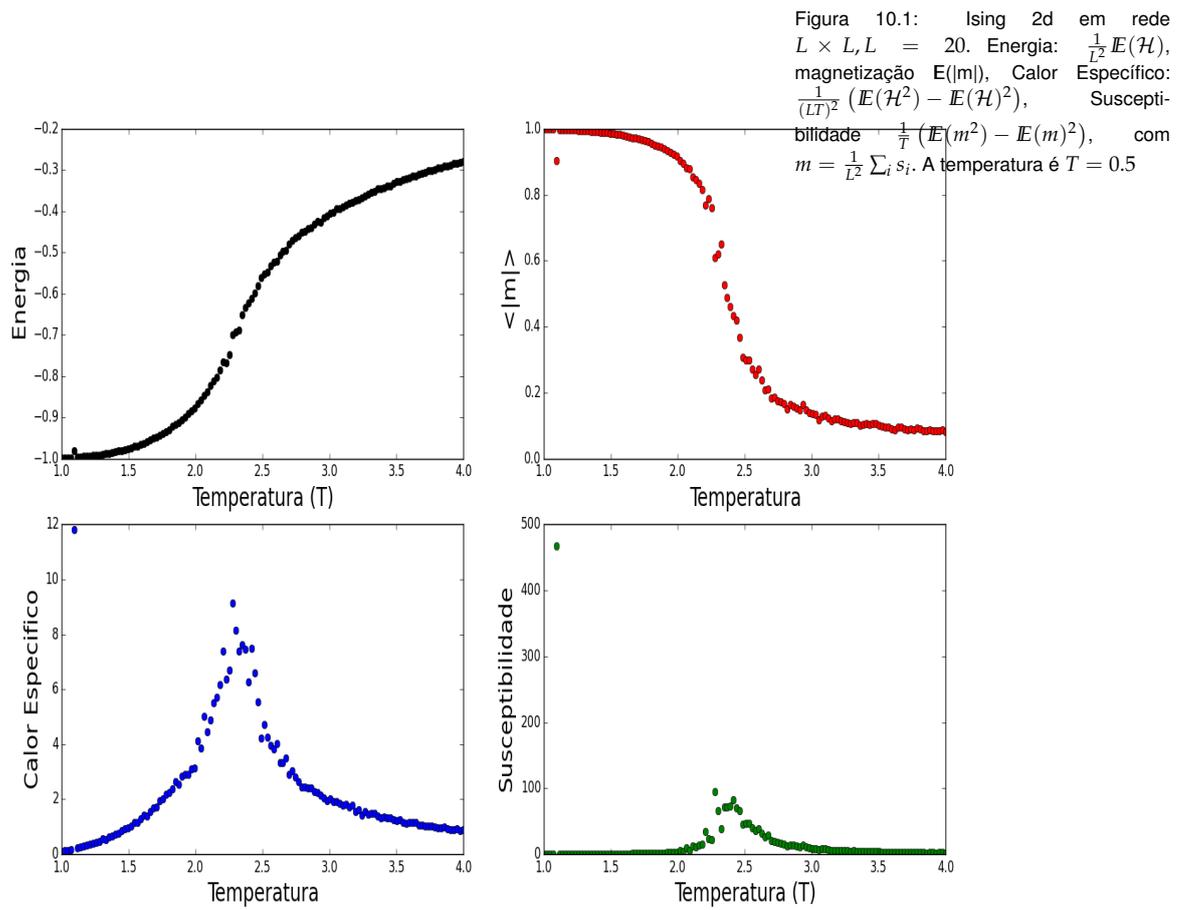
A rede tem tamanho L^2 , `Nterm`, `NMC*deltaNMC` são respectivamente o número de configurações geradas para permitir termalização e o número de configurações que serão geradas na fase de medidas. As medidas são feitas a cada `deltaNMC` configurações. Portanto o número de configurações que são medidas é `NMC`.

`interT` determina o conjunto de temperaturas em que o sistema será simulado.

A função `Novaconf(sigma, param)` recebe uma configuração e devolve outra. Este é o coração do algoritmo. Abaixo mostramos como exemplo uma implementação do algoritmo de Metropolis.

A função `Acumula(sigma, f1, f2, ... f1)` extrai informação de `config` e a acumula nos diversos `fs`. O objetivo da simulação é estimar quantidades de interesse termodinâmico através dos diferentes `fs`, por exemplo a energia, a magnetização, o calor

⁶ Pode pegar o programa em <http://rajeshrinet.github.io/blog/2014/ising-model/>



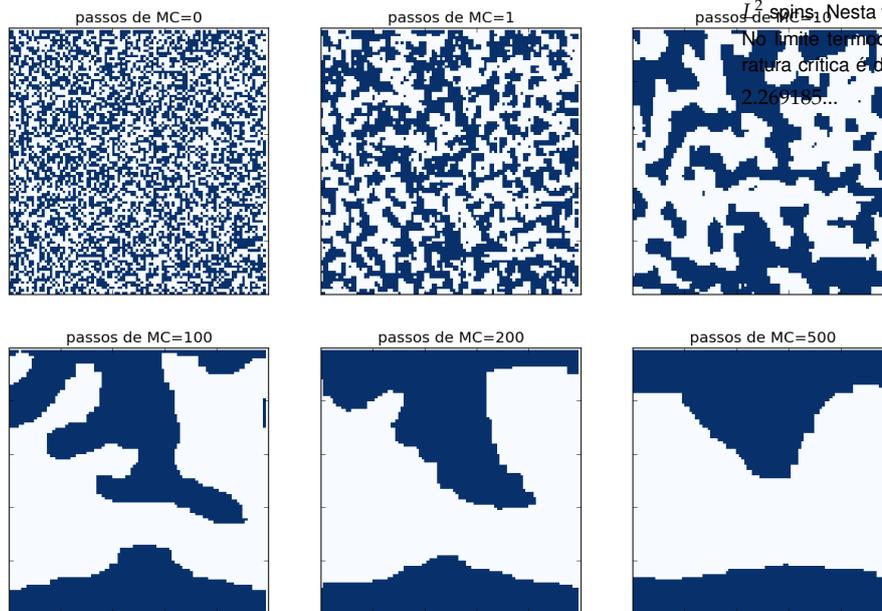


Figura 10.2: A matriz sigma em diferentes tempos. Unidade de tempo é 1 passo MC, que é dado pela tentativa de mudança de L^2 spins. Nesta figura $L = 100$ e $T = 0.50$. No limite termodinâmico, $L \rightarrow \infty$, temperatura crítica é dada por $T = \frac{2}{\log(1+\sqrt{2})} \approx 2.269185\dots$

Figura 10.3: O mesmo que na figura anterior, temperatura é $T = 2.26 \approx T_c$

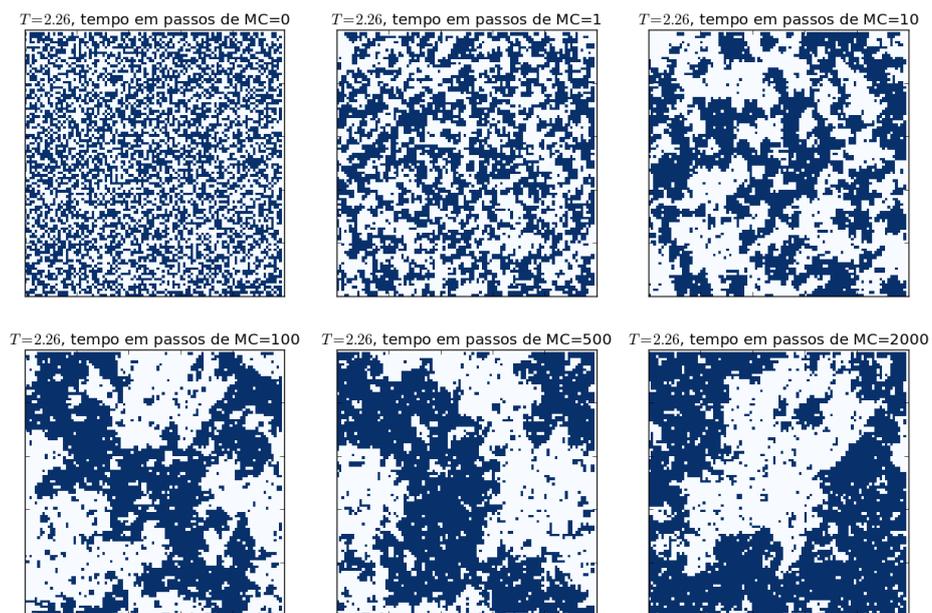
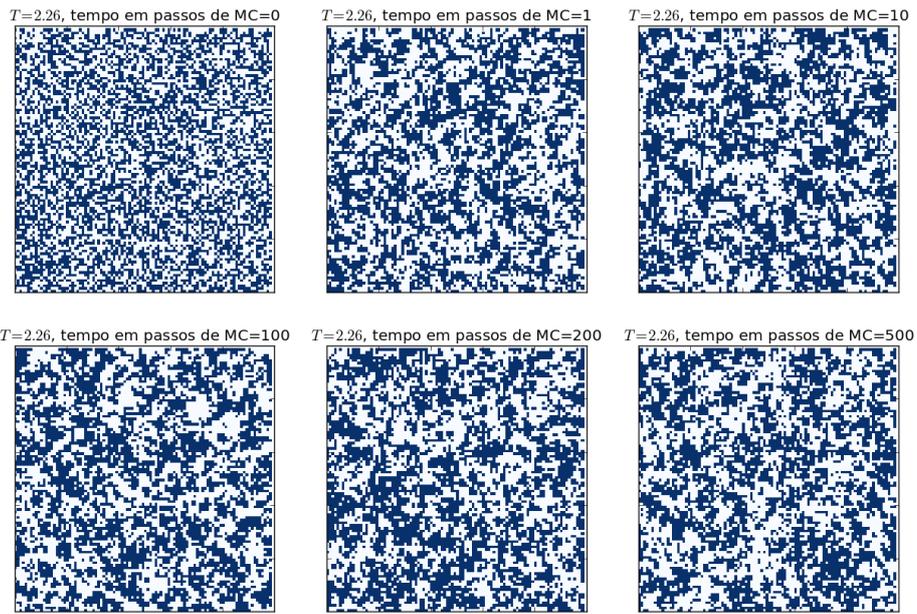


Figura 10.4: O mesmo que na figura anterior à temperatura é $T = 3.5$



específico, a susceptibilidade magnética e correlações, espaciais e temporais.

A função `GravaResultados` faz pequenas operações, como dividir pelo número de configurações, gravar os resultados. A função `Grafico` produz os gráficos para acabar de forma satisfatória. Fazer os gráficos é opcional, podem ser feitos em outro ambiente.

- `def Controle(param):`
 - para todo $1 \leq i, j \leq L$: $\sigma(i, j) = \pm 1$ com probabilidade meio
 - chama `Novaconf(sigma, param)` `Nterm` vezes para termalizar;
 - faz `NMC` vezes # coração do programa
 - * chama `Novaconf(sigma, param)` `deltaNMC` vezes
 - * chama `Acumula(sigma, f1, f2, ... fl)`
 - chama `GravaResultados`.

A rotina que realiza a mudança de configuração

- `def Novaconf(sigma, param):`
 - para todo $1 \leq i, j \leq L$:
 - * escolhe duas coordenadas k, l com probabilidade uniforme, e soma seus quatro vizinho $h_{kl} = \sum_{\text{vizinhos}} \sigma$ ($v(k), v(l)$). A definição de o que é um vizinho na borda da rede impõe a escolha de condições de contorno, por exemplo $v(k)$ toma valores $(k \pm 1) \bmod L$, para condições periódicas de contorno.
 - * A mudança de energia ao tentar inverter o spin $\sigma(k, l)$ é $\Delta E = -2h_{kl} \sigma(k, l)$
 - * Se $\Delta E \leq 0$,
 - a nova configuração terá $\sigma(k, l) = -\sigma(k, l)$
 - * else escolhe `aleat` aleatório uniforme entre 0 e 1.
 - se $\exp(-\Delta E/T) \geq \text{aleat}$ então $\sigma(k, l) = -\sigma(k, l)$

O programa é de forma esquemática

- `param = L, Nterm, NMC, deltaNMC, interT`; inicializa parâmetros.
- para `t` no conjunto de temperaturas `interT`
- `Controle(param)`
- `Grafico`