

Teoria: Sistemas de Processamento de Informação

Nestor Caticha

28 de outubro de 2015

Capítulo 1

Entropia

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

Bertrand Russell, in *Sceptical Essays*

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

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

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

1.1 Simetria

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

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

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

$$\begin{aligned} P(\sigma' = 1|IJ_2) &= P(\sigma = -1|IJ_1) \\ P(\sigma' = -1|IJ_2) &= P(\sigma = 1|IJ_1). \end{aligned} \quad (1.1)$$

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

Suponha agora que a informação que temos I_2 , que difere de I por muito pouco, nos leve a ser **indiferentes** sobre que jogo está sendo jogado. Suponha que não faça diferença dado I_2 se é $\sigma = 1$ que está associada à cara ou se é $\sigma' = 1$ que o está. Isto é o “+1” esta pintado numa dos lados da moeda e “-1” está pintado no outro, mas não sabemos qual. É claro que as equações 1.1 ainda valem com I_2 no lugar I . Mas agora há uma simetria entre os jogos J_1 e J_2 , que nos leva a concluir por consistência

$$\begin{aligned} P(\sigma' = 1|I_2J_2) &= P(\sigma = 1|I_2J_1) \\ P(\sigma' = -1|I_2J_2) &= P(\sigma = -1|I_2J_1). \end{aligned} \quad (1.2)$$

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

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

Note a sutil diferença entre as equações 1.1 (escrita com I_2 em vez de I) e 1.2. Dado que $P(\sigma = 1|IJ_1) + P(\sigma = -1|IJ_1) = 1$, devemos concluir que $P(\sigma = 1|IJ_1) = 1/2$ e $P(\sigma = -1|IJ_1) = 1/2$.

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

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

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

1.2 Incerteza, ignorância e Informação

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

A última imposição, que chamaremos de

- (5) aditividade sob agrupamento,

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

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

$$H[p_i] = H_G[P_g] + \sum_{g=1}^N P_g H_g[p(i|g)] \quad (1.3)$$

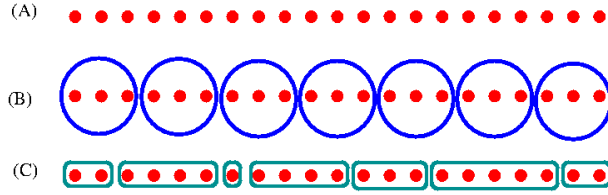


Figura 1.1: (A) Cada ponto (vermelho) representa um possível estado (mutuamente exclusivos) de uma variável. (B) Agrupamos em grupos com o mesmo tamanho. (C) Os agrupamentos são arbitrários.

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

(I) Começamos analisando o caso particular (figura 1.1.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 $p(i|g) = N/n = 1/m$

$$H[1/n \dots 1/n] = H_G[1/N, \dots, 1/N] + \sum_{g=1}^N \frac{1}{N} H_g[1/m, \dots, 1/m] \quad (1.4)$$

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

$$F(n) = F(N) + \sum_{g=1}^N \frac{1}{N} F(m) \quad (1.5)$$

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

$$F(mN) = F(N) + F(m) \quad (1.6)$$

A solução óbvia é dada por

$$F(n) = k \log n. \quad (1.7)$$

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

(II) Ainda não obtivemos a forma de H no caso geral. Passamos a analisar o caso em que os tamanhos dos agrupamentos m_g são arbitrários (figura 1.1.C), salvo o fato $\sum_{g=1}^N m_g = n$, mas ainda temos que $p_i = 1/n$ é uniforme.

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

$$H[1/n, \dots, 1/n] = H_G[P_g] + \sum_{g=1}^N P_g H_g[1/m_g, \dots, 1/m_g] \quad (1.8)$$

e substituímos a entropia da distribuição uniforme:

$$F(n) = H_G[P_g] + \sum_{g=1}^N P_g F(p(i|g)), \quad (1.9)$$

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

$$H_G[P_g] = F(n) - \sum_{g=1}^N P_g F(m_g). \quad (1.10)$$

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

$$H_G[P_g] = \left(\sum_{g=1}^N P_g \right) k \log n - k \sum_{g=1}^N P_g \log(m_g) \quad (1.11)$$

$$H_G[P_g] = -k \sum_{g=1}^N P_g \log(m_g/n) \quad (1.12)$$

e finalmente

$$H_G[P_g] = -k \sum_{g=1}^N P_g \log(P_g) \quad (1.13)$$

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

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

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

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

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

$$H_G[P|p_0] = -k \int dx P(x) \log\left(\frac{P(x)}{p_0(x)}\right) \quad (1.14)$$

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

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

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

⁵Não cabe no curso introdutório de Mecânica Estatística

⁶ver A.C

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

1.3 Exemplos

1.3.1 O problema variacional: variáveis discretas

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

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

$$\sum P_i = 1 \quad (1.15)$$

$$\sum x_i P_i = \mu \quad (1.16)$$

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

$$H[P] - \lambda_1 \left\{ \sum x_i P_i - \mu \right\} - \lambda_0 \left\{ \sum P_i - 1 \right\} \quad (1.17)$$

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

$$\delta \left[-\sum P_i \log P_i - \lambda_1 \left\{ \sum x_i P_i - \mu \right\} - \lambda_0 \left\{ \sum P_i - 1 \right\} \right] = 0 \quad (1.18)$$

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

$$(1) \quad \delta H[P] = H[P + \delta P] - H[P] = -\sum_i \delta P_i [\log P_i + 1] \quad (1.19)$$

$$(2) \quad \delta(-\lambda_1 \left\{ \sum_i x_i P_i - \mu \right\}) = -\lambda_1 \sum_i x_i \delta P_i \quad (1.20)$$

$$(3) \quad \delta(-\lambda_0 \left\{ \sum_i P_i - 1 \right\}) = -\lambda_0 \sum_i \delta P_i \quad (1.21)$$

Juntando os três termos:

$$-\sum_i \delta P_i \{\log P_i + 1 + \lambda_1 x_i + \lambda_0\} = 0 \quad (1.22)$$

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

$$\log P_i + 1 + \lambda_1 x_i + \lambda_0 = 0 \quad (1.23)$$

e segue que

$$P_i = e^{-1-\lambda_0-\lambda_1 x_i} \quad (1.24)$$

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

$$Z(\lambda_1) = \sum_i e^{-\lambda_1 x_i} \quad (1.25)$$

A seguir notamos uma utilidade de Z , chamada *função de partição*:

$$\frac{d \log Z}{d \lambda_1} = -\frac{1}{Z} \sum_i x_i e^{-\lambda_1 x_i} \quad (1.26)$$

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

$$\langle x \rangle = -\frac{d \log Z}{d \lambda_1} \quad (1.27)$$

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

1.3.2 Exemplo: Variável de dois estados

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

$$\delta \left[-\sum P_i \log P_i - \lambda_0 \left\{ \sum P_i - 1 \right\} \right] = 0 \quad (1.28)$$

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

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

$$H(p) = -p \log p - (1 - p) \log(1 - p) \quad (1.29)$$

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

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

1.3.3 3 estados

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

$$P_i = \frac{e^{-\lambda_1 x_i}}{Z(\lambda_1)} \quad (1.30)$$

com λ_1 determinado pela eq. 1.27 e Z pela eq. 1.25 . Portanto, chamando $u = \exp(-\lambda_1)$,

$$Z = 1 + u + u^2 \quad (1.31)$$

e

$$P_i = \frac{u^i}{1 + u + u^2} \quad (1.32)$$

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

$$u^2(m - 2) + u(m - 1) + m = 0 \quad (1.33)$$

$$u = \frac{1 - m \pm \sqrt{6m - 3m^2 + 1}}{2(m - 2)} \quad (1.34)$$

A equação 1.33 tem duas raízes. Discuta a escolha da solução correta. A figura 1.2 mostra o valor máximo da entropia H como função de m , e do lado esquerdo as probabilidades. Note que para o gráfico de $H[P]$ deveria ser desenhado num espaço tridimensional (bidimensional se o vínculo da normalização for incluído). Mas desenhamos o valor do máximo de $H[P]$ que é função de um único parâmetro.

1.3.4 Variáveis reais: a gaussiana

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

$$\begin{aligned} \int P(x) dx &= 1 \\ \int x P(x) dx &= \mu \\ \int x^2 P(x) dx &= \sigma^2 + \mu^2 \end{aligned} \quad (1.35)$$

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

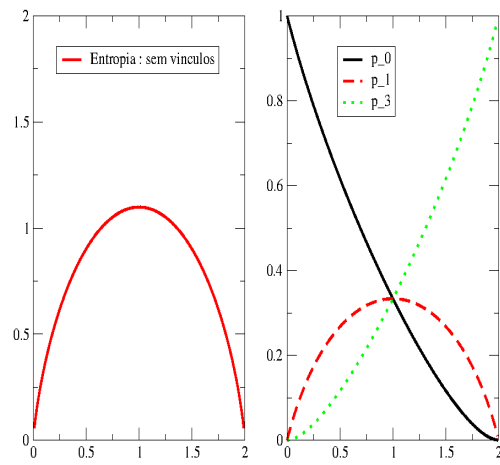


Figura 1.2: Esquerda: Máximo da entropia como função da média m . Direita: as probabilidades p_0 , p_1 e p_2 que maximizam a entropia de Shannon como função da informação $\langle x \rangle = m$.

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

$$\delta \left[- \int dx P(x) \log \left(\frac{P(x)}{p_0(x)} \right) + \lambda_0 \left\{ \int P(x) dx - 1 \right\} + \right. \quad (1.36)$$

$$\left. \lambda_1 \left\{ \int x P(x) dx - \mu \right\} + \lambda_2 \left\{ \int x^2 P(x) dx - \sigma^2 - \mu^2 \right\} \right] = 0 \quad (1.37)$$

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

$$\{1 + \log P(x) + \lambda_0 + \lambda_1 x + \lambda_2 x^2\} \delta P(x) = 0, \quad (1.38)$$

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

$$P(x) = e^{1 - \lambda_0 - \lambda_1 x - \lambda_2 x^2} \quad (1.39)$$

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

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1.40)$$

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

Exemplo: Bayes e Máxima Entropia

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

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

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

$$y_i = f_{\theta}(x_i) + \eta_i \quad (1.41)$$

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

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

$$P(\theta|(x_i, y_i), I) = \frac{p_0(\theta)P(\{y_i\}|\theta, \{x_i\}, I)}{P(D)}, \quad (1.42)$$

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

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

Se não houvesse ruído η :

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

Devido à equação 1.41 podemos ver que

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

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

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

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

1.4 Entropia e convexidade

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

$$K[p|q] = \sum_i p_i \log \frac{p_i}{q_i}. \quad (1.43)$$

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

$$K[p|q] = \sum_i p_i \log(np_i), \quad (1.44)$$

$$K[p|q] = \log n - H[p], \quad (1.45)$$

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

$$\log x \leq x - 1, \quad (1.46)$$

e que pode ser facilmente visto do gráfico das duas funções. Mas ver o gráfico não é suficiente, prove este resultado. A igualdade só ocorre se $x = 1$. A equação 1.46 usada na forma de $H[p]$ leva a

$$\begin{aligned} H[p] &= - \sum_i p_i \log(p_i), \\ &\geq - \sum_i p_i (p_i - 1) \\ &\geq 1 - \sum_i p_i^2 \geq 0, \end{aligned} \tag{1.47}$$

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

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

$$\begin{aligned} K[p|q] &= - \sum_i p_i \log \frac{1}{np_i}, \\ &\geq - \sum_i p_i \left(\frac{1}{np_i} - 1 \right) \\ &\geq \sum_i \frac{1}{n} - \sum_i p_i = 0, \end{aligned} \tag{1.48}$$

com a igualdade ocorrendo somente se $p_i = q_i$. Juntando as duas cotas com a relação entre K e H , equação 1.45, temos que

$$0 \leq H[p] \leq \log n, \tag{1.49}$$

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

1.4.1 Independência

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

$$P(X, Y|I) = P(X|YI)P(Y|I). \tag{1.50}$$

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

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

$$\begin{aligned} K[P(X, Y|I)|P(X|I)P(Y|I)] &= \sum_i P(X, Y|I) \log \frac{P(X, Y|I)}{P(X|I)P(Y|I)} \\ &= -H[P(X, Y|I)] + \sum_i P(X, Y|I) \log P(X|I) + \sum_i P(X, Y|I) \log P(Y|I) \\ &= -H[P(X, Y|I)] + H[P(X|I)] + H[P(Y|I)] \end{aligned} \quad (1.51)$$

e como $K \geq 0$ obtemos que

$$H[P(X, Y|I)] \leq H[P(X|I)] + H[P(Y|I)]. \quad (1.52)$$

A equação 1.51 mostra que $K[P(X, Y|I)|P(X|I)P(Y|I)]$ mede a diferença entre a informação que falta se só olharmos para X e Y separadamente ou conjuntamente. A equação 1.52, que mostra a propriedade chamada de subaditividade, mostra que a informação que falta quando as variáveis são analisadas conjuntamente é menor ou igual que quando o são em separado. A não ser que as variáveis sejam independentes a análise da distribuição conjunta incorpora mais informação ou seja, falta menos informação.

1.4.2 Exercício sobre independência: Variáveis Gaussianas correlacionadas

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

$$P(x, y|\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)}. \quad (1.53)$$

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

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

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

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

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

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

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

Suponha que queremos usar σ_x como estimador de σ_y . Calcule como função de ρ a probabilidade e_σ de fazer um *erro*, i.e. de errar o sinal de Y quando o sinal de X é usado no seu lugar ou que os sinais de X e Y sejam diferentes. A resposta $e_\sigma = \frac{1}{\pi} \arccos \rho$. Fica fácil de entender se fizermos um desenho. Trace duas linhas ξ_x e ξ_y que fazem um ângulo θ . Trace dois vetores perpendiculares 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 1.53.

O discutido neste exemplo é extremamente útil em problemas de aprendizagem de máquina, onde queremos aprender uma direção desconhecida $\vec{J}_y$ só com informação sobre σ_y . O que se faz é construir uma aproximação $\vec{J}_x$ a partir de σ_y para vários exemplos $\vec{S}$. Isto é, a máquina aprende e a medida da qualidade da aprendizagem, $P(\sigma_x|\sigma_y)$ é chamado de erro de generalização, que cai quanto maior for a correlação ρ . Sobre isto, muito mais será dito em capítulos posteriores⁸.

1.5 Caso Geral para informação sobre valores esperados

Em geral podemos incluir vários vínculos

$$\langle g_k(x) \rangle = G_k, \quad (1.54)$$

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

$$\delta \left[- \sum P_i \log P_i - \sum_{k=0}^K \lambda_k \left\{ \sum_i g_k(x_i) P_i - \mu \right\} \right] = 0 \quad (1.55)$$

que leva a

$$- \sum_i \delta P_i \{ \log P_i + 1 + \sum_{k=0}^K \lambda_k g_k(x_i) \} = 0 \quad (1.56)$$

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

$$P_i = \frac{e^{-\sum_{k=1}^K \lambda_k g_k(x_i)}}{Z} \quad (1.57)$$

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

⁸Mas não neste curso introdutório.

$$Z = \sum_i e^{-\sum_{k=1}^K \lambda_k g_k(x_i)} \quad (1.58)$$

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

$$\langle g_k \rangle = G_k = - \left(\frac{\partial \log Z}{\partial \lambda_k} \right)_{\lambda_i, i \neq k} \quad (1.59)$$

exercício Mostre o resultado acima.

As equações acima formam um conjunto de equações não lineares que determinam os valores de $\{\lambda_k\}$ se os G_k forem conhecidos. No entanto nos dá ainda mais. Suponha que os valores de G_k não sejam conhecidos, mas que só tenhamos o conhecimento que se os soubessemos teríamos determinado uma distribuição adequada para algum fim. Isso significa que se, em vez de G_k tivessemos 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.

1.6 Conexão com a Termodinâmica: estrutura matemática

1.6.1 Transformações de Legendre

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

$$S(G_1, G_2, \dots, G_K) = \log Z(\lambda_1, \lambda_2, \dots, \lambda_K) + \sum_{k=1}^K \lambda_k G_k \quad (1.60)$$

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

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

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

$$\frac{\partial S(G_1, G_2, \dots, G_K)}{\partial G_i} = \sum_{j=1}^K \frac{\partial \log Z}{\partial \lambda_j} \frac{\partial \lambda_j}{\partial G_i} + \sum_{j=1}^K \frac{\partial \lambda_k}{\partial G_i} G_i + \lambda_i \quad (1.61)$$

Usando a equação 1.59, obtemos finalmente que

$$\lambda_i = \frac{\partial S(G_1, G_2, \dots, G_K)}{\partial G_i} \quad (1.62)$$

Um caso particular desta relação é $\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_{V,N}$ que veremos na seção 1.9.1.

Os resultados 1.60 e 1.62 mostram como podemos descrever o sistema em termos dos G_k s ou dos λ_k s, ou seja $S(\{G_k\})$ e $\log Z(\{\lambda_k\})$ são obtidos um do outro através de uma transformada de Legendre.

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

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

1.6.2 Relações de Maxwell

Podemos continuar estudando as derivadas parciais superiores e obter a estrutura das relações de Maxwell. Olhemos para as segundas derivadas, a partir das derivadas da equação 1.62:

$$\frac{\partial \lambda_i}{\partial G_j} = \frac{\partial^2 S(G_1, G_2, \dots, G_K)}{\partial G_i \partial G_j} = \frac{\partial^2 S(G_1, G_2, \dots, G_K)}{\partial G_j \partial G_i} = \frac{\partial \lambda_j}{\partial G_i} \quad (1.63)$$

e da equação 1.59

$$\frac{\partial G_k}{\partial \lambda_i} = -\frac{\partial^2 \log Z}{\partial \lambda_k \partial \lambda_i} = \frac{\partial G_i}{\partial \lambda_k} \quad (1.64)$$

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

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

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

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

1.7 Conexão com a Termodinâmica: Micro e Macroestados

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

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

⁹Na época não era necessário o adjetivo clássico.

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

$$\{qp\}_{mol}\{qp\}_{atom}\{qp\}_{elet}\{qp\}_{nucl}\{qp\}_{prot}\{qp\}_{neutr}\{qp\}_{quark}\{qp\}_{gluon}\{qp\}? \quad (1.65)$$

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

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

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

¹⁰A.C.

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

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

$$P(\sigma_1 = +1, \dots, \sigma_N = -1, \theta_1, \dots, \theta_N) \prod_i d\theta_i = p \dots q P(\theta_1) \dots P(\theta_N) \prod_i d\theta_i \quad (1.66)$$

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

$$\begin{aligned} P(m|N\{p, q, P(\theta)\}) &= \sum_{\sigma_1, \dots, \sigma_N} \int \prod_i d\theta_i P(\sigma_1, \dots, \sigma_N, \theta_1, \dots, \theta_N) \delta_{m, (N + \sum_i \sigma_i)/2} \\ &= \frac{N!}{m!(N-m)!} p^m q^{N-m} \end{aligned} \quad (1.67)$$

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

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

$$P(\{\sigma_i\}_{i=1, N}|I) = \int \prod_i d\theta_i P(\{(\theta_i, \sigma_i)\}_{i=1, N}|I) \quad (1.68)$$

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

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

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

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

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

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

1.8 Conexão com a Termodinâmica: Física

1.8.1 O Ensemble Microcanônico

A importância do conceito de energia é devida à sua conservação. O célebre teorema de Noether mostra a relação entre simetrias e leis de conservação. Se sob certas condições o sistema tem uma simetria contínua, concluímos que numa experiência realizada nessas condições haverá uma quantidade conservada. Suponha um sistema isolado de tal maneira que as propriedades são invariantes antes translações temporais, a energia será conservada. A preparação do arranjo experimental é feita com o objetivo de responder certas perguntas. Isso define o macroestado. Pode ser que não saibamos como fixá-lo, mas essencialmente esse é o objetivo de nosso estudo, encontrar quais as condições experimentais, ou quais os vínculos impostos ao sistema, para que o 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

$$H[P] = -k \int_{E \leq \mathcal{H}(q_i, p_i) \leq E + \delta E, V} \prod_i dq_i dp_i P(q_i, p_i) \log P(q_i, p_i) \quad (1.69)$$

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

$$P(q_i, p_i) = c \quad (1.70)$$

como esperado pelo princípio da razão insuficiente.

O fluido simples, introduzido por Gibbs, é um sistema onde o estado termodinâmico é determinado por algumas poucas variáveis. Suponhamos um gás isolado numa caixa de paredes adiabáticas. O experimental pode controlar separadamente a energia E , o volume V e o número de constituintes (e.g moléculas) N . A constante c na equação 1.70 é determinada pela imposição do vínculo de normalização,

$$P(q_i, p_i) = \frac{1}{\Omega(E, N, V)} \quad (1.71)$$

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

$$\Omega(E, N, V) = \int_{E \leq \mathcal{H}(q_i, p_i) \leq E + \delta E} \prod_i dq_i dp_i. \quad (1.72)$$

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

$$H_{extr}[P] = \int_{E \leq \mathcal{H}(q_i, p_i) \leq E + \delta E} \prod_i dq_i dp_i \frac{1}{\Omega(E, N, V)} \log \Omega(E, N, V) \quad (1.73)$$

e o extremo da entropia é

$$S(E, N, V) = \log \Omega(E, N, V). \quad (1.74)$$

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

Não é incomum que os livros de Mecânica Estatística comecem por *postular* que todos os microestados com energia E são igualmente prováveis. A conexão entre a Mecânica e a Termodinâmica é também *postulada* através da equação 1.74.

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

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

1.9 Conexão com a Termodinâmica: T, P, μ

1.9.1 Temperatura

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

$$E_1 + E_2 = E \quad (1.75)$$

$$V_1 + V_2 = V, \quad (1.76)$$

$$N_1 + N_2 = N, \quad (1.77)$$

tal que todos os termos das equações 1.76 e 1.77 conhecidos enquanto só o lado direito da equação 1.75 é conhecida. Como é feita a distribuição de energia entre os dois subsistemas? Suponhamos que o Hamiltoniano é bem descrito por

$$\mathcal{H}(\{q_i, p_i, Q_i, P_i\}) = \mathcal{H}_1(\{q_i, p_i\}) + \mathcal{H}_2(\{Q_i, P_i\}) \quad (1.78)$$

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

assim

$$\begin{aligned}
H[P(\{q_i, p_i, Q_i, P_i\})] &= - \int_{E_1 \leq \mathcal{H}_1(q_i, p_i) \leq E_1 + \delta E_1, E_2 \leq \mathcal{H}_2(Q_i, P_i) \leq E_2 + \delta E_2} \\
&\times P(\{q_i, p_i, Q_i, P_i\}) \log P(\{q_i, p_i, Q_i, P_i\}) \prod_i dq_i dp_i dQ_i dP_i \\
&= - \int_{E_1 \leq \mathcal{H}_1(q_i, p_i) \leq E_1 + \delta E_1} \prod_i dq_i dp_i P(\{q_i, p_i\}) \log P(\{q_i, p_i\}) \\
&- \int_{E_2 \leq \mathcal{H}_2(Q_i, P_i) \leq E_2 + \delta E_2} \prod_i dQ_i dP_i P(\{Q_i, P_i\}) \log P(\{Q_i, P_i\}) \\
H[P(\{q_i, p_i, Q_i, P_i\})] &= H[P(\{q_i, p_i\})] + H[P(\{Q_i, P_i\})] \tag{1.79}
\end{aligned}$$

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

$$\begin{aligned}
\Omega(E, N, V) &= \int_{E_1 \leq \mathcal{H}_1(\{q_i, p_i\}) \leq E_1 + \delta E_1, E_2 \leq \mathcal{H}_2(\{Q_i, P_i\}) \leq E_2 + \delta E_2} \prod_i dq_i dp_i dQ_i dP_i \\
&= \int_{E_1 \leq \mathcal{H}_1(\{q_i, p_i\}) \leq E_1 + \delta E_1} \prod_i dq_i dp_i \int_{E_2 \leq \mathcal{H}_2(\{Q_i, P_i\}) \leq E_2 + \delta E_2} \prod_i dQ_i dP_i \\
&= \Omega(E_1, N_1, V_1) \Omega(E_2, N_2, V_2), \tag{1.80}
\end{aligned}$$

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

$$S(E_1, E_2, N_1, N_2, V_1, V_2) = S(E_1, N_1, V_1) + S(E_2, N_2, V_2) \tag{1.81}$$

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

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

$$S(E_1, E_2, N_1, N_2, V_1, V_2) = \log \Omega(E_1, N_1, V_1) \Omega(E_2, N_2, V_2), \tag{1.82}$$

$$S(E_1, E_2, N_1, N_2, V_1, V_2) = S(E_1, N_1, V_1,) + S(E_2, N_2, V_2) \tag{1.83}$$

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

$$0 = \frac{\partial S(E_1, E_2, N_1, N_2, V_1, V_2)}{\partial E_1} = \frac{\partial S(E_1, N_1, V_1,)}{\partial E_1} + \frac{\partial S(E_2, N_2, V_2)}{\partial E_1} \tag{1.84}$$

$$= \frac{\partial S(E_1, N_1, V_1,)}{\partial E_1} - \frac{\partial S(E_2, N_2, V_2)}{\partial E_2} \tag{1.85}$$

$$\left(\frac{\partial S(E_1, N_1, V_1)}{\partial E_1} \right)_{V_1, N_1} = \left(\frac{\partial S(E_2, N_2, V_2)}{\partial E_2} \right)_{V_2, N_2}. \quad (1.86)$$

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

$$\left(\frac{\partial S(E, N, V)}{\partial E} \right)_{V, N} = \frac{1}{T} \quad (1.87)$$

e reescrevemos a equação 1.86 como $1/T_1 = 1/T_2$, os sistemas 1 e 2 estarão em equilíbrio quando suas temperaturas forem iguais. Isso também nos permitirá interpretar (ver seções 1.6.1 e 1.11) do ponto de vista experimental o parâmetro de Lagrange associado à um vínculo, no caso a temperatura com um vínculo sobre o energia. Os parâmetros intensivos da termodinâmica serão associados a multiplicadores de Lagrange que impoem vínculos sobre valores de grandezas extensivas.

1.9.2 Pressão

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

$$E_1 + E_2 = E \quad (1.88)$$

$$V_1 + V_2 = V, \quad (1.89)$$

$$N_1 + N_2 = N, \quad (1.90)$$

mas do lado esquerdo, só os termos da equação 1.90 são conhecidos. Novamente procuramos o máximo da equação

$$S(E_1, E_2, N_1, N_2, V_1, V_2) = S(E_1, N_1, V_1) + S(E_2, N_2, V_2) \quad (1.91)$$

mas agora devemos satisfazer duas condições:

$$\left(\frac{\partial S(E_1, E_2, N_1, N_2, V_1, V_2)}{\partial E_1} \right)_{V_1, N_1, N_2} = 0; \quad \left(\frac{\partial S(E_1, E_2, N_1, N_2, V_1, V_2)}{\partial V_1} \right)_{E_1, N_1, N_2} = 0 \quad (1.92)$$

que levam a

$$\left(\frac{\partial S(E_1, N_1, V_1)}{\partial E_1} \right)_{V_1, N_1} = \left(\frac{\partial S(E_2, N_2, V_2)}{\partial E_2} \right)_{V_2, N_2}. \quad (1.93)$$

e

$$\left(\frac{\partial S(E_1, N_1, V_1)}{\partial V_1} \right)_{V_1, N_1} = \left(\frac{\partial S(E_2, N_2, V_2)}{\partial V_2} \right)_{V_2, N_2}. \quad (1.94)$$

que significam respectivamente que

$$\frac{1}{T_1} = \frac{1}{T_2} \text{ e } \frac{P_1}{T_1} = \frac{P_2}{T_2} \quad (1.95)$$

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

1.9.3 Potencial químico

Consideremos a situação em que a parede que separa os dois subsistemas é rígida mas não impermeável. Só os termos da equação 1.89 são determinados pelo arranjo experimental. Podemos supor que o sistema é composto por vários tipos de moléculas

$$N_1^a + N_2^a = N^a \quad (1.96)$$

$$N_1^b + N_2^b = N^b \quad (1.97)$$

$$N_1^c + N_2^c = N^c \quad (1.98)$$

mas a parede só permite a passagem de moléculas do tipo a , sendo os termos da equações 1.97 e 1.98 conhecidos e fixos. Obteremos, por maximização da entropia do sistema conjunto que

$$\left(\frac{\partial S(E_1, N_1^a, N_1^b, N_1^c, V_1)}{\partial N_1^a} \right)_{E_1, N_1^b, N_1^c, V_1} = \left(\frac{\partial S(E_2, N_2^a, N_2^b, N_2^c, V_2)}{\partial N_2^a} \right)_{E_2, N_2^b, N_2^c, V_2} \quad (1.99)$$

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

$$-\frac{\mu}{T} = \left(\frac{\partial S(E, N, V)}{\partial N} \right)_{E, V} \quad (1.100)$$

$$\frac{1}{T_1} = \frac{1}{T_2} \text{ e } -\frac{\mu_1^a}{T_1} = -\frac{\mu_2^a}{T_2} \quad (1.101)$$

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

Exercício

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

1.10 Conexão com a Termodinâmica

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

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

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

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

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

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

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

A equação 1.78 mostra sob que condições ocorre a extensividade. Consideremos um sistema num estado (E, V, N) com interações de curto alcance. Imaginemos, sem fazer, uma separação em dois sistemas 1 e 2 como nas seções anteriores. Façamos a separação em K partes

$$S(E, V, N) = \sum_{i=1, K} S(E_i, V_i, N_i) \quad (1.102)$$

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

$$S(E, V, N) = KS(E/K, V/K, N/K) \quad (1.103)$$

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

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

$$dS(E, V, N) = \left(\frac{\partial S}{\partial E} \right)_{VN} dE + \left(\frac{\partial S}{\partial V} \right)_{EN} dV + \left(\frac{\partial S}{\partial N} \right)_{EV} d\mu \quad (1.104)$$

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

e usando as definições dos parâmetros intensivos

$$dS(E, V, N) = \frac{1}{T}dE + \frac{P}{T}dV - \frac{\mu}{T}d\mu \quad (1.105)$$

1.11 Ensemble Canônico

Há vezes em que uma experiência termodinâmica é feita numa situação diferente daquela em que o sistema está dentro de paredes isolantes, rígidas e impermeáveis. Por exemplo podemos olhar uma transformação a temperatura constante, a pressão constante ou a potencial químico constante. Queremos investigar o que significa isto em termos de vínculos de informação. Devemos olhar de novo a seção 1.5. Obtivemos resultados gerais para casos onde o valor esperado de certas quantidades é conhecido. O ensemble canônico de Gibbs é obtido quando o sistema não está isolado, mas a temperatura é fixa. Como implementar o vínculo informacional que a temperatura é fixa? Usamos a equação 1.62,

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

que nos dá indícios de como prosseguir.

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

$$\langle \mathcal{H} \rangle = \int \prod_i dq_i dp_i P(q_i, p_i) \mathcal{H}(q_i, p_i) \quad (1.106)$$

é fixo, podemos supor que não o conhecemos, mas tem um valor definido. Se soubessemos esse valor poderíamos aplicar os resultados da seção 1.5 e obter a distribuição - de Boltzmann, de Boltzmann-Gibbs, canônica - em analogia à equação 1.57, que agora toma a forma

$$P(\{q_i, p_i\}) = \frac{e^{-\beta \mathcal{H}}}{Z} \quad (1.107)$$

onde o multiplicador de Lagrange que chamamos β em lugar de λ , pode ser obtido da equação 1.62

$$\beta = \frac{\partial S(E, V, N)}{\partial E} \quad (1.108)$$

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

$$\beta = \frac{1}{T} \quad (1.109)$$

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

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

A função de partição Z adquire uma interpretação importante. Note que a equação 1.25 toma a forma

$$Z = \int \prod_i dq_i dp_i e^{-\beta \mathcal{H}} \quad (1.110)$$

e a equação 1.59 fica assim

$$E = \langle \mathcal{H} \rangle = - \frac{\partial \log Z}{\partial \beta} \quad (1.111)$$

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

$$S(E, V, N) = - \int \prod_i dq_i dp_i P(q_i, p_i) \log P(q_i, p_i) \quad (1.112)$$

$$= - \int \prod_i dq_i dp_i \frac{e^{-\beta \mathcal{H}}}{Z} \log \left(\frac{e^{-\beta \mathcal{H}}}{Z} \right) \quad (1.113)$$

$$= - \int \prod_i dq_i dp_i \frac{e^{-\beta \mathcal{H}}}{Z} (-\beta \mathcal{H} - \log Z) \quad (1.114)$$

$$= \beta \langle \mathcal{H} \rangle - \beta \log Z \quad (1.115)$$

$$= E - \beta \log Z. \quad (1.116)$$

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

$$F(T, V, N) = - \frac{1}{\beta} \log Z(\beta, V, N) \quad (1.117)$$

$$F = E - TS \quad (1.118)$$

1.11.1 Comparação da Entropia nos diferentes Ensembles

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

Exercício: Ensemble Grande Canônico

Considere a situação em os vínculos de informação sobre a experiência são o valor esperado da energia e do número de partículas. A notação que usamos é que $P(q_i, p_i, n) = P(n)P(\{q_i, p_i\}|n)$ denota a probabilidade atribuímos a que o sistema tenha n partículas e elas estejam no estado $\{q_i, p_i\}$. Obviamente dado um valor de n o número de variáveis depende de n , e.g. poderia ser $6n$ se o sistema estiver em três dimensões. O equivalente aos vínculos $\langle g_k \rangle$ da equação 1.54 tomam a forma

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

Encontre a distribuição de probabilidades chamada grande canônica ou grã canônica que maximiza o funcional de entropia sob os vínculos acima. Mostre que o multiplicador de Lagrange para o vínculo sobre o número de partículas está relacionado ao potencial químico $\lambda = \mu/T$, veja a equação 1.100. Voltaremos ao ensemble grande canônico em problemas importantes como o gás ideal classico e quântico que veremos nos próximos capítulos.

1.12 Apêndice: Multiplicadores de Lagrange

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

Seja o problema

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

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

Queremos, a seguir resolver um caso mais difícil.

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

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

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

$$F_\lambda(x, y) = f(x, y) + \lambda(\phi(x, y) - c) \quad (1.119)$$

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

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

$$\frac{\partial F_\lambda}{\partial x} = 0 \quad (1.120)$$

$$\frac{\partial F_\lambda}{\partial y} = 0, \quad (1.121)$$

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

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

¹³a não ser que não seja....

Capítulo 2

Aplicações Simples

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

2.1 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:

$$\mathcal{H} = -\mu_0 h \sum_{i=1}^N \sigma_i \quad (2.1)$$

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

2.1.1 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

$$N = N_+ + N_- \quad (2.2)$$

$$E = -\mu_0 h (N_+ - N_-) \quad (2.3)$$

de onde segue que

$$N_+ = \frac{1}{2}(N - \mathcal{E}) = \frac{N}{2}(1 - \epsilon) \quad (2.4)$$

$$N_- = \frac{1}{2}(N + \mathcal{E}) = \frac{N}{2}(1 + \epsilon) \quad (2.5)$$

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

$$\Omega(E, N) = \frac{N!}{N_+! N_-!} = \frac{N!}{(\frac{1}{2}(N - \mathcal{E}))! (\frac{1}{2}(N + \mathcal{E}))!}. \quad (2.6)$$

Usando a fórmula de Stirling para a expansão do fatorial obtemos

$$\begin{aligned} S(E, N) &= N \log N - N_+ \log N_+ - N_- \log N_- \\ &= -N \left(\frac{1 - \epsilon}{2} \log(1 - \epsilon) + \frac{1 + \epsilon}{2} \log(1 + \epsilon) \right) \end{aligned} \quad (2.7)$$

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

$$\begin{aligned} \frac{1}{T} &= \frac{\partial S(E, N)}{\partial E} = \frac{\partial \epsilon}{\partial E} \frac{\partial S}{\partial \epsilon} \\ &= \frac{1}{2\mu_0 h} \log \left(\frac{1 - \epsilon}{1 + \epsilon} \right) \end{aligned} \quad (2.8)$$

que pode ser invertida, dando

$$E = \mu_0 h N \epsilon = -\mu_0 h N \tanh \mu_0 h T \quad (2.9)$$

e para a magnetização por grau de liberdade

$$m := \left\langle \frac{\sum_{i=1}^N \sigma_i}{N} \right\rangle = \mu_0 \tanh \frac{\mu_0 h}{T} \quad (2.10)$$

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

$$\chi_T = \left(\frac{\partial m}{\partial h} \right)_T = \frac{\mu_0^2}{T} \frac{1}{\cosh^2 \frac{\mu_0 h}{T}} \quad (2.11)$$

mostrada na figura 2.3. 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.

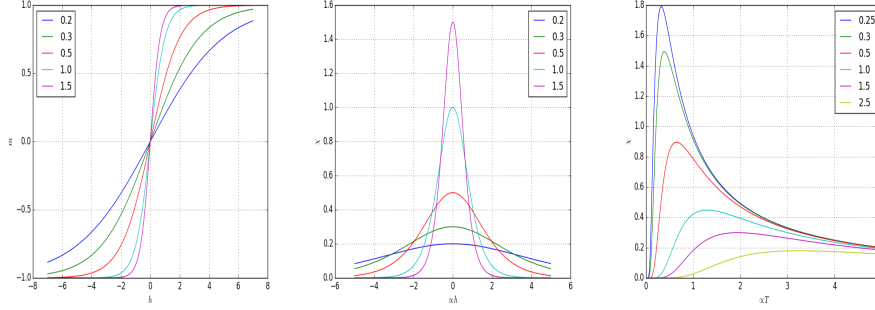


Figura 2.1: Esquerda: A magnetização por grau de liberdade m como função do campo h externo para diferentes valores da temperatura inversa β ($\mu_0 = 1$). Centro: Suscetibilidade magnética (equação 2.11) como função do campo h , para diferentes valores de β . Direita: $\chi(T)$ para diferentes valores de h .

2.1.2 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

$$P(\{\sigma_i\}) = \frac{e^{\beta\mu_0 h \sum_{i=1}^N \sigma_i}}{Z} \quad (2.12)$$

com a função de partição

$$Z(\beta, N, h) = \sum_{\{\sigma_i\}} e^{\beta\mu_0 h \sum_{i=1}^N \sigma_i} \quad (2.13)$$

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

$$\begin{aligned} Z(\beta, N, h) &= \prod_{i=1}^N \sum_{\sigma_i = \pm 1} e^{\beta\mu_0 h \sigma_i} \\ &= (2 \cosh \beta\mu_0 h)^N \end{aligned} \quad (2.14)$$

Note que da equação 2.13 obtemos

$$\frac{\partial \log Z}{\partial h} = \beta\mu_0 \left\langle \sum_{i=1}^N s_i \right\rangle \quad (2.15)$$

e da equação 2.14

$$\frac{\partial \log Z}{\partial h} = \beta \mu_0 N \tanh(\beta \mu_0 h) \quad (2.16)$$

e portanto chegamos a uma equação igual a equação 2.10.

Exercício Resfriamento magnético: Na figura 2.3 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 2.3. 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*.

2.2 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 :

$$E \leq \mathcal{H}(\{q_i, p_i\}) \leq E + \delta E, \quad (2.17)$$

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

$$\Omega_{dist}(E, V, N) = \int_{\mathcal{H}=E, V} \prod dq_i dp_i \quad (2.18)$$

mas

$$\Omega(E, V, N) = \int_{\mathcal{H}=E, V} \frac{\prod dq_i dp_i}{N! h^{3N}} \quad (2.19)$$

onde h é por agora uma constante em 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 2.19 é muito simples. Para cada partícula temos uma integral sobre os valores possíveis das coordenadas, portanto cada partícula contribui com um fator V .

$$\Omega(E, V, N) = \frac{V^N}{N!} \int_{\mathcal{H}=E} \prod dp_i, \quad (2.20)$$

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

$$\Omega(E, V, N) = \frac{V^N}{N!} \frac{2\pi^{\frac{3N}{2}}}{\Gamma(3N/2)} E^{\frac{3N-1}{2}} \delta E \quad (2.21)$$

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 1.74 que a entropia é

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

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

$$S(E, V, N) = N \left(\log \frac{V}{N} + \frac{3}{2} \log \frac{E}{N} - \frac{1}{2} + \frac{3}{2} \log \frac{3\pi}{2} \right) + \dots \quad (2.22)$$

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 $v = V/N$.

$$s(u, v) = \frac{S}{N} = \log v + \frac{3}{2} \log u - \frac{1}{2} + \frac{3}{2} \log \frac{3\pi}{2}. \quad (2.23)$$

¹Uma definição pragmática de paradoxo: algo que está errado mas é fruto de um raciocínio a primeira vista correto.

A seguir investigamos as equações de estado obtidas ao tomar as derivadas da entropia com respeito à energia, volume e número de partículas

$$\left(\frac{\partial S}{\partial E}\right)_{V,N} = \frac{1}{T} = \frac{3N}{2E} \quad (2.24)$$

$$\left(\frac{\partial S}{\partial V}\right)_{E,N} = \frac{P}{T} = \frac{N}{V} \quad (2.25)$$

$$\left(\frac{\partial S}{\partial N}\right)_{E,V} = \frac{-\mu}{T} = \log \frac{V}{N} + \frac{3}{2} \log \frac{E}{N} - 3 + \frac{3}{2} \log \frac{3\pi}{2} \quad (2.26)$$

(VERIFICAR CONTAS) onde reconhecemos as célebres equações de estado:

$$PV = NT \quad (2.27)$$

$$E = \frac{3}{2}NT \quad (2.28)$$

2.3 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 $\mathcal{H}(\{q_i, p_i\}) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m}$ fosse

$$E = \langle \mathcal{H}(\{q_i, p_i\}) \rangle. \quad (2.29)$$

A densidade de probabilidades é

$$P(\{q_i, p_i\}) = \frac{\exp(-\beta \mathcal{H}(\{q_i, p_i\}))}{Z} \quad (2.30)$$

onde

$$Z = \int \frac{\prod dq_i dp_i}{N!} \exp(-\beta \mathcal{H}), \quad (2.31)$$

que impõe a normalização

$$1 = \int \frac{\prod dq_i dp_i}{N!} P(\{q_i, p_i\}). \quad (2.32)$$

A equação relevante é

$$E = -\frac{\partial \log Z}{\partial \beta} \quad (2.33)$$

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 :

$$\zeta_i = \int dq_i dp_i \exp(-\beta \mathcal{H}_i), \quad (2.34)$$

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

$$Z = \frac{1}{N!} \zeta^N. \quad (2.35)$$

A integração sobre as coordenadas é análoga à secção anterior, dando V^N .

Para integrar os momentos usamos:

$$\int_{-\infty}^{\infty} e^{-\frac{\beta}{2m} p^2} dp = \sqrt{2\pi \frac{m}{\beta}} \quad (2.36)$$

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

$$Z = \frac{V^N}{N!} \left(2\pi \frac{m}{\beta}\right)^{\frac{3N}{2}} \quad (2.37)$$

$$\zeta = V \left(2\pi \frac{m}{\beta}\right)^{\frac{3}{2}} \quad (2.38)$$

e a equação 2.33 dá:

$$E = \frac{3N}{2\beta} = 3N \frac{T}{2}. \quad (2.39)$$

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 1.116 através de $-\beta F = \log Z$, portanto $S = \log Z + \beta E$. A pressão no ensemble canônico, obtida através a equação 2.25 $\left(\frac{\partial S}{\partial V}\right)_{V,N} = \frac{P}{T}$, 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 $\log N!$

$$-\beta F = N \log V + \frac{3N}{2} \log \frac{2\pi m}{\beta} - N \log N + N \quad (2.40)$$

²Não devemos usar a palavra teorema em Física, teoremas se aplicam a estruturas matemáticas e não a questões empíricas. A falha do *teorema* em prever corretamente resultados empíricos levou à conclusão, não que a matemática estava errada, mas que essa particular estrutura matemática não era útil para acomodar os fatos empíricos.

e definindo as densidades $f = F/N$, $v = V/N$ e $\epsilon = E/N$

$$-\beta f = \log v + \frac{3}{2} \log \frac{2\pi m}{\beta} + 1 \quad (2.41)$$

de onde fica claro que se o fatorial não estivesse presente as densidades não seriam intensivas.

2.4 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

$$\mathcal{H} = \sum_i \mathcal{H}_l(i), \quad (2.42)$$

e podemos supor que

$$\mathcal{H}_l(i) = \frac{\mathbf{p}_i^2}{2m_l} + \varepsilon_l(s_i) \quad (2.43)$$

onde s_i representa os graus de liberdade internos da molécula i que é do tipo l e $\varepsilon_l(s_i)$ a energia associada. A função de partição associada a essa molécula do tipo l é

$$\zeta_l = \int d\mathbf{q} d\mathbf{p} \sum_s e^{-\beta \mathcal{H}_l}, \quad (2.44)$$

e a função de partição do sistema

$$Z(\beta, V, N_1, \dots, N_m) = \prod_{l=1, \dots, m} \frac{\zeta_l^{N_l}}{N_l!}, \quad (2.45)$$

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

$$-\beta F = \log Z = \sum_l (N_l \log \zeta_l - \log N_l + 1). \quad (2.46)$$

A pressão pode ser facilmente calculada pois cada fator ζ_l contribui com um fator V , portanto

$$-\beta F = \sum_l (N_l \log V) + \dots, \quad (2.47)$$

onde não escrevemos termos independentes do volume, segue que

$$\beta p = \left(\frac{\partial \log Z}{\partial V} \right)_{T, N_1, \dots, N_l, \dots, N_m} \quad (2.48)$$

$$\beta p = \sum N_l \frac{1}{V} = \beta \sum_{l=1}^m p_l, \quad (2.49)$$

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} .

2.4.1 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

$$\frac{dN_1}{c_1} = \frac{dN_2}{c_2} = \dots = \frac{dN_l}{c_l} = \dots = \frac{dN_m}{c_m} \quad (2.50)$$

A energia livre $F(T, V, \{N_l\}_{l=1, \dots, m})$ é mínima no equilíbrio, portando flutuações dN_l devem satisfazer

$$0 = \Delta F = \sum_l \left(\frac{\partial F}{\partial N_l} \right)_{T, V, N_1, \dots, N_{l'} \neq l, \dots, N_m} dN_l. \quad (2.51)$$

Lembrando que estas derivadas de F são os potenciais químicos

$$\mu_l = \left(\frac{\partial F}{\partial N_l} \right)_{T, V, N_{l'} \neq l} \quad (2.52)$$

podemos ver que

$$\sum_l c_l \mu_l = 0. \quad (2.53)$$

E dada a relação entre a energia livre e a função de partição ($-\beta F = \log Z$), obtemos

$$\mu_l = -T \log \frac{\zeta_l}{N_l} = f_l - T \log N_l \quad (2.54)$$

onde introduzimos $f_l = -\beta^{-1} \log \zeta_l$, a soma $F_0 = \sum_l f_l$ e temos

$$0 = T \sum_l c_l \log \frac{\zeta_l}{N_l} = - \sum_l f_l - T \sum_l \log N_l^{c_l} \quad (2.55)$$

$$\prod_l N_l^{c_l} = e^{-\beta F_0} = K(T, V, \{C\}), \quad (2.56)$$

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

$$\frac{N_A^2 N_B}{N_C^2} = K(T, V, A, B, C) \quad (2.58)$$

2.5 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

$$\mathcal{H}(\{q_i, p_i\}) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} + V(\{q_i\}) \quad (2.59)$$

pode ser aproximado por

$$\mathcal{H}(\{q_i, p_i\}) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} + \sum_i V(q_i) \quad (2.60)$$

onde o potencial $V(q_i)$ só depende das coordenadas do átomo i e ainda mais, que é harmônico

$$\mathcal{H}(\{q_i, p_i\}) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} + \sum_i \frac{m\omega^2}{2} \mathbf{q}_i^2 \quad (2.61)$$

A função de partição é novamente gaussiana e além da equação 2.36 usamos

$$\int_{-\infty}^{\infty} e^{-\frac{\beta m \omega^2}{2} q^2} dq = \sqrt{2\pi \frac{1}{\beta m \omega^2}}. \quad (2.62)$$

A função de partição

$$Z = \left(\sqrt{2\pi \frac{m}{\beta}} \sqrt{2\pi \frac{1}{\beta m \omega^2}} \right)^{3N} \quad (2.63)$$

A energia, obtida derivando o logaritmo de Z dá:

$$E = \frac{6TN}{2} = 3NT (= 3nRT_K), \quad (2.64)$$

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 diatômicas. 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.

2.6 Sólido de Einstein

Einstein considera os $3N$ ³ osciladores descritos acima quantizados e independentes, o que significa que o microestado do sistema é descrito por $\{n_1, n_2, \dots, n_N\}$ e a energia

$$\mathcal{H}(\{n_1, n_2, \dots, n_{3N}\}) = \hbar\omega \sum_{i=1}^{3N} \left(\frac{1}{2} + n_i \right) \quad (2.65)$$

2.6.1 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

$$\mathcal{H}(\{n_1, n_2, \dots, n_{3N}\}) = E = \frac{3N}{2}\hbar\omega + K\hbar\omega \quad (2.66)$$

O número de microestados $\Omega(E, N)$ é dado pelas combinações indistinguíveis do ponto de vista experimental de $3N + K$ objetos, formados por duas classes de

³N osciladores tridimensionais

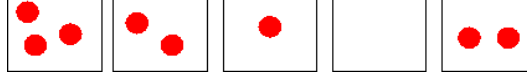


Figura 2.2: Exemplo: uma das 1287 configurações distintas com $K = 8$ quanta e $N = 5$ osciladores

objetos diferentes mas indistinguíveis entre si dentro da classe: $3N$ osciladores e K quanta de energia $\hbar\omega$:⁴

$$\Omega(E, N) = \frac{(3N + K)!}{(3N)!K!} \quad (2.67)$$

Definimos $\varepsilon = E/(N\hbar\omega)$, a energia por partícula medida em unidades de $\hbar\omega$. Para a entropia temos

$$S(E, N) = (3N + K) \log(3N + K) - 3N \log 3N - K \log K \quad (2.68)$$

$$= N \left(\left(\varepsilon + \frac{3}{2} \right) \log \left(\varepsilon + \frac{3}{2} \right) - \left(\varepsilon - \frac{3}{2} \right) \log \left(\varepsilon - \frac{3}{2} \right) \right) \quad (2.69)$$

onde usamos $\varepsilon = 3/2 + K/N$. Assim podemos calcular a temperatura

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E} \right)_N = \frac{1}{\hbar\omega} \log \frac{\varepsilon + \frac{3}{2}}{\varepsilon - \frac{3}{2}} \quad (2.70)$$

e invertendo temos

$$\varepsilon = \frac{3}{2} + \frac{3}{e^{\frac{\hbar\omega}{T}} - 1} \quad (2.71)$$

$$E = \frac{3N\hbar\omega}{2} + \frac{3N\hbar\omega}{e^{\frac{\hbar\omega}{T}} - 1} \quad (2.72)$$

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

$$E = \frac{3N\hbar\omega}{2} + 3N\hbar\omega \langle n \rangle \quad (2.73)$$

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

$$c_v = \frac{1}{N} \frac{\partial E}{\partial T} = 3 \frac{(\hbar\omega)^2 e^{\frac{\hbar\omega}{T}}}{T^2 (e^{\frac{\hbar\omega}{T}} - 1)^2}, \quad (2.74)$$

e tomando os limite de pequenas temperaturas vemos que o calor específico não é constante mas cai com a temperatura

$$c_v \approx 3 \frac{(\hbar\omega)^2 e^{-\frac{\hbar\omega}{T}}}{T^2}, \quad (2.75)$$

⁴Na realidade devemos considerar $3N - 1$ osciladores mas não faz diferença

enquanto que para valores altos da temperatura

$$c_v \approx 3 \quad (2.76)$$

devemos colocar as unidades costumeiras para obter $c_v \approx 3R$ por mol.

2.6.2 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

$$P(\{n_i\}) = \frac{e^{-\beta \mathcal{H}(\{n_i\})}}{Z} \quad (2.77)$$

onde

$$Z(\beta, N) = \sum_{n_1 \dots n_{3N}} e^{-\beta \mathcal{H}} \quad (2.78)$$

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

$$Z(\beta, N) = \left[\sum_{n=0}^{\infty} e^{-\beta(n+\frac{1}{2})\hbar\omega} \right]^{3N}. \quad (2.79)$$

$$Z(\beta, N) = e^{-\frac{3N\beta}{2}\hbar\omega} [1 + a + a^2 + a^3 + \dots]^{3N}, \quad (2.80)$$

onde $a = e^{-\beta\hbar\omega}$ e somando a série geométrica

$$Z(\beta, N) = e^{-\frac{3N\beta}{2}\hbar\omega} \left[\frac{1}{1 - e^{-\beta\hbar\omega}} \right]^{3N}. \quad (2.81)$$

Usando $E = -\partial \log Z / \partial \beta$, obtemos

$$E = \frac{3N\hbar\omega}{2} + \frac{3N\hbar\omega}{e^{\beta\hbar\omega} - 1} \quad (2.82)$$

que é exatamente o resultado obtido no ensemble microcanônico

2.6.3 Gases ideais clássicos com estrutura interna quântica

a ser escrito ...

2.7 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

$$E = \langle E_\alpha \rangle, \quad N = \langle N_\alpha \rangle \quad V = \text{fixo}. \quad (2.83)$$

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:

$$E = \sum_{\alpha} P_{\alpha} E_{\alpha} \quad N = \sum_{\alpha} P_{\alpha} N_{\alpha}, \quad 1 = \sum_{\alpha} P_{\alpha}, \quad (2.84)$$

que inclui a normalização.

Maximizamos a entropia de Shannon sujeita a estes vínculos

$$H[P_{\alpha}] = - \sum_{\alpha} P_{\alpha} \log P_{\alpha} + \lambda_0 (1 - \sum_{\alpha} P_{\alpha}) + \beta (E - \sum_{\alpha} P_{\alpha} E_{\alpha}) + \lambda_1 (N - \sum_{\alpha} P_{\alpha} N_{\alpha}), \quad (2.85)$$

e obtemos

$$P_{\alpha} = e^{1-\lambda_0} e^{-\beta E_{\alpha} - \lambda_1 N_{\alpha}}. \quad (2.86)$$

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

$$P_{\alpha} = \frac{1}{\Xi(\beta, \lambda_1, V)} e^{-\beta E_{\alpha} - \lambda_1 N_{\alpha}} \quad (2.87)$$

onde

$$\Xi(\beta, \lambda_1, V) = \sum_{\alpha} e^{-\beta E_{\alpha} - \lambda_1 N_{\alpha}}. \quad (2.88)$$

Usamos, como é costume, a letra ξ maiúscula Ξ para enfatizar que isto levará a um potencial termodinâmico diferente da energia livre de Helmholtz F , que aparece no ensemble canônico. A entropia termodinâmica será identificada com o máximo da entropia de Shannon, ou seja substituímos a distribuição de probabilidades, equação 2.87, na entropia de Shannon 2.85

$$S(E, N, V) = - \sum_{\alpha} P_{\alpha} \log P_{\alpha}, \quad (2.89)$$

para obter

$$S(E, N, V) = \log \Xi + \beta E + \lambda_1 N. \quad (2.90)$$

O leitor deveria ver isto como semelhante à equação 1.116. Não é a mesma equação porque estamos descrevendo um sistema em situação experimental diferente, onde o número de partículas está fixo somente através do seu valor médio. Para interpretar o significado de λ_1 , devemos fazer o análogo à equação 1.60. Primeiro notemos que

$$\left. \frac{\partial \log \Xi}{\partial \beta} \right|_{\lambda_1, V} = -E \quad (2.91)$$

$$\left. \frac{\partial \log \Xi}{\partial \lambda_1} \right|_{\beta, V} = -N \quad (2.92)$$

$$\left. \frac{\partial \log \Xi}{\partial V} \right|_{\beta, \lambda_1} = -\frac{P}{T} \quad (2.93)$$

Agora, usando a regra da cadeia

$$\left. \frac{\partial S}{\partial E} \right|_N = \frac{\partial \log \Xi}{\partial \beta} \frac{\partial \beta}{\partial E} + \frac{\partial \log \Xi}{\partial \lambda_1} \frac{\partial \lambda_1}{\partial E} + \frac{\partial \beta}{\partial E} E + \frac{\partial \lambda_1}{\partial E} N + \beta \quad (2.94)$$

$$\left. \frac{\partial S}{\partial E} \right|_N = \beta, \quad (2.95)$$

obtemos novamente a interpretação que β é o inverso da temperatura. Derivando com respeito a N :

$$\left. \frac{\partial S}{\partial N} \right|_E = \frac{\partial \log \Xi}{\partial \beta} \frac{\partial \beta}{\partial N} + \frac{\partial \log \Xi}{\partial \lambda_1} \frac{\partial \lambda_1}{\partial N} + \frac{\partial \beta}{\partial N} E + \frac{\partial \lambda_1}{\partial N} N + \lambda_1 \quad (2.96)$$

$$\left. \frac{\partial S}{\partial N} \right|_E = \lambda_1, \quad (2.97)$$

de onde obtemos, usando a definição do potencial químico (ver equação 1.100)

$$\lambda_1 = -\frac{\mu}{T} = -\beta\mu. \quad (2.98)$$

Segue que

$$\Xi(\beta, \mu) = \sum_{\alpha} e^{-\beta(E_{\alpha} - \mu N_{\alpha})}. \quad (2.99)$$

Para referência futura note que

$$E = \langle E_{\alpha} \rangle = -\left. \frac{\partial \log \Xi}{\partial \beta} \right|_{\mu, V} \quad (2.100)$$

$$N = \langle N_{\alpha} \rangle = \beta^{-1} \left. \frac{\partial \log \Xi}{\partial \mu} \right|_{\beta, V} \quad (2.101)$$

$$P = -\beta^{-1} \left. \frac{\partial \log \Xi}{\partial V} \right|_{\beta, \mu} \quad (2.102)$$

2.8 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:

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r}) + V(\mathbf{r})\psi(\mathbf{r}) = \epsilon\psi(\mathbf{r}) \quad (2.103)$$

obtendo os autoestados de energia $\epsilon_{\mathbf{q}}$, associados ao momento $\mathbf{p} = \hbar\mathbf{q}$

$$\epsilon_{\mathbf{q}} = \frac{\hbar^2\mathbf{q}^2}{2m}. \quad (2.104)$$

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

$$\mathbf{q} = (q_x, q_y, q_z), \text{ com } q_x = \frac{2\pi}{L}l_x, q_y = \frac{2\pi}{L}l_y, q_z = \frac{2\pi}{L}l_z \quad (2.105)$$

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:

$$\left(-\frac{\hbar^2}{2m}\sum_i^{N_\alpha}\nabla_i^2 + V(\mathbf{r})\right)\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{N_\alpha}) = E_\alpha\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{N_\alpha}), \quad (2.106)$$

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 2.103 é solução da equação 2.106.

Mas isto não resolve o problema. Suponha duas partículas e duas soluções de 2.103, $\psi_a(\mathbf{r}_1)$ e $\psi_b(\mathbf{r}_2)$. A função $\Psi_{a1,b2} = \psi_a(\mathbf{r}_1)\psi_b(\mathbf{r}_2)$ é solução de 2.106

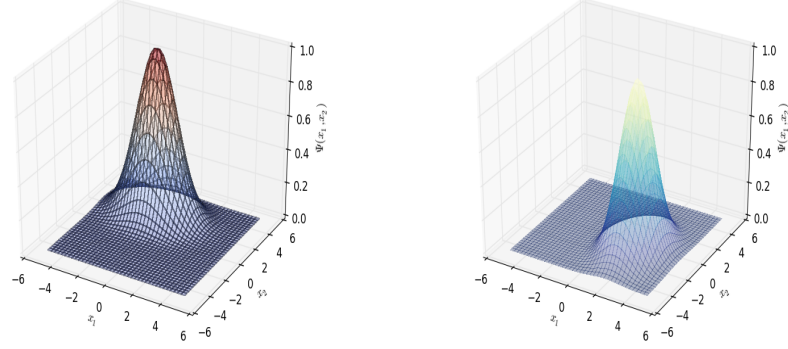


Figura 2.3: (linhas grossas) $\Psi_{a1,b2} = \psi_a(x_1)\psi_b(x_2) \neq \psi_a(x_2)\psi_b(x_1) = \Psi_{a2,b1}$ (linhas finas)

mas não leva em conta a indistinguibilidade das partículas. Note que colocar a partícula 1 no estado a e a 2 no b é diferente de colocar a 1 em b e a 2 em a , i.e.

$$\Psi_{a1,b2} \neq \Psi_{a2,b1} \quad (2.107)$$

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

$$\Psi_{\pm}(\mathbf{r}_1, \mathbf{r}_2) \propto \psi_a(\mathbf{r}_1)\psi_b(\mathbf{r}_2) \pm \psi_a(\mathbf{r}_2)\psi_b(\mathbf{r}_1) \quad (2.108)$$

satisfaz $\Psi_{\pm}(\mathbf{r}_1, \mathbf{r}_2) = \pm \Psi_{\pm}(\mathbf{r}_2, \mathbf{r}_1)$, logo as densidades de probabilidade permanecem iguais quando as duas partículas são intercambiadas.

$$|\Psi_{\pm}(\mathbf{r}_1, \mathbf{r}_2)|^2 = |\Psi_{\pm}(\mathbf{r}_2, \mathbf{r}_1)|^2 \quad (2.109)$$

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 2.109 deva ser satisfeita. Isto em princípio poderia levar ao aparecimento de uma fase

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_k, \mathbf{r}_l, \dots) = e^{i\phi} \Psi(\mathbf{r}_1, \dots, \mathbf{r}_l, \mathbf{r}_k, \dots) \quad (2.110)$$

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 quais partículas estão nesse estado mas quantas. Portanto o estado $-N_\alpha$ será caracterizado por o conjunto de números de ocupação. Como cada estado -1 é caracterizado por um conjunto de números quânticos $\mathbf{q}$, então o estado $-N_\alpha$ é representado por $:(n_{q_1}, n_{q_2}, \dots, n_{q_l}, \dots)$, com

$$N_\alpha = \sum_{\mathbf{q}} n_{\mathbf{q}}. \quad (2.111)$$

No caso do gás ideal o $\mathbf{q}$ são simplesmente os momentos permitidos na caixa pelas condições de contorno e portanto

$$E_\alpha = \sum_{\mathbf{q}} n_{\mathbf{q}} \epsilon_{\mathbf{q}}, \quad (2.112)$$

pois a energia total do estado $-N_\alpha$ é a soma das energias dos estados -1 ocupados e simplesmente juntamos os termos iguais.

Exercício Mostre a equação 2.112 a partir da equação 2.106.

O que significa que uma função de onda seja simétrica (+) ou antisimétrica (-) ao intercambiar o índice de duas partículas? Consideremos

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_k, \mathbf{r}_l, \dots) = \pm \Psi(\mathbf{r}_1, \dots, \mathbf{r}_l, \mathbf{r}_k, \dots) \quad (2.113)$$

para o caso de duas partículas. Segue que

$$\Psi_{BE}(\mathbf{r}_1 \mathbf{r}_2) = \frac{1}{\sqrt{2}} (\psi_a(\mathbf{r}_1) \psi_b(\mathbf{r}_2) + \psi_a(\mathbf{r}_2) \psi_b(\mathbf{r}_1)) \quad (2.114)$$

$$\Psi_{FD}(\mathbf{r}_1 \mathbf{r}_2) = \frac{1}{\sqrt{2}} (\psi_a(\mathbf{r}_1) \psi_b(\mathbf{r}_2) - \psi_a(\mathbf{r}_2) \psi_b(\mathbf{r}_1)) \quad (2.115)$$

Para o sinal +, Ψ_{BE} não apresenta nenhum comportamento estranho se $a = b$, isto é se as duas partículas ocupam o mesmo estado-1. Já para o sinal -, Ψ_{FD} é zero se $a = b$, portanto os estados aceitáveis serão aqueles em não há dois férmions no mesmo estado. Temos o resultado que os estados $-N_\alpha$, que são descritos pelo conjunto de números de ocupação devem ser

$$\text{Bósons } \alpha \leftrightarrow (n_1, n_2, \dots) \text{ sem nenhuma restrição: } n_{\mathbf{q}} = 0 \dots \infty, \quad (2.116)$$

$$\text{Férmions } \alpha \leftrightarrow (n_1, n_2, \dots) \text{ restrição: } n_{\mathbf{q}} = 0, 1. \quad (2.117)$$

A generalização para mais de dois férmions foi feita por Slater que percebeu que a expressão 2.115 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 $\Xi(\beta, \mu, V) = \sum_\alpha e^{-\beta(E_\alpha - \mu N_\alpha)}$,

$$\begin{aligned} \Phi(\beta, \mu, V) &= -T \log \Xi \\ &= -T \log \sum_{n_{q_l}} e^{-\beta \sum_{q_l} n_{q_l} (\epsilon_{q_l} - \mu)} \\ &= -T \log \prod_{q_l} \sum_{n_{q_l}} e^{-\beta n_{q_l} (\epsilon_{q_l} - \mu)} \\ &= -T \sum_{q_l} \log \sum_{n_{q_l}} e^{-\beta n_{q_l} (\epsilon_{q_l} - \mu)} \end{aligned} \quad (2.118)$$

$$= \sum_{q_l} \Phi_{q_l}. \quad (2.119)$$

Ainda não sabemos como fazer a soma sobre os q e trataremos disso mais tarde. Por agora nos concentraremos nas somas sobre n_{q_l} , levando em conta 2.116 e 2.117. Para o caso de Bósons, a estatística Bose-Einstein será governada pelo potencial

$$\begin{aligned} \Phi_{q_l}^{BE} &= -T \log \sum_{n_{q_l}=0}^{\infty} e^{-\beta n_{q_l} (\epsilon_{q_l} - \mu)} \\ &= -T \log(1 + x + x^2 + x^3 + \dots) \\ &= -T \log\left(\frac{1}{1-x}\right) \\ &= T \log(1 - e^{-\beta(\epsilon_{q_l} - \mu)}) \end{aligned} \quad (2.120)$$

A convergência da série geométrica só ocorrerá caso $x = e^{-\beta(\epsilon_{q_l} - \mu)}$ seja menor que 1. Como o menor valor de ϵ_{q_l} , no limite de volume infinito, é zero, devemos ter um potencial químico negativo

$$x < 1 \rightarrow \mu_{BE} < 0 \quad (2.121)$$

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:

$$\begin{aligned} \Phi_{q_l}^{FD} &= -T \log \sum_{n_{q_l}=0}^1 e^{-\beta n_{q_l} (\epsilon_{q_l} - \mu)} \\ &= -T \log(1 + x) \\ &= -T \log(1 + e^{-\beta(\epsilon_{q_l} - \mu)}), \end{aligned} \quad (2.122)$$

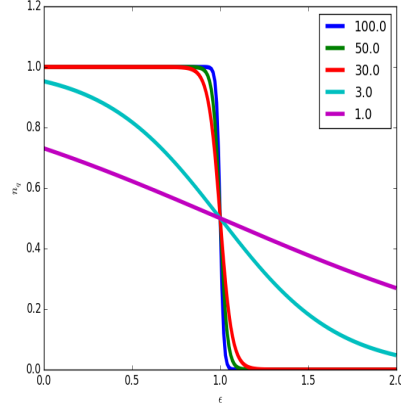


Figura 2.4: A função de Fermi: $n = (\exp(\beta(\epsilon - 1)) + 1)^{-1}$, para diferentes valores de β .

Juntando os resultados

$$\Phi_{\mathbf{q}_l} = \mp T \log(1 \pm e^{-\beta(\epsilon_{\mathbf{q}_l} - \mu)}) \quad (2.123)$$

com $-+$ para férmions e $+-$ para bósons. Se a exponencial na expressão acima for pequena $e^{-\beta(\epsilon_{\mathbf{q}_l} - \mu)} \ll 1$ a expansão do logaritmo ($\log(1 + x) \approx x$) leva ao mesmo resultado tanto para férmions quanto bósons:

$$\Phi_{\mathbf{q}_l}^C = -T e^{-\beta(\epsilon_{\mathbf{q}_l} - \mu)} \quad (2.124)$$

que veremos, recupera os resultados para o limite clássico.

Uma quantidade central na discussão que segue é o *número médio de ocupação* $\bar{n}_{\mathbf{q}}$ que pode ser calculado a partir de (ver equação 2.101)

$$\bar{n}_{\mathbf{q}} = \langle n_{\mathbf{q}} \rangle = \frac{\partial \Phi_{\mathbf{q}}}{\partial \mu} \Big|_{\beta, V} . \quad (2.125)$$

Obtemos para bósons

$$\bar{n}_{\mathbf{q}}^{BE} = \frac{1}{e^{\beta(\epsilon_{\mathbf{q}_l} - \mu)} - 1} \quad (2.126)$$

e para férmions

$$\bar{n}_{\mathbf{q}}^{FD} = \frac{1}{e^{\beta(\epsilon_{\mathbf{q}_l} - \mu)} + 1} \quad (2.127)$$

Estas expressões são aparentemente parecidas, mas como veremos são muito diferentes para baixas temperaturas. A diferença ocorre devido a que o potencial químico se comporta de maneira muito diferente nos dois casos, para ambos os

casos é claro que pode ser determinado de forma implícita pelo vínculo que o número médio de partículas é dado

$$N_{BE} = \sum_{\mathbf{q}} \bar{n}_{\mathbf{q}} = \sum_{\mathbf{q}} \frac{1}{e^{\beta(\epsilon_{\mathbf{q}} - \mu)} - 1} \quad (2.128)$$

$$N_{FD} = \sum_{\mathbf{q}} \bar{n}_{\mathbf{q}} = \sum_{\mathbf{q}} \frac{1}{e^{\beta(\epsilon_{\mathbf{q}} - \mu)} + 1} \quad (2.129)$$

Na figura 2.4 mostramos a função de Fermi. Não é o número de ocupação porque ainda não sabemos como depende o potencial químico. O número de ocupação para BE é mais difícil de desenhar, e será deixado para após a análise do potencial químico. Para esta discussão precisamos entender a soma sobre os estados de uma partícula e para tanto é necessário introduzir a idéia de densidade de estados.

2.9 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

$$\mathcal{N}_1 = \Delta l_x = \frac{L}{2\pi} \Delta q_x = \frac{L}{2\pi\hbar} \Delta p_x \quad (2.130)$$

Ao fazer a conta integrando sobre momentos num intervalo pequeno Δp_x

$$\int_{L, \Delta p_x} \frac{dx dp_x}{h} = \frac{L \Delta p_x}{h}. \quad (2.131)$$

Na equação 2.19 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 autovaleiros 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

$$\mathcal{N}_3 = \frac{L^3 \Delta p_x \Delta p_y \Delta p_z}{(2\pi\hbar)^3}, \quad (2.132)$$

e em coordenadas esféricas, tomando os Δp como diferenciais, teremos a medida de integração

$$\frac{4\pi V p^2 dp}{(2\pi\hbar)^3}, \quad (2.133)$$

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 2.104

$$\epsilon_{\mathbf{q}} = \frac{\hbar^2 \mathbf{q}^2}{2m}, \quad (2.134)$$

leva, usando $d\epsilon = \frac{d\epsilon}{dp} dp = \frac{p dp}{m}$, a

$$\mathcal{N}_3 = \frac{4\pi V}{(2\pi\hbar)^3} \sqrt{2m^{\frac{3}{2}}} \epsilon^{\frac{1}{2}} d\epsilon, \quad (2.135)$$

e introduzindo $\rho(\epsilon)$, a densidade de estados de energia de uma partícula não relativística em 3 dimensões

$$\rho(\epsilon) d\epsilon = \frac{4\pi V}{(2\pi\hbar)^3} \sqrt{2m^{\frac{3}{2}}} \epsilon^{\frac{1}{2}} d\epsilon = CV \epsilon^{\frac{1}{2}} d\epsilon, \quad (2.136)$$

onde

$$C = \frac{4\pi}{(2\pi\hbar)^3} \sqrt{2m^{\frac{3}{2}}} \quad (2.137)$$

Exploraremos a seguir a aproximação de somas por integrais:

$$\sum_{\mathbf{q}} (\cdot) \rightarrow \int (\cdot) \rho(\epsilon) d\epsilon. \quad (2.138)$$

Exercício Encontre a densidade de estados para partículas confinadas a uma região bidimensional no caso não relativístico ($\epsilon = p^2/2m$).

Exercício Para partículas relativísticas, ou excitações sem massa, $\epsilon = cp$. Obtenha $\rho(\epsilon)$ em duas e três dimensões.

2.10 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 2.119 e 2.122. O limite de temperatura $T \rightarrow 0$ é o ponto de partida, pois permite passar ao estudo de pequenas temperaturas. Isso parece óbvio, mas o que não é óbvio é qual o regime de valores para considerar temperaturas altas ou baixas. Um resultado surpreendente é que em muitos problemas a escala da temperatura que determina pequeno ou grande é da ordem de alguns milhares de graus Kelvin. Ou seja, quase todo o regime de interesse é em muitos casos o de baixa temperatura.

O número de partículas médio do sistema é dado por uma integral sobre os estados de energia com densidade $\rho(\epsilon)$

$$\begin{aligned} N &= \sum_{\mathbf{q}} \bar{n}_{\mathbf{q}} = \sum_{\mathbf{q}} \frac{1}{e^{\beta(\epsilon_{\mathbf{q}} - \mu)} + 1} \\ &\rightarrow \int_0^\infty \rho(\epsilon) \frac{1}{e^{\beta(\epsilon - \mu)} + 1} d\epsilon, \end{aligned} \quad (2.139)$$

onde lembramos que junto com vínculo

$$E = \int_0^\infty \rho(\epsilon) \frac{\epsilon}{e^{\beta(\epsilon-\mu)} + 1} d\epsilon, \quad (2.140)$$

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

$$N = \lim_{T \rightarrow 0} \int_0^\infty \rho(\epsilon) \frac{1}{e^{\beta(\epsilon-\mu)} + 1} d\epsilon, \quad (2.141)$$

usando que o número de ocupação se torna um função degrau, 1 para energias abaixo de μ_0 e zero acima

$$N = \int_0^{\mu_0} \rho(\epsilon) d\epsilon, \quad (2.142)$$

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 de o que é grande ou pequeno. Para partículas não relativísticas em três dimensões é fácil calcular usando a expressão 2.136. Podemos definir o módulo do momento da partícula com maior energia, o momento de ... Fermi = p_F , definido por $\mu_0 = p_F^2/2m$. Mas falta um detalhe. O elétron tem spin 1/2, isso faz com que haja dois estados diferentes com a mesma energia. Em geral se o spin for s o número será $g = 2s + 1$. portanto a densidade deve ser multiplicada por g :

$$\begin{aligned} N &= g \int_0^{\mu_0} \rho(\epsilon) d\epsilon, \\ &= gV \int_0^{\mu_0} C \epsilon^{\frac{1}{2}} d\epsilon = gVC \frac{2}{3} \epsilon_F^{\frac{3}{2}}, \end{aligned} \quad (2.143)$$

$$\begin{aligned} &= g \frac{4\pi V}{(2\pi\hbar)^3} \sqrt{2m}^{\frac{3}{2}} \frac{2}{3} \epsilon_F^{\frac{3}{2}} \\ &= \frac{gV(4\pi p_F^3)}{3(2\pi\hbar)^3}. \end{aligned} \quad (2.144)$$

Segue que

$$p_F = 2\pi\hbar \left(\frac{3}{4\pi g}\right)^{\frac{1}{3}} \left(\frac{N}{V}\right)^{\frac{1}{3}} \propto \left(\frac{N}{V}\right)^{\frac{1}{3}} \quad (2.145)$$

e

$$\epsilon_F = \left(\frac{3}{2gC}\right)^{\frac{2}{3}} \left(\frac{N}{V}\right)^{\frac{2}{3}} \quad (2.146)$$

$$\mu_0 = \epsilon_F = \frac{(2\pi\hbar)^2}{2m} \left(\frac{3}{4\pi g}\right)^{\frac{2}{3}} \left(\frac{N}{V}\right)^{\frac{2}{3}} \propto \left(\frac{N}{V}\right)^{\frac{2}{3}} \quad (2.147)$$

A energia média por partícula

$$\bar{\epsilon}(T, N) = \frac{g}{N} \int_0^\infty \rho(\epsilon) \frac{\epsilon}{e^{\beta(\epsilon-\mu)} + 1} d\epsilon \rightarrow \bar{\epsilon} = \frac{\int_0^{\epsilon_F} \epsilon \rho(\epsilon) d\epsilon}{\int_0^{\epsilon_F} \rho(\epsilon) d\epsilon}, \quad (2.148)$$

$$\bar{\epsilon} = \frac{\int_0^{\epsilon_F} \epsilon^{\frac{3}{2}} d\epsilon}{\int_0^{\epsilon_F} \epsilon^{\frac{1}{2}} d\epsilon} = \frac{3}{5} \epsilon_F. \quad (2.149)$$

A energia total

$$E = \frac{3}{5} N \epsilon_F = 3 \frac{(2\pi\hbar)^2}{10m} \left(\frac{3}{4\pi g}\right)^{\frac{2}{3}} \left(\frac{N}{V}\right)^{\frac{2}{3}} N \propto (V)^{-\frac{2}{3}} N^{\frac{5}{3}} \quad (2.150)$$

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:

$$P = - \frac{\partial E}{\partial V} \Big|_N = \frac{2E}{3V}. \quad (2.151)$$

No caso clássico tínhamos $E = 3NT/2 = \frac{3}{2}PV$ zero energia a temperatura zero, agora também temos

$$E = \frac{3}{2}PV, \quad (2.152)$$

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 10⁵ graus Kelvin.

2.10.1 Gás ideal de Férmions a temperatura finita

Pelo exercício acima vemos que temperaturas da ordem de 300K ainda são muito baixas. Olhe para a figura 2.4 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 2.5. 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 2.5 é diferente de zero.

As equações relevantes são 2.139 e 2.140 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 e devemos calcular

$$A(T, \mu) = \int_0^\infty \frac{a(\epsilon)}{e^{\beta(\epsilon-\mu)} + 1} d\epsilon, \quad (2.153)$$

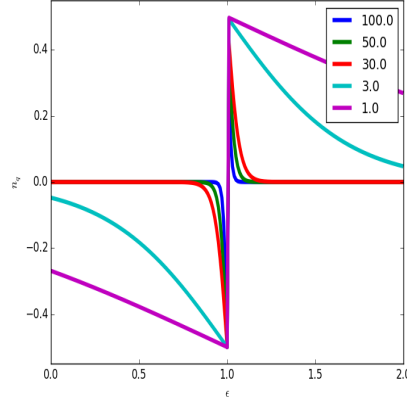


Figura 2.5: A função de Fermi a temperatura finita menos a função de Fermi a $T = 0$ (degrau): $f(\epsilon, \beta) = (\exp(\beta(\epsilon - 1)) + 1)^{-1} - \Theta(1 - \epsilon)$, para diferentes valores de β . A energia é medida em unidades do potencial químico.

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

$$A(T, \mu) = A(0, \mu) + (A(T, \mu) - A(0, \mu)) \quad (2.154)$$

onde, usando a função f (ver figura 2.5)

$$A(0, \mu) = \int_0^\mu a(\epsilon) d\epsilon \quad (2.155)$$

$$\begin{aligned} \delta A = (A(T, \mu) - A(0, \mu)) &= \int_0^\infty a(\epsilon) \left(\frac{1}{e^{\beta(\epsilon - \mu)} + 1} - \Theta(\mu - \epsilon) \right) d\epsilon \\ &= \int_{-\infty}^\infty a(\epsilon) f(\epsilon, T) d\epsilon. \end{aligned} \quad (2.156)$$

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

$$\int_{-\infty}^\infty a(\epsilon) f(\epsilon, T) d\epsilon = \sum_k \frac{T^{k+1}}{k!} \frac{d^k a}{d\epsilon^k} \Big|_\mu \int_{-\infty}^\infty \left(\frac{\epsilon - \mu}{T} \right)^k f(\epsilon, T) \frac{d\epsilon}{T} \quad (2.157)$$

onde colocamos alguns fatores de T para poder introduzir a variável $u = (\epsilon - \mu)/T$. Devido a que f é uma função impar para reflexões em torno de $u = 0$ teremos que as ordens pares se anulam

$$\int_{-\infty}^\infty u^{2k} f(u, T) du = 0 \quad (2.158)$$

se anulam e o primeiro termo na expansão que contribui é:

$$\begin{aligned}
 \int_{-\infty}^{\infty} u f(u, T) du &= \int_{-\infty}^{\infty} u \left(\frac{1}{e^u + 1} - \Theta(-u) \right) du \\
 &= \int_{-\infty}^0 u \left(\frac{1}{e^u + 1} - 1 \right) du + \int_0^{\infty} \frac{u}{e^u + 1} du \\
 &= - \int_{-\infty}^0 u \left(\frac{1}{1 + e^{-u}} \right) du + \int_0^{\infty} \frac{u}{e^u + 1} du \\
 &= 2 \int_0^{\infty} \frac{u}{e^u + 1} du = \frac{\pi^2}{6}.
 \end{aligned} \tag{2.159}$$

O próximo termo na expansão será o cúbico

$$2 \int_0^{\infty} \frac{u^3}{e^u + 1} du = \frac{7\pi^4}{60}. \tag{2.160}$$

Note que o expoente de T em 2.157 é $k + 1$, portanto a série em potências de T para $A(\beta, \mu)$ na equação 2.153 tem expoentes pares:

$$A(T, \mu) = \int_0^{\mu} a(\epsilon) d\epsilon + \frac{da}{d\epsilon} \Big|_{\mu} \frac{\pi^2}{6} T^2 + \frac{d^3 a}{d\epsilon^3} \Big|_{\mu} \frac{7\pi^4}{360} T^4 + \dots \tag{2.161}$$

Agora vamos a brincar com a expressão acima:

O potencial termodinâmico Φ e a energia E

Lembre que

$$\Phi(\beta, \mu, V) = -T \int_0^{\infty} \rho(\epsilon) \log(1 + e^{-\beta(\epsilon - \mu)}) d\epsilon.$$

Suponha que $\rho = CV\epsilon^b$, como é nos casos de interesse. Integramos por partes:

$$\Phi = -\frac{CV}{b+1} \epsilon^{b+1} \frac{1}{e^{\beta(\epsilon - \mu)} + 1} \Big|_0^{\infty} - \int_0^{\infty} \frac{CV}{b+1} \epsilon^{b+1} \frac{1}{e^{\beta(\epsilon - \mu)} + 1} d\epsilon$$

o termo de superfície é nulo, e o segundo termo (que muda duas vezes de sinal) é proporcional à energia

$$\begin{aligned}
 \Phi &= - \int_0^{\infty} \frac{\epsilon}{b+1} \rho(\epsilon) \frac{1}{e^{\beta(\epsilon - \mu)} + 1} d\epsilon \\
 &= - \frac{1}{b+1} \int_0^{\infty} \epsilon \rho(\epsilon) \bar{n}(\epsilon) d\epsilon \\
 &= - \frac{E}{b+1}
 \end{aligned} \tag{2.162}$$

$$= -\frac{2}{3} E \quad (b = 1/2) \tag{2.163}$$

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

$$\begin{aligned}\int_0^\mu CV\epsilon^{b+1}d\epsilon &= \frac{CV}{b+2}\mu^{b+2} = \frac{2CV}{5}\mu^{\frac{5}{2}} \\ \frac{da}{d\epsilon}|_\mu &= (b+1)CV\mu^b = \frac{5}{2}CV\mu^{\frac{1}{2}} \\ \frac{d^3a}{d\epsilon^3}|_\mu &= (b+1)b(b-1)CV\mu^{b-2} = -\frac{5}{8}CV\mu^{-\frac{3}{2}}\end{aligned}$$

e a expansão 2.161 até a primeira ordem não trivial (primeira derivada de a) fica, para $b = 1/2$:

$$\begin{aligned}\Phi(\beta, \mu, V) &= -\frac{2CV}{3} \left(\frac{2}{5}\mu^{\frac{5}{2}} + \frac{\pi^2}{4}\mu^{\frac{1}{2}}T^2 \right) \\ &= -\frac{4CV}{15}\mu^{\frac{5}{2}} \left(1 + \frac{5\pi^2}{8}\left(\frac{T}{\mu}\right)^2 \right)\end{aligned}\quad (2.164)$$

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 2.136

$$\Phi(\beta, \mu, V) = -\frac{4}{15} \frac{4\pi V g}{(2\pi\hbar)^3} \sqrt{2} m^{\frac{3}{2}} \mu^{\frac{5}{2}} \left(1 + \frac{5\pi^2}{8}\left(\frac{T}{\mu}\right)^2 \right) \quad (2.165)$$

O número de partículas N e o potencial químico μ

O número de partículas N para um dado valor de μ pode ser calculado com a função $a(\epsilon)$ apropriada ou usando a relação termodinâmica

$$\begin{aligned}N = -\frac{\partial\Phi}{\partial\mu}|_T &= \frac{2CV}{3} \left(\mu^{\frac{3}{2}} + \frac{\pi^2}{8}\mu^{-\frac{1}{2}}T^2 \right) \\ &= \frac{2CV}{3} \mu^{\frac{3}{2}} \left(1 + \frac{\pi^2}{8}\left(\frac{T}{\mu}\right)^2 \right).\end{aligned}\quad (2.166)$$

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

$$N(T = 0) = N_0 = \frac{2CV}{3} \mu^{\frac{3}{2}}, \quad (2.167)$$

ou para um dado N o potencial químico será μ_0 , obtido invertendo 2.167

$$\mu_0 = \left(\frac{3N}{2CV} \right)^{\frac{2}{3}}. \quad (2.168)$$

Considerando N fixo temos duas expressões (2.166 e 2.168) que permitem, ao eliminar N obter uma relação entre os potenciais químicos a temperatura zero

e finita:

$$N = \frac{2CV}{3} \mu_0^{\frac{3}{2}} = \frac{2CV}{3} \mu^{\frac{3}{2}} \left(1 + \frac{\pi^2}{8} \left(\frac{T}{\mu} \right)^2 \right). \quad (2.169)$$

Mantendo a expansão até segunda ordem em T , esta expresso pode ser invertida para obter

$$\mu = \mu_0 \left(1 + \frac{\pi^2}{8} \left(\frac{T}{\mu} \right)^2 \right)^{-\frac{2}{3}} \quad (2.170)$$

$$= \mu_0 \left(1 - \frac{\pi^2}{12} \left(\frac{T}{\mu} \right)^2 \right) \quad (2.171)$$

$$= \mu_0 \left(1 - \frac{\pi^2}{12} \left(\frac{T}{T_F} \right)^2 \right) \quad (2.172)$$

onde na segunda linha expandimos novamente $(1+x)^a \approx 1+ax$, e na terceira linha substituímos $\mu \approx \mu_0 = T_F$, pois o erro cometido ao fazer isso é de terceira ordem. O sinal negativo da correção mostra que o potencial químico diminui quando aumenta a temperatura, o que é esperado, pois a altas temperaturas no limite clássico. o resultado deve ser igual ao de bósons, que deve ser sempre negativo. Lembramos que $\mu = \partial E / \partial N$ a entropia constante, isto é a energia necessária para adicionar uma partícula a entropia constante. A altas temperaturas, se apenas se coloca uma nova partícula, o número de estados possíveis do sistema aumenta. Logo calor deve ser retirado para que a entropia desça até seu valor inicial. A baixas temperaturas a nova partícula só poderá entrar acima da energia de Fermi, sem que isso aumente o número de estados distinguíveis do sistema. Portanto $\mu > 0$.

A entropia S e o calor específico C_V

Usamos

$$S = - \frac{\partial \Phi}{\partial T} \Big|_{\mu, V}$$

e a expressão 2.164

$$S = \frac{CV\pi^2}{3} \mu^{\frac{1}{2}} T. \quad (2.173)$$

Mas por 2.167, $\mu^{\frac{1}{2}} = (3/2C)^{1/3} (N/V)^{1/3}$ e $\epsilon_F = (\frac{3}{2C})^{\frac{2}{3}} (\frac{N}{V})^{\frac{2}{3}}$

$$S = \frac{\pi^2 N}{2} \frac{2C}{3} \frac{V}{N} (3/2C)^{1/3} (N/V)^{1/3} T = \frac{\pi^2 N}{2} \frac{T}{T_F} \quad (2.174)$$

e o calor específico

$$C_V = T \left(\frac{\partial S}{\partial T} \right)_V = \frac{\pi^2 N}{2} \frac{T}{T_F}. \quad (2.175)$$

Podemos escrevê-lo de uma forma que lembra a equipartição

$$C_V = \frac{3}{2} N_{ef} \quad (2.176)$$

onde $N_{ef} = \frac{\pi^2}{3} \frac{T}{T_F}$ é o número efetivo de elétrons que contribui para o calor específico. São aqueles que podem ser excitados a estados desocupados por energias da ordem de $k_B T$ ⁵, os outros estão muito afundados nos estados de menor energia para poder sair da superfície de Fermi.

Exercício Calcule para cobre o valor de previsto de $C_V/Nk_B T$ 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 [?].

2.11 Bósons: 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 2.132:

$$\rho(\epsilon)d\epsilon = \mathcal{N}_3 = 2 \frac{L^3 \Delta p_x \Delta p_y \Delta p_z}{(2\pi\hbar)^3}, \quad (2.177)$$

só que agora a relação de dispersão entre energia e momento é diferente e o prefatr 2 é devido aos dois estados de polarização da luz⁶.

$$\epsilon = \hbar\omega, \quad \mathbf{p} = \hbar\mathbf{q}, \quad |\mathbf{p}| = \frac{\hbar\omega}{c} \quad (2.178)$$

$$\rho(\epsilon)d\epsilon = 2 \frac{V 4\pi p^2}{(2\pi\hbar)^3} dp, \quad (2.179)$$

⁵Essencialmente nestas notas, como medimos temperatura em unidades de energia, a constante de Boltzmann $k_B = 1$.

⁶O campo do fóton é um vetor, tem spin 1, mas não são 3 estados porque a massa é nula

$$\rho(\epsilon)d\epsilon = \frac{8\pi V \omega^2}{(2\pi c)^3} d\omega, \quad (2.180)$$

O número de ocupação médio de um modo com frequência ω é obtido da expressão 2.126, 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:

$$\bar{n}_\omega^P = \frac{1}{e^{\beta\hbar\omega} - 1}. \quad (2.181)$$

A energia total será a integral sobre todas as frequências, da energia associada ao estado de frequência $\hbar\omega$, vezes o número de ocupação $\bar{n}_\omega^P$ e densidade de estados de frequência ω :

$$\begin{aligned} E &= \int_0^\infty \frac{8\pi V \omega^2}{(2\pi c)^3} \frac{\hbar\omega}{e^{\beta\hbar\omega} - 1} d\omega \\ &= \frac{8\pi V \hbar}{(2\pi c)^3} \int_0^\infty \frac{\omega^3}{e^{\beta\hbar\omega} - 1} d\omega \end{aligned} \quad (2.182)$$

Podemos escrever $E = \int u(\omega, T) d\omega$, definindo a densidade de energia espectral:

$$u(\omega, T) = \frac{8\pi V \hbar}{(2\pi c)^3} \frac{\omega^3}{e^{\beta\hbar\omega} - 1} \quad (2.183)$$

que é o famoso resultado de Planck. Mudando variáveis $x = \beta\hbar\omega$ temos

$$\begin{aligned} E &= T^4 \frac{V}{\pi^2 (\hbar c)^3} \int_0^\infty \frac{x^3}{e^x - 1} dx \\ \frac{E}{V} &= \sigma_{SB} T^4 \end{aligned} \quad (2.184)$$

onde a última linha expressa o que é chamado de lei de Stephan-Boltzmann e a constante σ_{SB} pode ser calculada usando

$$\int_0^\infty \frac{x^3}{e^x - 1} dx = \frac{\pi^4}{15} \quad (2.185)$$

o que leva, após recolocar a constante de Boltzmann, a

$$\sigma_{SB} = \frac{\pi^2 k_B^4}{15 (\hbar c)^3} \quad (2.186)$$

Exercício Mostre que a pressão do gás de fótons é $P = E/(3V)$.